菰米集

卷八

野蕃哥哥

南非比勒陀利亚 Pretoria

2021.09

目录

卷八：

[J190] X. Xia, Control problems in building energy retrofit and maintenance planning, *Annual Reviews in Control*, vol. 44, 2017, pp. 78-88.

[J191] E. Wanjiru and X. Xia, Sustainable energy-water management for residential houses with optimal integrated grey and rain water recycling, *Journal of Cleaner Production* 170 (2018), pp. 1151-1166.

[J192] H. Carstens, X. Xia and S. Yadavalli, Measurement uncertainty in energy monitoring: present state of the art, *Renewable and Sustainable Energy Reviews*, 82 (2018), pp. 2791-2805.

[J193] H. Carstens, X. Xia and S. Yadavalli, Bayesian energy measurement and verification analysis, *Energies* 2018, 11, 380; doi:10.3390/en11020380.

[J194] Q. Li, Z. Wu and X. Xia, Estimate and characterize PV power at demand-side hybrid system, *Applied Energy* 218 (2018), pp. 66-77.

[J195] Y. Fan and X. Xia, Energy-efficiency building retrofit planning for green building compliance, *Building and Environment* 136 (2018), pp. 312-321.

[J196] J. Mei, X. Xia and M. Song, An autonomous hierarchical control for improving indoor comfort and energy efficiency of a direct expansion air conditioning system, *Applied Energy*, 221 (2018), pp. 450-463.

[J197] J. Yan, H. Yu and X. Xia, Distributed optimization of multi-agent systems with delayed sampled-data, *Neurocomputing*, 296 (2018), pp. 100-108.

[J198] C. Kagiri, E. M. Wanjiru, L. Zhang and X. Xia, Optimized response to electricity time-of-use tariff of a compressed natural gas fuelling station, *Applied Energy*, 222 (2018), pp. 244-256.

[J199] Z. Wu and X. Xia, Tariff-driven demand side management of green ship, *Solar Energy* 170 (2018), pp. 991-1000.

[J200] L. Zhang, M. Chennells and X. Xia, A power dispatch model for a ferrochrome plant heat recovery cogeneration system, *Applied Energy* 227 (2018), pp. 180-189.

[J201] Y. Fan and X. Xia, Building retrofit optimization models using notch test data considering energy performance certificate compliance, *Applied Energy* 228 (2018), pp. 2140-2152.

[J202] M. S. Masaki, L. Zhang and X. Xia, A design approach for multiple drive belt conveyors minimizing life cycle costs, *Journal of Cleaner Production* 201 (2018), pp. 526-541.

[J203] Z Wu, K. Zhao and X Xia, Lighting retrofit and maintenance models with decay and adaptive control, *IET Control Theory & Applications*, vol. 12, no. 5, 2018, pp. 593-600.

[J204] F. Barzegar, L. Zhang, A. Bello, N. Manyala and X. Xia, Three dimension modeling of the components in supercapacitors for proper understanding and contribution of each parameter to the final electrochemical performance, *Journal of Materials Chemistry A*, 2018, DOI: 10.1039/C8TA04736G

[J205] B. Zhu, Z. Ren, W. Xie, F. Guo and X. Xia, Active nonlinear partial-state feedback control of contacting force for a pantograph-catenary system, *ISA Transactions*, 91, 2019, pp. 78-89. https://doi.org/10.1016/j.isatra.2019.01.033

[J206] M. S. Masaki, L. Zhang and X. Xia, A hierarchical predictive control for supercapacitor-retrofitted grid-connected hybrid renewable systems, *Applied Energy*, 242 (2019), pp. 393-402.

[J207] C. Kagiri, L. Zhang and X. Xia, A hierarchical optimisation of a compressed natural gas station for energy and fuelling efficiency under a demand response program, *Energies* 2019, 12, 2165; doi: 10.3390/en12112165.

[J208] B. Wang, Y. Li, F. Yang, and X. Xia, A competitive swarm optimizer-based technoeconomic optimization with appliance scheduling in domestic PV-battery hybrid systems, *Complexity*, vol. 2019 |Article ID 4824837 | https://doi.org/10.1155/2019/4824837

[J209] J. Mei and X. Xia, Distributed control for a multi-evaporator air conditioning system, *Control Engineering Practice*, vol. 90, 2019, pp. 85 -100.

Annual Reviews in Control 44 (2017) 78–88

Contents lists available at ScienceDirect

Annual Reviews in Control

journal homepage: www.elsevier.com/locate/arcontrol

Review article

Control problems in building energy retrofit and maintenance planning☆

Xiaohua Xia

Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa

ARTICLE INFO

Article history:
Received 16 January 2017
Accepted 15 April 2017
Available online 20 April 2017

Keywords:
Energy efficiency
Optimal control
Maintenance planning

ABSTRACT

This paper presents a series of control problems in prioritizing building energy retrofit and maintenance plans through a review of recent studies. The building energy retrofits can be strategically performed on policy level, management level, system level and unit level. Based on existing research efforts, this study casts the optimal building maintenance planning problem into a general control system framework. Unlike traditional control applications, this study argues that the control system framework is also applicable to the building energy management level, which will significantly improve the sustainability of realized energy savings and cost-effectiveness of building energy retrofits. In a general control framework, a number of research problems in the control systems are formulated, namely 1) control system decay dynamics modeling; 2) control system inputs and model uncertainties; 3) control system outputs; 4) control system uncertainties and disturbances; 5) control system algorithms; and 6) grouping and modeling. The proposed control problems bring out the intrinsic relationship of reliability engineering, maintenance engineering and control engineering in the broad directions of energy efficiency and optimization. Investigations into the proposed control problems will contribute to further improvements in the building energy retrofit and maintenance plans than the currently prevailing engineering practice.

1. Introduction

Developments of building energy efficiency technologies enable various optional energy conservation measures (ECMs) to improve the building energy performance. However selection of prioritized ECMs for a building retrofit plan is very challenging due to technical barriers and financial barriers. The building energy retrofits can be strategically performed on policy, management, system and unit levels with different aspects of addressing a spectrum of financial and technical barriers. Technically, building energy retrofit is a lengthy process that includes energy audit, baseline development, retrofit planning, implementation and commission, operations and maintenance (O&M), and measurement and verification (M&V). The complexity of an effective building energy retrofit demands a non-trivial amount of information and expert knowledge about building construction, operation, and energy consumption before and after the retrofit. Financially, cost-effectiveness is usually the first concern of a building energy retrofit plan. The achievable energy and cost savings are the primary attractions to the building retrofit investors and building owners. In buildings, such energy and cost savings can be achieved from many components or subsystems. As a complex system, there are many components that pertain to power generation, building materials and envelope, electricity appliances, water consuming appliances, etc. One or several of such components can comprise a subsystem that provides additional and enhanced functionality to the building. In the building context, many such subsystems can be identified to present energy efficiency opportunities. These energy efficiency opportunities are roughly categorized into four technical functional layers, namely the power electronics layer, smart appliance layer, energy flow layer and planning layer. The power electronics layer involves energy optimization that focus on the maintaining and improving the power quality (Abo-Al-Ez, Elaiw, & Xia, 2014; Esmaeli, 2016; Liu, Zhang, Wang, & Wang, 2014; Mokgonyana, Zhang, Zhang, & Xia, 2016; Nikkhajoei & Lasseter, 2009; Sao & Lehn, 2005; Wilson, Robinett, Weaver, Byrne, & Young, 2016; Yu, Khambadkone, Wang, & Terence, 2010), which is essential to guarantee the performances of all electricity consuming components in the building. The smart appliance layer improves the building energy efficiency by bringing in energy efficiency intelligence to the appliances in addition

☆ A semi-plenary was presented at 10th IFAC Symposium on Nonlinear Control Systems, Monterey, USA, 23–25 August 2016 based on a preliminary version of the paper.

E-mail address: xxia@up.ac.za

http://dx.doi.org/10.1016/j.arcontrol.2017.04.003
1367-5788/

to the built-in control logic (Arens, Federspiel, Wang, & Huizenga, 2005; Bijker, Xia, & Zhang, 2009; Catherine, Wheeler, Wilkinson, & de Jager, 2012; Mei, Zhu, & Xia, 2015a; 2015b; Portmess & Tower, 2015; Setlhaolo & Xia, 2015; 2016; Setlhaolo, Xia, & Zhang, 2014; Stavropoulos et al., 2015; Stavropoulos, Koutitas, Vrakas, Kontopoulos, & Vlahavas, 2016; Wang, Zhang, & Xia, 2013). The energy flow layer focuses on the energy efficiency opportunities from balancing different energy sources (Elaiw, Xia, & Shehata, 2012; 2013; Ntsaluba, Zhu, & Xia, 2016; Nwulu & Xia, 2015a; 2015b; 2015c; 2017; Sichilalu, Tazvinga, & Xia, 2016; Sichilalu & Xia, 2015a; 2015b; Tazvinga, Xia, & Zhang, 2013; Tazvinga, Zhu, & Xia, 2014; 2015; Wu, Tazvinga, & Xia, 2015a; Xia & Elaiw, 2010; Xia, Zhang, & Elaiw, 2009; 2011; Zhang & Xia, 2011; Zhu, Tazvinga, & Xia, 2015). The planning level actually contributes a series of investment decisions and budget competitions at the building energy management level to improve the overall cost effectiveness, or overcoming the financial barrier of an energy efficiency project, e.g., a retrofit project (Malatji, Zhang, & Xia, 2013; Wang, Xia, & Zhang, 2014; Wu, Wang, & Xia, 2016; Wu, Xia, & Wang, 2015b). Although huge amount of studies have been conducted at the first three layers, in particular, in the framework of a control system (Xia & Zhang, 2010; 2011; 2015; Xia, Zhang, & Cass, 2012; Xia & Zhang, 2016), there lacks a systematic method to model, evaluate and optimize the building retrofit plans at the management level. Furthermore, perceived uncertainty in realized energy savings and the risk of underachieving the projected savings prevent investors and building owners from pursuing a building retrofit. At the current stage, the building energy guideline (USDOE, 2011) indicates that the energy savings of the building energy retrofit actions are quantified by the M&V process. But the verified energy savings usually couple with uncertainties from measurement, sampling and modeling efforts during the M&V process (Carstens, Xia, Zhang, & Ye, 2013; Xia & Zhang, 2013; Ye & Xia, 2014; 2016; Ye, Xia, & Zhang, 2013; 2014).

In order to facilitate the building energy retrofit, a number of energy efficiency (EE) incentive programmes and policies have been implemented to address both the technical and financial barriers, such as clean development mechanism (CDM) (Michaelowa & Jotzo, 2005), tradable white certificate (TWC) scheme (Bertoldi & Rezessy, 2008; Mundaca, 2007), demand side management (DSM) programmes (Eskom, 2011), and performance contracting (Mozzo, 1999). Taking advantage of the EE programme incentives usually improves the building energy retrofit projects' cost-effectiveness when regulations of the EE programmes are properly followed. In general, accuracy and persistency of the achieved energy savings are the major concerns in these EE programmes' crediting period. However, the energy savings from most building retrofit projects are often not sustainable given that the retrofitted EE devices will fail over time. On identification of the device failures, some existing EE programme guidelines apply a penalty factor during the energy savings accounting process instead of requiring direct maintenance activities to correct the devices failures. For instance, the CDM guidelines (UNFCCC, 2007; 2010) apply a penalty factor, which is called lamp failure rate (LFR) to the energy savings calculation and further restrict that no project rebates will be issued to the implemented projects when 50% of the initial population is failed during the project crediting period. Under these rules, the lighting projects are only considered sustainable when the survived lighting population is equal to or greater than 50% of their initial population by proper maintenance. Some latest designed lighting project guidelines UNFCCC (2011; 2012) request to perform continuous replacements of all the failed lamps. Practically, the following barriers hold the investors and building owners back from performing such a full maintenance policy. Firstly, the full maintenance is not easily implementable due to the demand of continuously monitoring and sampling the lighting devices' working conditions. Secondly, the maintenance activities also require additional investments for the procurement and installation of the new lighting devices. The extra investments sometimes contribute to a tighter project budget.

Since neither the "no maintenance" nor the "full maintenance" policy is preferable to the investors and building owners, it is thus interesting to design an optimal maintenance plan to the whole building energy retrofit process to improve its cost-effectiveness. The optimal maintenance planning (OMP) problem can be aptly formulated under the control system framework as a control problem. The control system framework is applicable for this purpose since the population decay dynamics of the retrofitted EE devices are characterized and modeled as state space equations. The population decay dynamics are taken as the plant of the control system. In order to achieve sustainable energy savings and maximum project profits, it is recommended to optimally control/replace a number of failed EE devices during each maintenance interval. The number of failed items to be replaced is taken as the control variable of the control system. As different EE technologies have different population decay dynamics and different rebate tariffs, the control inputs can be optimally decided based on the investors and building owners' budget availability.

Formulating the OMP problem into control problems exhibits following advantages. Firstly, under the control system framework, classic control theories and methodologies can be applied to improve the designed maintenance strategy. Secondly, applying the control system approach to solve the OMP problem on the building energy management level significantly improve the sustainability of realized energy savings and cost-effectiveness of building energy retrofits when comparing to traditional control applications in the building energy retrofit areas. Thirdly, the proposed control system approach also brings out the intrinsic relationship of reliability engineering and control engineering. One major issue to design the optimal maintenance plan is to characterize the population decay and performance deterioration dynamics of the building EE devices, where some deterministic or stochastic models of energy device reliability can be found in the existing reliability engineering studies. For instance, a series of common failure distributions, reliability and hazard rate functions for EE devices with various reliability characteristics are provided by O'Connor and Kleyner (2011), according to which the population degradation of various types of retrofitted items, e.g., the non-repairable products and the repairable products, can be characterized. In addition, the new applications of maintenance activities for EE purposes by the control system approach also bring new developments in the reliability engineering field, such as investigations and model developments on the population decay or performance degradation of various EE building appliances. The relevant research progress in the reliability engineering area will facilitate the control applications in improving the building EE managements, and vice versa.

Based on existing research efforts in the literature, this study casts the optimal building maintenance planing problem into a general control system framework. From the general control formulation, a number of research problems in the control systems are systematically discovered, namely 1) control system decay dynamics modeling; 2) control system inputs and model uncertainties; 3) control system outputs; 4) control system uncertainties and disturbances; 5) control objective function formulation; 6) control system algorithms; and 7) grouping and modeling. Furthermore, a case study is given to illustrate the application of control system framework in practical OMP problems. Detailed research proposals will be given in next section.

2. Control problems in building energy retrofit and maintenance

In this section, the OMP problem is mathematically formulated under the control system framework. Thereafter, the control prob-

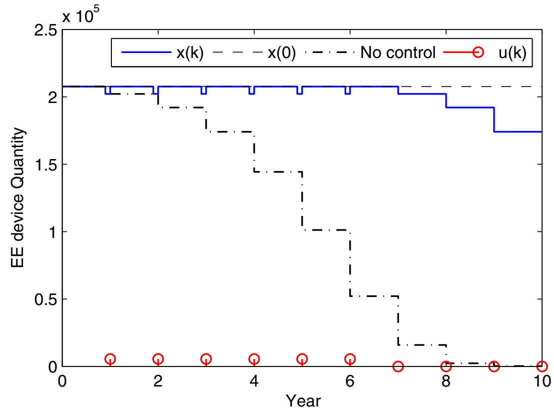


Fig. 1. Optimal maintenance strategy of a lighting retrofit project.

lems that related to the OMP problems are identified systematically, which are introduced in detail in the following subsections.

2.1. Beginning of the story

The idea to use control approach to solve the building energy retrofit and maintenance problem starts with a large-scale lighting retrofit project. To keep the story simple but interesting, only part of the whole project is introduced below. A lighting retrofit project that aims to reduce the lighting load in a fleet of commercial buildings is going to be implemented. A number of 207 693 energy efficient LEDs will be installed to replace existing inefficient halogen downlighters (HDLs). The LEDs must have very high quality with a rated life of equal to or more than 6 years. As an energy efficient project with new technologies, project developers (PDs) will receive a rebate of R[1] 0.55 per kWh savings realized from the implementation of this project. More project details from the energy audit of this project are found in Ye, Xia, Zhang, and Zhu (2015).

Scope of this project sounds easy, however PDs must comply with following general project regulation policies in order to receive their project rebates.

1. PDs will have implement the project at their own cost.
2. The crediting period of this project is 10 years during which PDs can receive their rebate on annual basis. All newly installed EE devices must be properly maintained. If more than 50% of the LEDs is malfunctioned, the rebate will be ceased.
3. The performance of the project will be reported once a year by a third-party M&V inspection company. This M&V company verifies the number of survived lamps by sampling and surveys. Once device failures are observed, PDs' are allowed to replace some (or all) of the failed EE devices at the end of each crediting year.

The first item on the programme policy makes the project a risky project since the PDs have to deposit a big amount initial investment. According to intensive project performance evaluation experience, professional M&V practitioners help the PDs design an optimal maintenance plan to replace a number of failed lamps at end of each year. The replacements keep the project rebate sustainable and ensure the savings' persistency of this project. At end of the project, PDs receive their maximum profits and the project produces its maximum energy savings. The M&V practitioner's strategy is illustrated in Fig. 1, in which the circled stems (in Red) denotes the number of replacement of failed LEDs at end of each year. The solid step lines (in Blue) represents the survived LEDs with replacements. The dash-dotted lines (in Black) show the survived LEDs would have been observed without replacements of failed ones. When looking at this figure, the PDs realize that they may only be able to claim rebates on project savings over the first 5 years due to unattended LED failures. The replacements of the failed lamps save the life of this project but necessary cost will be incurred for the rescue. To make final decisions on the project investments, the PDs asked the following immediate questions:

1. How many lamps have to be replaced? How do they cost, and when to replace?
2. If the replacements are helpful to generate more energy savings and rebates, then shall all the failed LEDs to be replaced at end of each year to produce the maximum benefits in terms of both energy savings and financial rebates?
3. If the M&V professional's maintenance plan helps with this lighting retrofit project, could more maintenance plans to be designed for other building energy retrofit projects with water heating devices, HVAC systems, plug-loads, and building envelopes?

2.2. General control system framework

Formulation of the general framework is introduced as follows. Let a large-scale building energy retrofit project to be implemented in a fleet of buildings, this project aims to replace N units of inefficient building appliances by energy efficient ones. The project is financially supported by local government through an incentive EEDSM programme, which awards an energy saving crediting period of 10 years for each implemented project. The programme regulations further request that the survived project population needs to be carefully maintained to guarantee the sustainability of the projected energy savings. Let t_0 and t_f denote the beginning and end of the project crediting period, respectively. I denoted the number of homogenous groups of the lighting population. $x_i(0)$ denotes the quantity of the initial installation of the EE devices in the ith group. Generally, the OMP problem is to find the optimal control sequences $\mathbf{u} = [u_1, u_2, \ldots, u_I]^T$ within the project crediting period. Here u_i is the control system input, which represents the number of replacements of the failed EE devices during each maintenance interval in the ith group. Then the OMP problem under the control system framework is formulated as:

$$\begin{cases} \dot{x}_i = f_i(\mathbf{x}, \mathbf{u}) + d_i, \\ y_j = h_j(\mathbf{x}) + \omega_j, \end{cases} \tag{1}$$

where $\mathbf{x}$ denotes the state variable that corresponds to the number of survival EE devices for each maintenance interval. The control system output y_j can be expressed by the sampling and measurement result of $\mathbf{x}$. $f_i(\cdot)$ denotes the function to characterize the project population decay dynamics. In addition, d_i and ω_j denote the modeling uncertainties and measurement disturbances, respectively. In the following subsections, each component in Eq. (1) will be discussed and relevant control problems will be identified for future research.

2.3. Control system decay dynamics modeling

In order to solve the OMP problem by a control system approach, the population decay dynamics model $f(\mathbf{x}, \mathbf{u})$ in Eq. (1) needs to be characterized. This links to the classic control system dynamics modeling problem, but in a new engineering field. Desired research efforts in this area can be summarized as to develop and validate a series of population decay dynamics and energy performance degradation models for various building appliances. Generally, building energy appliance are categorized into reparable and non-reparable ones. A repairable appliance can have

[1] South African Rand

multiple minor failures and be repaired before becoming salvaged. Air conditioners, heat pumps or printers are repairable appliances. A non-repairable item can only experience one catastrophic failure before the salvage. A replacement is required to remove such failure. CFLs or motion sensors are non-repairable appliances. The failure rates of the repairable and non-repairable items are usually different. The repairable and non-repairable classification is investigated at the current stage, and it is believed that there are many other available classifications in different scenarios, which remain uninvestigated. Consequently, two types of models can be developed to characterize the control system decay dynamics, namely population decay models for non-reparable failures and energy performance degradation models for reparable failures. The population decay dynamics models investigated here are merely a small part of a broad field of the reliability engineering. There are many other available models, corresponding to different categories of EE devices. It is expected that the research progress in the reliability engineering area will facilitate the advance of building energy optimization studies, and vice versa.

In our studies, we have come across non-reparable population decay models, such as the clean development mechanism (CDM) linear lamp population decay model (UNFCCC, 2010), CFL population decay model from the PELP study (Navigant, 1999), as well as reparable population decay models, such as the solar panel performance degradation model (Fan & Xia, 2015; 2017) and the exponential population decay model (Wang & Xia, 2015b), and possibly hybrid models such as the population decay models including interactive energy systems (Wang & Xia, 2015a) and a multi-stage performance degradation model (Wang, Wu, & Xia, 2017).

A linear lamp population decay model is proposed in the AMS-II.J (UNFCCC, 2010) as given in Eq. (2)

$$f(t) = \begin{cases} t \times H \times \frac{100-Y}{100 \times L}, & \text{if} \quad t \times H < L, \\ 100\%, & \text{if} \quad t \times H \geq L, \end{cases} \tag{2}$$

where $f(t)$ denotes the percentage of lamps that fails to work in the t-th year since installation, H is the annual average operating hours, L is the rated life span (in hours), and when $t \times H \geq L$, $f(t) = 100\%$, all lamps are deemed to be failed.

In Fan and Xia (2015), a function $E(t)$ is applied to characterize the solar panel linear output degradation over years, where

$$E(t) = -0.007(t-1) + 0.98. \tag{3}$$

An exponential degradation model is investigated to model the repairable failures in O'Connor and Kleyner (2011) and applied in the studies (Wang & Xia, 2014; 2015b), as shown in Eq. (4),

$$x(k) = x(0)e^{-\zeta_i k}. \tag{4}$$

The state space form of Eq. (4) is

$$x(k+1) = x(k)(1-\zeta_i), \tag{5}$$

or in continuous time,

$$\dot{x} = -\zeta_i x,$$

where θ_i denotes the Mean Time between Failures (MTBF) of the EE items, and ζ_i is calculated by:

$$\zeta_i = (\theta_i)^{-1}. \tag{6}$$

Although widely used, the models (2)–(5) are not accurate enough to characterize the lamp population decay dynamics due to unrevealed model uncertainties. For instance, the model (2) assumes a constant failure of the lighting devices, which turned out to be inaccurate from the PELP study report (Navigant, 1999). The model (3) assumes a constant performance degradation rate of the solar panels given that model does not consider the actual installation position and weather conditions of the solar panels. As commented in Carstens et al. (2013) and Carstens, Xia, and Ye (2014), the model (5) is also inappropriate to assume a constant hazard rate of the EE lighting devices.

In order to improve the model accuracy of the population decay dynamics of the EE devices, studies Carstens et al. (2013), Carstens et al. (2014) offer informative reviews on the existing lamp population decay dynamics models (Navigant, 1999), and also proposed a general form of the population decay dynamics model by re-calibrating existing models established from biological population dynamics study or from reliability engineering experiments. The general form of the model is provided in Eq. (7).

$$s(t) = \frac{1}{c + ae^{bt}}, \tag{7}$$

where $s(t)$ is the percentage of survived devices at time t for a lighting project, t is counted from the implementation of a lighting retrofit project. $a = e^{-L}$ and L is the rated average life span of a certain model of the EE devices. The rated average life span is declared by the manufacturer or responsible vendor as being the expected time at which 50% of any large number of EE devices reach the end of their individual lives (UNFCCC, 2010). b is the slope of decay and c is the initial percentage lamp survival at $t = 0$. Thus, with a given L, b and c can be obtained by solving the following equations:

$$\begin{cases} s(0) = 1, \\ s(L) = 0.5. \end{cases} \tag{8}$$

The model (7) is more advantageous than the models (2)–(5) as it has a validated model uncertainty quantified by R^2=0.996. The model (7) has acceptable accuracy level to describe the population decay dynamics due to non-reparable failures. An equivalent biological population dynamic model was proposed in Carstens et al. (2014, 2013) as the following:

$$\frac{\mathrm{d}s(t)}{\mathrm{d}t} = -bs(t)(1 - cs(t)), \tag{9}$$

or in its discrete-time form,

$$s(k+1) = bc(s(k))^2 \Delta t - bs(k)\Delta t + s(k). \tag{10}$$

In practice, energy performance of some EE devices does not simply drop from a good condition to a sudden failure. Failure mechanisms of the EE devices may experience a series of performance degradation process in real cases. For instance, energy performance of EE devices may have multiple functional stages such as good status, average status, bad status, and failed status. In addition, proper maintenance actions have to ability to restore the degraded performance into a better condition if taken before the salvage. In order to characterize the performance degradation process of the EE devices, the study (Wang et al., 2017) employs a state-transition model of items from homogenous groups, as shown in Eq. (11). Eq. (11) and Fig. 2 depict a following state transition mechanism. Given a series of discrete time instants t_k, $k = 1, 2, \ldots,$ the working state of an EE device has a possibility to jump to another state over interval $[t_k, t_{k+1})$. In Fig. 2, $P_{l,\,i}(t_k)$, $i \in [1, M_l]$ denotes the probability that this item works under state i at instant t_k. $\lambda^l_{i,i-1}(t_k)$ indicates the state-transition from state i to state $i-1$ over the interval $[t_k, t_{k+1})$. The circle F denotes the malfunctioning state and $P_{l,\,F}(t_k)$ the probability of this item being malfunctioning. $\lambda^l_{i,F}(t_k)$ indicates the state-transition from state i to malfunctioning. As shown in Fig. 2, $P_{l,\,i}(t_k)$ increases due to transition $\lambda^l_{i+1,i}(t_k)$, decreases due to transition $\lambda^l_{i,i-1}(t_k)$ and transition $\lambda^l_{i,F}(t_k)$ simultaneously. $P_{l,M_l}(t_k)$ continuously decreases and $P_{l,\,F}(t_k)$ continuously increases without maintenance. An early study on the production and maintenance control for manufacturing systems (Boukas & Liu, 2001) formulates such state-transition as a partially observable Markov decision process (POMDP), with a hypothesis that the

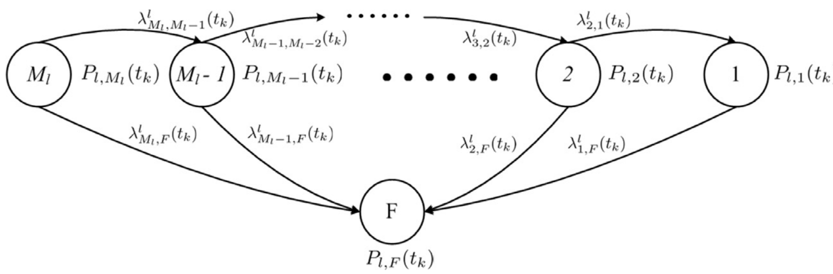


Fig. 2. The state transition of an item with M_l working states (Wang et al., 2017).

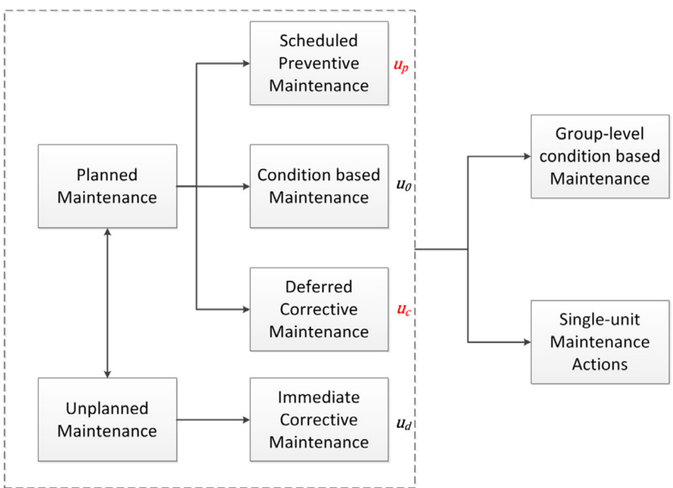


Fig. 3. Control system inputs.

transition rate to the next state depends on the current state. In Wang et al. (2017), it is assumed that for a homogeneous group l of such devices, the population dynamics of group l is commensurate with the individual item state-transition. Taking advantage of the POMDP formulation in Boukas and Liu (2001), the population changes $\triangle x_{l,\,i}(t_k)$ with $i = 1, 2, \ldots, M_l$ in group l are formulated in (11), where $f^l_{i,i-1}(\mathbf{x}_{l,i}(t_k))$ denotes the population change from subset i to subset $i-1$ that is resulted from the transition $\lambda^l_{i,i-1}(t_k)$.

$$\begin{cases} \triangle x_{l,M_l}(t_k) = -f^l_{M_l,M_l-1}(\mathbf{x}_{l,M_l}(t_k)) - f^l_{M_l,F}(\mathbf{x}_{l,M_l}(t_k)) + \sum_{i=1}^{M_l-1} u^l_i(t_k) + u^l_C(t_k), \\ \triangle x_{l,M_l-1}(t_k) = f^l_{M_l,M_l-1}(\mathbf{x}_{l,M_l}(t_k)) - f^l_{M_l-1,Ml-2}(\mathbf{x}_{l,M_l-1}(t_k)) - f^l_{M_l-1,F}(\mathbf{x}_{l,M_l-1}(t_k)) - u^l_{M_l-1}(t_k), \\ \vdots \\ \triangle x_{l,2}(t_k) = f^l_{3,2}(\mathbf{x}_{l,3}(t_k)) - f^l_{2,1}(\mathbf{x}_{l,2}(t_k)) - f^l_{2,F}(\mathbf{x}_{l,2}(t_k)) - u^l_2(t_k), \\ \triangle x_{l,1}(t_k) = f^l_{2,1}(\mathbf{x}_{l,2}(t_k)) - f^l_{1,F}(\mathbf{x}_{l,1}(t_k)) - u^l_{1,M_l}(t_k), \end{cases} \tag{11}$$

The introduced system decay dynamics models (2)–(11) are capable of characterizing the population decay or energy performance degradation dynamics for homogeneous group of EE devices, despite that there are modeling uncertainties involved in these models. In practice, interactive effects are sometimes observed across different EE device groups. For instance, existing study (Ahn, Jang, Leigh, Yoo, & Jeong, 2014) shows that heat gain from lights can significantly influence the energy consumptions of the air conditioning system. In this case, the population dynamics of the lighting group can pose significant impact on the energy performances of the air conditioners. The impact of such interaction is worthy taking into account in the control system dynamics formulation. As introduced in Wang and Xia (2015a), the population decay dynamics of interacting energy systems such as lighting and HVAC systems are formulated as

$$\begin{cases} \dot{x}_L = \hat{f}_1(x_L) + u_L, \\ \dot{x}_{AC} = \hat{f}_2(x_L, x_{AC}) + u_{AC}, \end{cases} \tag{12}$$

where x_L and x_{AC} are the state variables representing the survived lighting and HVAC systems, respectively. $\hat{f}_1(x_L)$ denotes the lamp population decay dynamics and $\hat{f}_2(x_L, x_{AC})$ denotes the HVAC system decay dynamics with the interaction of heating and cooling load from the lighting systems. An assumption is made in Eq. (12) that the HVAC systems have minimum impact to the lighting systems' life span. More details of the formulation and parameter identification in Eq. (12) can be found in Wang and Xia (2015a). Besides the lighting and HVAC system interactions, there are also other interactive effects observed to influence the reliability performance and modeling accuracy of the system decay dynamics. For instance, interactive reliability performance can be observed for a group of EE devices including both the newly retrofitted EE devices and old ones.

2.4. Control system inputs

The OMP problem is indeed an optimal control system input design problem. In the OMP scenario, the control inputs refer to the maintenance actions, which are described by the maintenance intensity and maintenance instant. The term 'maintenance intensity' describes the count of the restored items from one item group at a specific instant. Such instant is referred to as the 'maintenance instant', i.e., a time point at which the maintenance actions are scheduled to take place. The collection of maintenance instants comprise the maintenance time schedule. The maintenance intensity and time schedule are both promising optimization variables to improve the energy efficiency and cost-effectiveness of a retrofitting project.

According to the reliability engineering, there are several types of maintenances, corresponding to different purposes and strategies. At the current stage, a maintenance action classification from BSI (1984) is employed, where maintenance actions are grouped into several categories. As a fact to accommodate various EE devices' failure characteristics and mechanisms, this study proposes four types of control inputs that are commensurate with the maintenance categories, denoted as $u_p(t)$, $u_c(t)$, $u_0(t)$, and $u_d(t)$ as shown in Fig. 3 (BSI, 1984). Apparently, one may also take other classifications for the maintenance actions, which may result in more than 4 types of control inputs but also applicable for the development of an optimal maintenance plan. In addition, other possible control inputs can also be identified rather than maintenance in the building energy retrofit practice.

In Fig. 3, there are unplanned maintenance and planned maintenance. The unplanned maintenance is denoted by $u_d(t)$ that refers to the emergency maintenance action, which usually has to be carried out as an unplanned event after the failure. As a result, the maintenance time schedule are ignored for unplanned maintenance, and it might be beneficial to simply take $u_d(t)$ as input disturbances. Under the planned maintenance category, there are corrective maintenance and preventive maintenance. The planned corrective maintenance (CM) is only conducted after the occurrence of a failure, in order to restore the system into a specific working condition. The CM is denoted by $u_c(t)$. The planned CM is deferrable should the failure not affect the whole production process, i.e., the CM is carried out according to the prescribed time schedule, rather than immediately after the failure. Therefore, the planned corrective maintenance is also called deferred corrective maintenance. Unlike the CM, the preventive maintenance (PM) is carried out before the occurrence of a failure in order to reduce the probability of failure or restore the system from a degraded state to a better working condition. The PM includes both the scheduled maintenance and condition-based maintenance, which are denoted by $u_p(t)$ and $u_0(t)$, respectively. The main difference between $u_p(t)$ and $u_0(t)$ is that $u_0(t)$ must be performed at prescribed time intervals or under pre-set conditions to some fatally important systems, while the $u_p(t)$ can be deferred or scheduled. t specifies the time instant when a maintenance action takes place. Mathematically, a key problem related to the optimal control inputs for the building energy retrofit is to identify the value of $\mathbf{u}(t)$, which is a set of $\{u_p(t), u_c(t), u_0(t), u_d(t)\}$ that tells the intensity, maintenance type and schedule of the required maintenance. In fact, as $u_p(t)$ and $u_c(t)$ are subject to the prescribed maintenance plan, their intensity and schedules can be optimized simultaneously. However, $u_c(t)$ and $u_d(t)$ must be performed either at prescribed time intervals or on occurrence of emergency failures. In the literature, selected research activities to design building OMP by the control system approaches are briefly introduced below.

In studies Wang and Xia (2014; 2015b), optimized corrective maintenance activities are designed of for a broad category of failed EE devices in buildings at pre-decided maintenance time schedule. As illustrated by the case study in Wang and Xia (2015b), maintenance plays an important role to the sustainability of the EE device group population. Comparing the optimal maintenance strategy to the full maintenance strategy, the maintenance cost is reduced up to 30.7% with a loss of 1.5% of the energy savings achieved by applying the optimal maintenance.

As discussed in Section 2.3, the functional conditions of an EE device may range from a number of transition stages from perfect to failure. in Wang et al. (2017), homogeneous group population dynamics and the aggregate performance dynamics under the impacts of multi-state deteriorations and maintenances are formulated as a control system model. In this way, both the corrective maintenance and preventive maintenance are introduced into the OMP problem. Fig. 4 depicts a maintenance plan involving both maintenance actions (Wang et al., 2017). The dashed line (in Blue) indicates the CM actions and the dash-dotted lines (in Red) the PM actions. These maintenance actions are subject to pre-decided maintenance time schedule, and the optimization variables are the CM and PM intensities. In the case study from Wang et al. (2017), when comparing to the maintenance plan without preven-

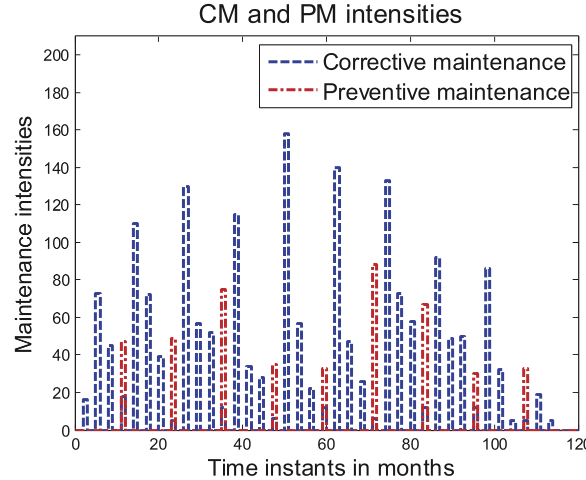


Fig. 4. Optimal maintenance intensities of PM and CM (Wang et al., 2017).

tive maintenance, the optimal maintenance strategy exhibits 5% additional energy savings and 7.5% improvements on the internal rate of return (IRR).

Apart from the maintenance type and intensities, the maintenance time scheduling is another major concern for an OMP problem. Based on the proposed multi-state based performance degradation models in Wang et al. (2017), a maintenance planning taking into account both the maintenance intensities and instants optimization is investigated in Wang, Wu, Zhu, and Xia (2015). Fig. 5 depicts the optimal maintenance intensities under both the scheduled and fixed maintenance intervals according to the case study in Wang et al. (2015). When comparing to the fixed maintenance schedule, the building energy retrofit project can achieve up to 21.7% additional energy savings and 5.7% of improvement on the IRR with the optimal maintenance plan if sufficient budget is provided.

To address the interactions among various building energy systems, the study (Wang & Xia, 2015a) further improves the previous developed maintenance plans. With the considerations of both the energy consumption and reliability interactions between building energy systems, the study (Wang & Xia, 2015a) finds that the optimal maintenance plan is able to provide 8.9% more energy savings and 9.6% improvements on the IRR when comparing to the maintenance activities without considering the interactive effects among building energy systems. The population decay dynamics and the maintenance intensities for the interactive lighting and HVAC systems are provided in Fig. 6.

2.5. Control system outputs

The control system outputs are related to the components y_j and $h_j(\cdot)$ in Eq. (1). For the building energy retrofits, research efforts required for the output y_j may refer to the measurement and sampling of the quantity of survived lamp population, or the M&V of the energy savings, carbon emission reductions, or cost savings of a specific building energy retrofit project. The function $h_j(\cdot)$ can be as simple as a sampling or measurement reading of the state variable $\mathbf{x}$, or a sampling approach, a set of metering instruments to observe y_j, or a performance evaluation process like M&V to determine the energy or cost savings of the building energy retrofit project. In this category, one key research problem is to identify y_j accurately with minimum measurement and sampling efforts, as accurate readings of y_j will contribute to reduce the control system disturbance of d_i and ω_j. More details are elaborated in the next subsection.

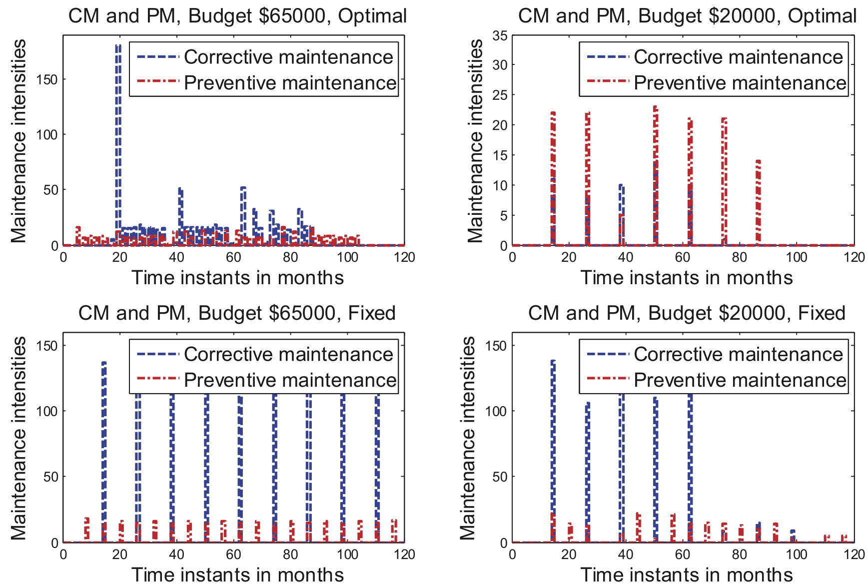


Fig. 5. Optimal maintenance intensities under scheduled and fixed maintenance intervals (Wang et al., 2015).

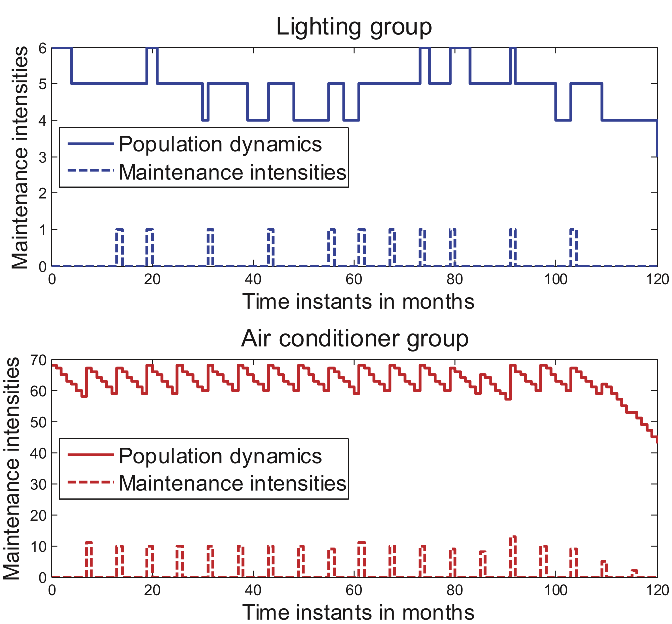


Fig. 6. Optimal maintenance intensities for interactive building energy systems (Wang & Xia, 2015a).

2.6. Control system uncertainties and disturbances

The control system uncertainties and disturbances are denoted by ω_j and d_i respectively in Eq. (1). The ASHRAE guideline (ASHRAE, 2002) introduces that quantifiable uncertainties of energy savings are categorized as modeling uncertainties, measurement uncertainties and sampling uncertainties. Such classifications are also applicable to the control systems as the modeling uncertainties delivers significant impacts to the control system performances. At the current stage, disturbances ω_j and d_i are introduced as simplified interpretations of such impact. These disturbances hereby refer to modeling mismatch of the system dynamics, which are due to the improper mathematical function form, inclusion of the irrelevant variables or exclusion of relevant variables. The measurement and sampling uncertainties are usually observed from the identification of the system outputs. The measurement uncertainties usually come from the inappropriate calibration of the measurement equipment, inexact measurement, or improper meter selection, installation or operation. The sampling uncertainties are resulted from inappropriate sampling approaches or insufficient sample sizes.

In the field of building energy retrofit, uncertainties can deliver further impacts, rather than merely to the control system performances. The relevant research activities focus on cost-effective approach to handle the three independent uncertainties that prevent M&V professionals from precisely evaluating the performances of EE device groups. Existing studies not only address the three uncertainties separately but also in combination. For instance, optimal sampling plans have been designed in Ye and Xia (2016); Ye et al. (2013; 2014) to accurately measure the daily energy consumptions of lighting systems with minimum sample sizes and cost. In Carstens and Xia (2015), the relative contribution of measurement uncertainty to combined measurement and sampling uncertainty is investigated in the context of M&V projects where the whole population is not metered. The study (Olinga, 2015) presents an M&V cost minimization model to handle M&V sampling and modelling uncertainties cost-effectively. The proposed models provide flexibility in designing optimal and easily implementable M&V plans, which either apply more accurate baseline models and fewer sample sizes or less accurate baseline models and greater sample sizes to achieve the same level of M&V accuracy. The research outcomes in the energy field can also be borrowed in the control field, for the purpose of reducing the control system disturbances and uncertainties, which further improves the accuracy and robustness of the control system.

2.7. Objective functions

In the aforementioned OMP problems, the decision maker often takes into account several contradictory considerations that leads to conflicting objectives (Evins, 2013), i.e., the OMP problems are often multi-objective optimizations. The involved objectives usually include maximizing energy savings, minimizing capital costs or maximizing financial paybacks, subject to a series of constraints, e.g., the targeted energy saving limit, budget limit, payback period limit, etc. In order to apply the control system framework in the OMP problems, a weighted sum approach is employed in the aforementioned studies to formulate the objec-

Table 1
Characteristics of retrofitted EE devices.

Retrofits	Quantities	Unit Price ($)	Unit Energy Saving (kWh)	Unit Cost Saving ($)	Preventive Cost ($)	Corrective Cost ($)
15W retrofit CFL	338	14	105.6	11.9	N/A	14
New fan coil units 3	42	380	4320	486.65	N/A	175
New fan coil units 2	0	380	3542.3	397.95	52	N/A
New fan coil units 1	0	380	2651.75	278.35	70	N/A

tive function, where the multi-objective optimization problem is translated into a minimization problem, i.e., a weighted sum of the objectives associated with a non-stationary penalty function. A general form of the objective function formulation is indicated in the following equation:

$$J = -\lambda_1 f_e(\mathbf{x}, \mathbf{u}) - \lambda_2 f_r(\mathbf{x}, \mathbf{u}) + \omega \sum_{n=1}^{k} max(0, P_n), \tag{13}$$

where λ_1 and λ_2 denote the weighting factors. $f_e(\mathbf{x}, \mathbf{u})$ denotes the energy performance indicator, e.g., the overall energy savings during the crediting period. $f_r(\mathbf{x}, \mathbf{u})$ denotes the economic performance indicator, e.g., the net present value (NPV) or internal rate of return (IRR). P_n with $n = 1, 2, \ldots, k$ denotes the penalty functions and ω is a large positive constant that emphasize the effects of the penalty functions. It a constraint is violated, $P_n > 0$. For example, assuming that the targeted energy saving limit, budget limit and payback period limit is involved in an OMP problem. P_n are accordingly defined as following:

$$P_n = \begin{cases} \alpha - ES|_{all}, & n = 1, \\ h|_{all} - \beta, & n = 2, \\ T_p - T', & n = 3 \end{cases} \tag{14}$$

where α denotes the targeted energy saving amount and $ES|_{all}$ the overall energy savings. β denotes the maintenance budget limit and $h|_{all}$ the overall maintenance costs. T_p denotes the actual payback period and T' the payback period limit. According to Eq. (14), $\sum_{n=1}^{k} max(0, P_n) > 0$ if a constraint is violated.

2.8. Control system algorithms

A great advantage to formulate the OMP problem under the control system framework is the applicability of various control system algorithms in finding the optimal controllers with tolerance of a certain level of control system uncertainties and disturbances. For instance, the results obtained by the MPC approach in Ye et al. (2015),Wang and Xia (2015b),Wang et al. (2017) exhibits better economic benefits and energy savings than those obtained by the open loop control approach in response to the added uncertainties in the control system state variables.

Due to different complexity of the control problems, other control system algorithms may also be used to solve the OMP problems. For instance, generic algorithm (GA) has been applied in Malatji et al. (2013) to identify the optimal building energy retrofit proposal, while the differential evolution (DE) approach is used in a series of articles (Wang & Xia, 2015a; Wang et al., 2014) to solve different types of OMP problems under the control system framework. In addition, neighbourhood field optimization (NFO) algorithm is adopted in Wang and Xia (2015b) to solve an multi-objective building energy retrofit and maintenance planning problem.

2.9. Grouping and modeling

Ideally, working status of each EE device over the crediting period would be continuously monitored to enable an opportunity of an immediate replacement on occurrence of a device failure. If the failure dynamics of each involved EE devcie can be monitored and observed, then a control system can be formulated based on the failure dynamics of an individual device. In this case, an N-dimensional control system can be developed that has N failure dynamics models for each device and a number of N control system inputs to record the replacement actions of the EE devices. In addition, the failure dynamics of each EE device involved in the project must also be observed to ensure the operation of the control system. The N-dimensional control system accurately reflects the device population dynamics since all the N units of the EE devices are continuously monitored. Consequently, the optimal maintenance strategy can also be designed and easily implementable to ensure sustainable project savings. However in practice, it is not feasible to continuously monitor the entire project population over 10 years, especially when the project population is large and decentralized. Worse still, the N dimension control system also brings computational burdens in finding the optimal solution.

In order to reduce the modeling cost and complexity of the unit-based control system, it is proposed develop a group-based control system. For each lighting retrofit project, it is recommended to find homogeneous lighting groups according to the devices' technical specification (i.e., model, make, rated power, life span, etc.), energy consumption patterns, and working conditions. For instance, there may be I types of EE devices involved in one building energy retrofit project, and each type of EE device exhibits the same specifications and energy usage patterns, which results in the same lamp population decay dynamics. Then the N lamps can be classified into I lighting groups, and $I \leq N$. Each of such a group consists of devices that are considered to be homogeneous ones, i.e., with the same inherent energy and reliability performances, the same operating schedules and similar operational environment.

The grouping method raises a new question that how different groupings influence the optimization results. Obviously, grouping is an inherently subjective approach. Different decision makers can have different opinions on how to implement groupings. For example, a collection of lamps can be categorized into two groups according to the geographic information or three groups according to the operating schedules. There are many possible relationships between different groupings. For two different groupings corresponding to the same collection of items, there can be overlap, containment or separation among the categorized groups. The number of groups can also be different. As a result of the common utilization of grouping methods in the aforementioned studies, a question is thereby asked: How will different groupings influence the results of the OMP problems? A preliminary theoretical analysis as to the performance robustness of the grouping method is proposed in Wang and Xia (2016). The concept 'performance robustness' is hereby introduced to facilitate the evaluation of the impacts from applying different grouping. For the OMP problems, performance robustness refers to the ability that the control system output sustains when an alternative grouping is applied. More specifically, given a set of same retrofitted items and two different groups, if the results (performances) of an arbitrary maintenance plan based on one grouping remain accessible when the other grouping is applied, the performance robustness is satisfied, and the two groupings are considered equivalent. The satisfaction

Table 2
Maintenance plan performances with optimal and full maintenance strategies.

Cases	Energy savings (kWh)	Percentage saved	IRR	Payback period (years)	NPV ($)	Maintenance cost ($)	Total investment ($)
Optimal maintenance	1395785.8	133.92%	30.95%	2.57	47724.70483	41984	70676
Full maintenance	1306983.15	125.40%	30.74%	2.58	44152.68	41959	70651

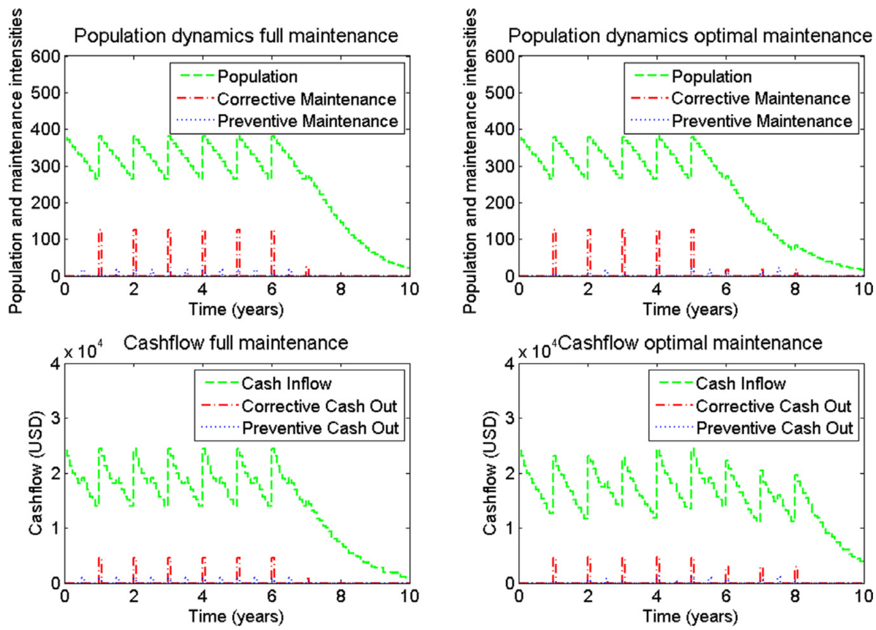


Fig. 7. The population and cash flows from the optimal maintenance and full maintenance strategies.

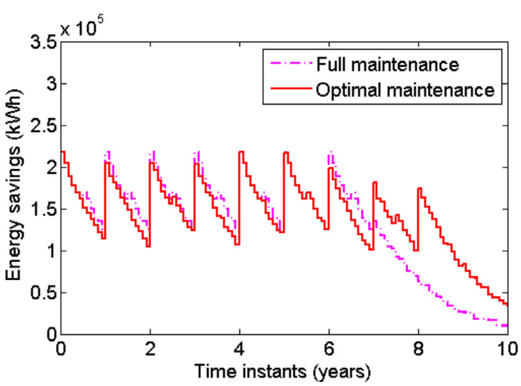


Fig. 8. The timely energy savings over the crediting period.

of the performance robustness provides the decision makers a method to evaluate alternative groupings. In Wang and Xia (2016), a mathematical description of the grouping as well as the grouping based control system formulation is proposed, and a theoretical characterization of grouping robustness is given. Taking advantage of the control system framework of the OMP problems, a distance is defined to evaluate the impacts from applying the grouping method, and a set of alternative groupings can be compared to identify the equivalence between each of them.

Moreover, there might exist an optimal I as the best grouping criterion. More research efforts are expected to find the optimal grouping criterion to minimize the modeling complexity but also ensure the accuracy of the population decay dynamics models.

3. Further discussions

Unlike most of the research articles, this paper formulates the optimal building energy retrofit and maintenance planning problems under the control system framework. Instead of giving more detailed answers to the building energy retrofit planning, this study identifies a number of control system problems that are worthy of future research and investigations. Major contribution of this study is to cast the optimal building maintenance planing problem into a general control system framework. From the general control formulation, the following major research problems in the control systems are discovered, namely

- Control system decay dynamics modeling;
- Control system inputs and model uncertainties;
- Control system outputs;
- Control system uncertainties and disturbances;
- Control system algorithm;
- Grouping and modeling.

The discovered control problems for the building energy retrofit and maintenance planning have been introduced separately in Section 2. However, intrinsic linkages are also observed among these control problems. Given a building energy retrofit project with massive EE devices involved, the grouping criteria to categorize the population into different homogeneous subgroups will influence most of the key factors in the control system formulation. For instance, the homogeneity of each subgroup does influence the system dynamics and will further affect the system state variable selection and system dynamics modeling accuracy. However, quantification of the impacts (i.e., complexity, accuracy) to the control system from different grouping criteria remains an unsolved problem. In addition, the grouping of the project population also decides the measurement and sampling plans to monitoring the projects' energy and financial performance.

3.1. A case study

A case study is given in this section to illustrate the control system framework in practical OMP problems. The case study is selected from a practical building energy retrofit project. There are two groups of retrofitted EE devices. One group consists of a set of compact fluorescent lamps (CFLs) that manifest binary working state. The CFLs are non-repairable items. The other group consists of the air conditioner fan coil units, where three working states are involved. The air conditioners are repairable items. As a result, the multi-state system model introduced in Wang et al. (2017) is employed to characterize the population decay. Due to the space limit, the detailed formulations of the population decay are excluded in this paper and can be found in Wang et al. (2017). Both the planned corrective and preventive maintenance are involved as the control inputs, where the maintenance intensities are the control variables. In this case study, maintenance instants are prescribed by fixed maintenance time schedule.

The specifications and some performance characteristics of the involved retrofitted items are illustrated in Table 1. The new fan coil unit 3, 2 and 1 denote the three working states that correspond with different savings and maintenance costs. The energy saving and cost saving are the annual average value obtained from the energy auditing. The preventive cost indicates the costs of restoring a fan coil unit from working state 2 or 1 to the best working state 3. The corrective cost indicates the costs of restoring one item from failure to normal working.

The crediting period is 10 years. An inspection is performed every month over the crediting period. From the inspection, the The targeted energy saving is 1,042,237.44 kWh. The initial cost is \$28,692. The discount rate for NPV calculation is 11% per year, and the payback period limit is 3 years. The employed budget limit in this case study is \$42,000, which is insufficient for full maintenance strategy. There are 19 preventive maintenance instants and 9 corrective maintenance instants, and the fixed preventive and corrective time schedules $Q_p = \{0.5, 1, 1.5, 2, \ldots, 9.5\}$ and $Q_c = \{1, 2, 3, \ldots, 9\}$. The unit of the maintenance instants is year. According to the time schedule, the maintenance instants are evenly distributed over the sustainability period. The weighted sum of two objectives: overall energy savings and IRR is employed to be the objective function. The adopted weight factors are $\lambda_1 = 0.5$ and $\lambda_2 = 0.5$, implying that the two objectives are equally considered. More detailed formulations can be found in Wang et al. (2017).

A model predictive control (MPC) controller is designed to solve the OMP problem in the case study. A DE algorithm based numerical solver is employed for the MPC controller. The solutions are illustrated in Table 2. The full maintenance strategy is the comparative baseline. In this case study, the full maintenance will restore all degraded devices to the normal working states until the all budget is consumed. The following performances are selected in Table 2: the overall energy savings during the crediting period (given in kWh), the percentage savings that indicate the ratios against the targeted energy savings, the IRR, the payback period (given in years), the NPV, the total maintenance cost and the overall investment (given in USD). According to Table 2, the optimal maintenance achieves much higher energy savings than the full maintenance strategy. The economic performances from optimal maintenance are also better. This implies that the full maintenance strategy cannot make the best use of the budget. The optimal maintenance provides further opportunities to achieve energy efficiency and cost effectiveness to decision makers. Fig. 7 depicts the population dynamics and cash flows over the crediting period from the two maintenance strategies. Fig. 8 depicts their timely energy savings. Generally, the difference between the two strategies is resulted from their maintenance actions with the air conditioners fan coil units. Due to their high savings and high maintenance costs, the optimal strategy devotes more budget to maintain the working state of the fan coil units. The optimal maintenance strategy appears to be 'smarter' to figure out more efficient options.

4. Conclusions

The ongoing and near future research in the building energy retrofit and maintenance planning by the control system approach are planned as follows: 1) to improved the modeling accuracy of the population decay and energy performance degradation dynamics of various EE devices; 2) to expand the existing population decay and/or energy performance degradation dynamics models for different types of EE devices in the same boundary whose energy usage pattern are interactive and coupling with each other, where decoupling control approaches must be used for the optimal maintenance planning; 3) to investigate the existence of an optimal grouping criterion, which results in minimal dimensions of the control system state variables and maximum control system performance; 4) to design optimal maintenance plans for building envelope retrofits; 5) to develop a software platform that designs optimal building retrofit and maintenance plans for different types of building blocks, for the purpose of maximizing energy savings and minimizing initial investment and payback periods.

References

Abo-Al-Ez, K. M., Elaiw, A., & Xia, X. (2014). A dual-loop model predictive voltage control/sliding-mode current control for voltage source inverter operation in smart microgrids. *Electric Power Components and Systems, 42*(3–4), 348–360.

Ahn, B.-L., Jang, C.-Y., Leigh, S.-B., Yoo, S., & Jeong, H. (2014). Effect of led lighting on the cooling and heating loads in office buildings. *Applied Energy, 113*, 1484–1489.

Arens, E., Federspiel, C., Wang, D., & Huizenga, C. (2005). How ambient intelligence will improve habitability and energy efficiency in buildings. In *Ambient intelligence* (pp. 63–80). Springer.

ASHRAE (2002). ASHRAE Guideline 14: Measurement of energy and demand savings. *Technical report*.

Bertoldi, P., & Rezessy, S. (2008). Tradable white certificate schemes: Fundamental concepts. *Energy Efficiency, 1*, 237–255.

Bijker, A. J., Xia, X., & Zhang, J. (2009). *Active power residential non-intrusive appliance load monitoring system*. IEEE AFRICON 2009, Nairobi, Kenya, 23–25 September.

Boukas, E., & Liu, Z. (2001). Production and maintenance control for manufacturing systems. *IEEE Transactions on Automatic Control, 46*(9), 1455–1460.

BSI (1984). Glossary of maintenance management terms in terotechnology. *Technical Report*. British Standards Institution.

Carstens, H., & Xia, X. (2015). Measurement uncertainty and risk in measurement and verification projects. *Iepec 2015*. Long Beach, California.

Carstens, H., Xia, X., & Ye, X. (2014). Improvements to longitudinal clean development mechanism sampling designs for lighting retrofit projects. *Applied Energy, 126*, 256–265.

Carstens, H., Xia, X., Zhang, J., & Ye, X. (2013). *Characterising compact fluorescent lamp population decay*. IEEE AFRICON 2013, Mauritius, 09–12 September.

Catherine, Q., Wheeler, J., Wilkinson, R., & de Jager, G. (2012). Hot water usage profiling to improve geyser efficiency. *Journal of Energy in Southern Africa, 23*(1), 39–45.

Elaiw, A., Xia, X., & Shehata, A. (2012). Application of model predictive control to optimal dynamic dispatch of generation with emission limitations. *Electric Power Systems Research, 84*(1), 31–44.

Elaiw, A., Xia, X., & Shehata, A. (2013). Hybrid de-sqp and hybrid pso-sqp methods for solving dynamic economic emission dispatch problem with valve-point effects. *Electric Power Systems Research, 103*, 192–200.

Eskom (2011). The measurement and verification guideline for demand-side management projects. *Technical report*.

Esmaeli, A. (2016). Stability analysis and control of microgrids by sliding mode control. *International Journal of Electrical Power & Energy Systems, 78*, 22–28.

Evins, R. (2013). A review of computational optimisation methods applied to sustainable building design. *Renewable and Sustainable Energy Reviews, 22*, 230–245.

Fan, Y., & Xia, X. (2015). A multi-objective optimization model for building envelope retrofit planning. *Energy Procedia, 75*, 1299–1304.

Fan, Y., & Xia, X. (2017). A multi-objective optimization model for energy-efficiency building envelope retrofitting plan with rooftop pv system installation and maintenance. *Applied Energy, 189*, 327–335.

Liu, Y., Zhang, Q., Wang, C., & Wang, N. (2014). A control strategy for microgrid inverters based on adaptive three-order sliding mode and optimized droop controls. *Electric Power Systems Research, 117*, 192–201.
Malatji, E. M., Zhang, J., & Xia, X. (2013). A multiple objective optimisation model for building energy efficiency investment decision. *Energy and Buildings, 61*, 81–87.
Mei, J., Zhu, B., & Xia, X. (2015a). *Model predictive control for optimizing indoor air temperature and humidity in a direct expansion air conditioning system*. The 27th Chinese Control and Decision Conference (CCDC 2015), Qingdao, China, 23–25 May.
Mei, J., Zhu, B., & Xia, X. (2015b). *Predictive control for thermal comfort and energy efficiency in a direct expansion air conditioning system*. 2015 Chinese Automation Congress (CAC 2015), Wuhan, China, 25–27 December.
Michaelowa, A., & Jotzo, F. (2005). Transaction costs, institutional rigidities and the size of the clean development mechanism. *Energy Policy, 33*, 511–523.
Mokgonyana, L., Zhang, J., Zhang, L., & Xia, X. (2016). Coordinated two-stage volt/var management in distribution networks. *Electric Power Systems Research, 141*, 157–164.
Mozzo, M. A. (1999). Measurement and verification of savings in performance contracting. *Energy Engineering, 96*(2), 33–45.
Mundaca, L. (2007). Transaction costs of tradable white certificate schemes: The energy efficiency commitment as case study. *Energy Policy, 35*, 4340–4354.
Navigant (1999). Evaluation of the IFC/GEF Poland efficient lighting project CFL subsidy program. *Final report, Edition 2*. Philadephia: Navigant Consulting, PA, USA.
Nikkhajoei, H., & Lasseter, R. H. (2009). Distributed generation interface to the certs microgrid. *IEEE Transactions on Power Delivery, 24*(3), 1598–1608.
Ntsaluba, S., Zhu, B., & Xia, X. (2016). Optimal flow control of a forced circulation solar water heating system with energy storage units and connecting pipes. *Renewable Energy, 89*, 108–124.
Nwulu, N. I., & Xia, X. (2015a). A combined dynamic economic emission dispatch and time of use demand response mathematical modelling framework. *Journal of Renewable and Sustainable Energy, 7*(4), 043134.
Nwulu, N. I., & Xia, X. (2015b). Implementing a model predictive control strategy on the dynamic economic emission dispatch problem with game theory based demand response programs. *Energy, 91*, 404–419.
Nwulu, N. I., & Xia, X. (2015c). Multi-objective dynamic economic emission dispatch of electric power generation integrated with game theory based demand response programs. *Energy Conversion and Management, 89*, 963–974.
Nwulu, N. I., & Xia, X. (2017). Optimal dispatch for a microgrid incorporating renewables and demand response. *Renewable Energy, 101*, 16–28.
O'Connor, P., & Kleyner, A. (2011). *Practical reliability engineering*. Wiley.
Olinga, Z. (2015). A cost-effective approach to handle measurement and verification sampling and modelling uncertainties. *MEng Thesis, University of Pretoria, South Africa*.
Portmess, L., & Tower, S. (2015). Data barns, ambient intelligence and cloud computing: the tacit epistemology and linguistic representation of big data. *Ethics and Information Technology, 17*(1), 1–9.
Sao, C. K., & Lehn, P. W. (2005). Autonomous load sharing of voltage source converters. *IEEE Transactions on Power Delivery, 20*(2), 1009–1016.
Setlhaolo, D., & Xia, X. (2015). Optimal scheduling of household appliances with a battery storage system and coordination. *Energy and Buildings, 94*, 61–70.
Setlhaolo, D., & Xia, X. (2016). Combined residential demand side management strategies with coordination and economic analysis. *International Journal of Electrical Power & Energy Systems, 79*, 150–160.
Setlhaolo, D., Xia, X., & Zhang, J. (2014). Optimal scheduling of household appliances for demand response. *Electric Power Systems Research, 116*, 24–28.
Sichilalu, S., Tazvinga, H., & Xia, X. (2016). Optimal control of a fuel cell/wind/pv/grid hybrid system with thermal heat pump load. *Solar Energy, 135*, 59–69.
Sichilalu, S. M., & Xia, X. (2015a). Optimal energy control of grid tied pv–diesel—battery hybrid system powering heat pump water heater. *Solar Energy, 115*, 243–254.
Sichilalu, S. M., & Xia, X. (2015b). Optimal power dispatch of a grid tied-battery-photovoltaic system supplying heat pump water heaters. *Energy Conversion and Management, 102*, 81–91.
Stavropoulos, T. G., Kontopoulos, E., Bassiliades, N., Argyriou, J., Bikakis, A., Vrakas, D., & Vlahavas, I. (2015). Rule-based approaches for energy savings in an ambient intelligence environment. *Pervasive and Mobile Computing, 19*, 1–23.
Stavropoulos, T. G., Koutitas, G., Vrakas, D., Kontopoulos, E., & Vlahavas, I. (2016). A smart university platform for building energy monitoring and savings. *Journal of Ambient Intelligence and Smart Environments, 8*(3), 301–323.
Tazvinga, H., Xia, X., & Zhang, J. (2013). Minimum cost solution of photovoltaic–diesel–battery hybrid power systems for remote consumers. *Solar Energy, 96*, 292–299.
Tazvinga, H., Zhu, B., & Xia, X. (2014). Energy dispatch strategy for a photovoltaic–wind–diesel–battery hybrid power system. *Solar Energy, 108*, 412–420.
Tazvinga, H., Zhu, B., & Xia, X. (2015). Optimal power flow management for distributed energy resources with batteries. *Energy Conversion and Management, 102*, 104–110.
UNFCCC (2007). Approved baseline and monitoring methodology AM0046, distribution of efficient light bulbs to household. *Version 02, Technical Report*.
UNFCCC (2010). Approved small scale methodology AMS II.J, demand-side activities for efficient lighting technologies. *Version 04, Technical Report*.
UNFCCC (2011). Approved small scale methodology AMS II.L, demand-side activities for efficient outdoor and street lighting technologies. *Version 01, Technical Report*.
UNFCCC (2012). Approved small scale methodology AMS II.N, demand-side energy efficiency activities for installation of energy efficient lighting and/or controls in buildings. *Version 01.0, Technical Report*.
USDOE (2011). Advanced energy retrofit guides: Office buildings. *Technical Report*. Pacific Northwest National Laboratory and PEC with U.S. Department of Energy.
Wang, B., Wu, Z., & Xia, X. (2017). A multi-state based control system approach towards the optimal maintenance for building energy efficiency retrofits. *IEEE Transactions on Control System Technology, 25*(1), 374–381.
Wang, B., Wu, Z., Zhu, B., & Xia, X. (2015). *Optimal control of maintenance instants and intensities in building energy efficiency retrofitting project*. The 54th IEEE Conference on Decision and Control (CDC 2015), Osaka, Japan, 15–18 December.
Wang, B., & Xia, X. (2014). *A control system approach to corrective maintenance planning of building retrofitted facilities*. The 19th World Congress of the International Federation of Automatic Control (IFAC 2014), Cape Town, South Africa, 24–29 August.
Wang, B., & Xia, X. (2015a). *Maintenance plan optimization in building retrofitting with interacting energy efficiency effects*. 2015 Chinese Automation Congress (CAC 2015), Wuhan, China, 25–27 December.
Wang, B., & Xia, X. (2015b). Optimal maintenance planning for building energy efficiency retrofitting from optimization and control system perspectives. *Energy and Buildings, 96*, 299–308.
Wang, B., & Xia, X. (2016). *A preliminary study on the robustness of grouping based maintenance plan optimization in building retrofitting*. 2016 International Conference on Applied Energy (ICAE 2016), Beijing, China, 08–11 October.
Wang, B., Xia, X., & Zhang, J. (2014). A multi-objective optimization model for the life-cycle cost analysis and retrofitting planning of buildings. *Energy and Buildings, 77*, 227–235.
Wang, N., Zhang, J., & Xia, X. (2013). Energy consumption of air conditioners at different temperature set points. *Energy and Buildings, 65*, 412–418.
Wilson, D. G., Robinett, R. D., Weaver, W. W., Byrne, R. H., & Young, J. (2016). Nonlinear power flow control design of high penetration renewable sources for ac inverter based microgrids. In *2016 international symposium on power electronics, electrical drives, automation and motion (speedam)* (pp. 701–708). IEEE.
Wu, Z., Tazvinga, H., & Xia, X. (2015a). Demand side management of photovoltaic–battery hybrid system. *Applied Energy, 148*, 294–304.
Wu, Z., Wang, B., & Xia, X. (2016). Large-scale building energy efficiency retrofit: concept, model and control. *Energy, 109*, 456–465.
Wu, Z., Xia, X., & Wang, B. (2015b). Improving building energy efficiency by multiobjective neighborhood field optimization. *Energy and Buildings, 87*, 45–56.
Xia, X., & Elaiw, A. (2010). Optimal dynamic economic dispatch of generation: A review. *Electric Power Systems Research, 80*(8), 975–986.
Xia, X., & Zhang, J. (2010). Energy efficiency and control systems–from a POET perspective. *1st IFAC conference on Control Methodologies and Technology for Energy Efficiency, IFAC Proceedings Volumes, 43*(1), 255–260.
Xia, X., & Zhang, J. (2011). Modeling and control of heavy-haul trains [applications of control]. *IEEE Control Systems, 31*(4), 18–31.
Xia, X., & Zhang, J. (2013). Mathematical description for the measurement and verification of energy efficiency improvement. *Applied Energy, 111*, 247–256.
Xia, X., & Zhang, J. (2015). Operation efficiency optimisation modelling and application of model predictive control. *IEEE/CAA Journal of Automatica Sinica, 2*(2), 166–172.
Xia, X., Zhang, J., & Cass, W. (2012). Energy management of commercial buildings-a case study from a POET perspective of energy efficiency. *Journal of Energy in Southern Africa, 23*(1), 23–31.
Xia, X., Zhang, J., & Elaiw, A. (2009). *A model predictive control approach to dynamic economic dispatch problem* (pp. 1–7). PowerTech, 2009 IEEE, Bucharest, Romania, 28 June - 2 July.
Xia, X., Zhang, J., & Elaiw, A. (2011). An application of model predictive control to the dynamic economic dispatch of power generation. *Control Engineering Practice, 19*(6), 638–648.
Xia, X., & Zhang, L. (2016). Industrial energy systems in view of energy efficiency and operation control. *Annual Reviews in Control, 42*, 299–308.
Ye, X., & Xia, X. (2014). Optimal sampling and maintenance plans for the lighting retrofit projects towards sustainable energy savings. *Energy Procedia, 61*, 648–651.
Ye, X., & Xia, X. (2016). Optimal metering plan for measurement and verification on a lighting case study. *Energy, 95*, 580–592.
Ye, X., Xia, X., & Zhang, J. (2013). Optimal sampling plan for clean development mechanism energy efficiency lighting projects. *Applied Energy, 112*, 1006–1015.
Ye, X., Xia, X., & Zhang, J. (2014). Optimal sampling plan for clean development mechanism lighting projects with lamp population decay. *Applied Energy, 136*, 1184–1192.
Ye, X., Xia, X., Zhang, L., & Zhu, B. (2015). Optimal maintenance planning for sustainable energy efficiency lighting retrofit projects by a control system approach. *Control Engineering Practice, 37*, 1–10.
Yu, X., Khambadkone, A. M., Wang, H., & Terence, S. T. S. (2010). Control of parallel–connected power converters for low-voltage microgrid - part i: A hybrid control architecture. *IEEE Transactions on Power Electronics, 25*(12), 2962–2970.
Zhang, J., & Xia, X. (2011). A model predictive control approach to the periodic implementation of the solutions of the optimal dynamic resource allocation problem. *Automatica, 47*(2), 358–362.
Zhu, B., Tazvinga, H., & Xia, X. (2015). Switched model predictive control for energy dispatching of a photovoltaic-diesel-battery hybrid power system. *IEEE Transactions on Control Systems Technology, 23*(3), 1229–1236.

Journal of Cleaner Production 170 (2018) 1151–1166

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

Sustainable energy-water management for residential houses with optimal integrated grey and rain water recycling

Evan Wanjiru*, Xiaohua Xia

Centre of New Energy Systems, Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa

ARTICLE INFO

Article history:
Received 3 July 2017
Received in revised form
17 August 2017
Accepted 23 September 2017
Available online 25 September 2017

Keywords:
Demand side management
Energy-water nexus
Grey water
Life cycle cost analysis
Optimal control
Rain water harvesting

ABSTRACT

South Africa is a semi-arid developing country facing water and energy insecurity. There are colossal challenges in reliably providing these resources amid growing population, increased urbanization and improved living standards causing increased demand for these resources. Development of new supply centralized systems comes at an exorbitant cost, whereas decentralized systems are touted as an attractive alternative. Grey water recycling and rain water harvesting at buildings level is such an alternative that can provide water for non-potable uses. However, there are technological challenges of optimally operating such systems while ensuring efficient use of associated energy. This paper introduces two control strategies; open loop optimal control and closed-loop model predictive control (MPC) strategies aimed at ensuring safe and reliable operation of the grey water recycling and rain water harvesting system while efficiently using associated energy. From the case study, the proposed system with either control strategy can save the cost of water and waste water by up to 32.3% and 29.5% respectively, while leading to 35.7% in energy cost savings and 31.5% in total operational cost savings in a month. Adoption of these systems would have a huge environmental effect in reducing demand for sewerage services, conservation of water hence reducing demand for potable water as well as increasing the energy efficiency. Furthermore, the system would increase the reliability and security of water supply. Despite the benefits, the system does not pay within its lifetime and therefore, government intervention is required so as to make it economically attractive. High cost of implementation coupled with low potable and waste water tariffs harbour adoption of these systems. Appropriate regulations, policies, incentives and public education are necessary to support such novel technologies in ensuring resource conservation, efficiency and security are achieved.

1. Introduction

Most developing nations are struggling to provide water and energy, two resources that are greatly connected, hence the name energy-water nexus Wanjiru et al. (2017a). In South Africa, various factors such as economic growth, improved standards of living, population increase, rural-urban migration, frequent droughts, greater connectivity, insufficient and seasonal rainfall have increased the insecurity of the two resources Odhiambo (2009); Cobbinah et al. (2015). The country suffered from serious power shortages in 2008 that necessitated authorities to come up with various initiatives to improve the situation. According to Xia and Zhang (2016), supply side initiatives such as building of new coal power plants were not only harmful to the environment but they also came at exorbitant cost. On the other hand, demand side management (DSM) initiatives that bridge the gap between supply and demand were more desirable, encironmentally friendly and cheaper to implement. Consequently, energy DSM research across various sectors has successfully been carried out. In commercial and industrial sectors, energy DSM initiatives include energy efficiency in coal plants through optimal sizing and operation of pumps studied by Zhang et al. (2014), while Badenhorst et al. (2011) looked at scheduling of deep mine rock winders, Numbi and Xia (2016) analysed optimal energy control of coal crushers and Chatterjee et al. (2015) looked into optimally operating of ventilation systems with the aim of managing energy consumption. Others like Mathaba and Xia (2015) explored energy management of conveyor belts, water pumping stations Zhang et al. (2012), including those with multiple pumps Zhuan and Xia (2013). Optimal dynamic power dispatch by utilities has been studied by

* Corresponding author.
E-mail address: murimev@gmail.com (E. Wanjiru).

https://doi.org/10.1016/j.jclepro.2017.09.212
0959-6526/

Nomenclature

A_t^1, A_t^2, A_t^3	Cross-sectional area of potable, grey and holding water tank (m^2) respectively
D_{grey}, D_{pot}	Potable and grey water demand (m^3) respectively
h_1, h_2, h_3	Height of water in potable, grey and holding tank (m) respectively
p_e	Price of electricity using TOU tariff (currency/kWh)
P_1^m, P_3^m	Potable and grey water pump's motor rating (kW) respectively
s_1, s_3	Auxiliary variable for potable and grey water pump respectively
S_{grey}, S_{rain}	Collected untreated grey and rain water respectively (m^3)
TOU	Time-of-use tariff
t_s and j	Sampling period (h) and j^{th} sampling interval
u_1, u_2	Potable water pump's switch and valve respectively
u_3, u_4	Grey water pump's switch and drainage valve respectively
Q_1, Q_3	Flow rate of water across potable and grey water pump (m^3/h) respectively
Q_2, Q_4	Flow rate of water across potable and drainage valve (m^3/h) respectively
V_1, V_2, V_3	Volume of water in respective tanks (m^3)
Rand (R)	South African currency (1 Rand = 0.076 USD, as at 17 Aug. 2017)

Elaiw et al. (2012) with the aim of limiting emissions, while Chou and Xia (2007) investigated on efficient and optimal operation of heavy haul trains. Application of renewable energy as a suitable alternative source of clean energy has also been examined. A minimum cost solution model for photovoltaic-diesel-battery power systems was designed by Tazvinga et al. (2013) to provide optimal power flow for remote customers, and later extended to cater for distributed energy systems Tazvinga et al. (2015). Further, Tazvinga et al. (2014) have proposed an energy dispatch strategy for photovoltaic-wind-diesel-battery power system. In residential buildings, Malatji et al. (2013) examined an optimal energy-efficient building retrofitting decision model whereas Wang et al. (2014) considered life cycle cost analysis and retrofitting planning in their model and Fan and Xia (2017) developed an optimization model for building envelope retrofitting planning. Setlhaolo and Xia (2015) have delved in demand response research involving optimal scheduling of household appliances. Energy management in hot water supply has been studied by Ntsaluba et al. (2016) who looked into optimal operation of a solar water heating systems, whilst Sichilalu et al. (2017) examined suitability of heat pump water heaters powered by remote hybrid energy systems as well as grid-tied systems Sichilalu and Xia (2015).

South Africa has a constant water stress of about 40%–60% resulting from low amount of rainfall averaging at about 500 *mm* per annum Roy and Rouault (2013), and high evaporation of about 1700 *mm* per annum Adewumi et al. (2010). Just like energy, water management can be carried out on supply or demand side. With extraction of fresh water sources nearing yield point, development of new centralized water supply and sewage disposal systems comes at astronomical costs, negative environmental impact and is also incompatible with modern complex requirements, especially in developing cities. The solution has been in development of cheaper decentralized systems that minimize the amount of pollution generated and discharged through using and reusing water very near to the point of origin. Decentralized systems, as pointed out by Wanjiru and Xia (2017), have further been supported by technological advancement in water treatment technologies, change of attitude and awareness on the importance of water conservation. Water insecurity is forcing municipalities to explore alternative means of supplying and managing water consumption. They include water restrictions, pressure management Hoy (2009), monitoring water consumption pattern, management of meters, installation and retrofitting with efficient devices, planting of water efficient vegetation Adewumi et al. (2010), communication and education Bohensky (2006), and promotion of waste water reuse Lam et al. (2015). Evidently, as Brandoni and Bošnjaković (2017) elaborate, most DSM initiatives have exclusively focused on either energy or water despite them being heavily intertwined both in production and consumption levels. In fact, decisions made on one resource have been shown to have undesired effect on the other since decision makers have little understanding on scientific or policy complexities of one resource on the other Stillwell et al. (2010). It is therefore important to consider energy-water nexus DSM in order to bridge the gap between supply and demand through conservation and efficient utilisation of both resources. This nexus has been appreciated well in developed countries, while developing countries are still catching up, with South Africa leading in the research in Africa. A number of energy-water DSM research in residential buildings in the country have been carried out. Wanjiru et al. (2016a) developed an open-loop optimal controller to operate heat pump water heaters and instantaneous heaters powered using integrated renewable energy systems for off-grid and later expanded for grid tied application in Wanjiru et al. (2017c), as well as control using closed-loop model predictive control (MPC) in Wanjiru et al. (2017b). The studies sought maximize the use of renewable energy while minimizing the amount of cold water wasted at end uses. In another study, Wanjiru and Xia (2015) explored harvesting rooftop rain water and optimizing the operation of autonomous lawn irrigation. For buildings with intermittent or insufficient water supply that rely on pump-storage systems, Wanjiru et al. (2016b) examined the optimal operation of the system using both open-loop optimal and MPC controllers. Further, the two controllers have been designed for grey water recycling system by Wanjiru and Xia (2017) suitable for urban residential buildings with the aim of minimizing the amount of pumping energy used as well as ensure conservation of potable water takes place.

Grey water is the waste water generated from households that has no faecal matter, for instance, water from washing machines, shower, baths and dishwasher. This water is easier to treat as it contains fewer pathogens, to be reused on-site for non-potable uses such as lawn irrigation and toilet flushing. Domestic grey water recycling and rain water harvesting systems not only have environmental benefits but also economic benefits to end users and the country at large. Most research has mainly focused on separate rain water harvesting systems. For instance, Kahinda and Taigbenu (2011) highlight the challenges and opportunities of rain water harvesting in South Africa while Morales-Pinzón et al. (2015) looked at their economic and environmental benefits. Al-Jayyousi (2003) reviewed the suitability of grey water recycling systems to achieve sustainable water management. However, Kim et al. (2007) argued that combined grey and rain water recycling systems could have more benefits than exclusive water recycling systems, though, the analysis done by Stec et al. (2017) shows that financial efficiency on water and energy savings would vary with the size of the system and region. Ghisi and de Oliveira (2007) looked into the potential of grey water recycling and rain water harvesting system in Brazil. First, the performance and payback period of the grey water and rain water harvesting systems was analysed separately, and later, the combined grey and rain water system's performance was analysed. It was found out that though the three systems have huge potential in conserving water, they had a high payback period of more than 20 years. Another study in Beijing by Zhang et al. (2009) revealed that grey water recycling is more suitable in the area than rain water harvesting due to severe pollution in the city that could reduce the quality of rain water. A localized water recycling and rain water harvesting scheme was designed by Rozos et al. (2009) and the analysis showed that although both rain water harvesting and grey water recycling lead to water conservation, schemes based on recycling grey water are less susceptible to climatic changes, while those based on rain water harvesting are more susceptible to changes. Therefore, a combined grey water recycling and rain water harvesting is suitable for a country like South Africa with low rainfall, in order to achieve maximum benefits and reliability. Three land uses, that is, single-family house, apartment cluster and mixed use site, were analysed for viability of use of grey water recycling and rain water harvesting systems as an alternative source of water supply for non-potable uses. Though water is conserved in all the three uses, the largest impediment to their adoption is cost. In fact, Loux et al. (2012) assert that the cost is much higher for single-family house, while the discrepancy between cost and savings levels out at higher densities. Despite the important benefits that such systems have on the environment and utilities, there are still technological challenges on their optimal operation so as to ensure both water and energy efficiency are achieved. In existing systems, some operations have to be carried out manually by home owners. Solutions to these challenges must be provided in consideration of energy-water nexus so as to maximize the combined results from the system.

Previous research has focused on designing grey and rain water systems, with little attention being given to optimal, reliable and autonomous operation of such systems. Various control strategies can be used to operate as well as minimize the system's dependence on human operator. Classical control techniques such as proportional-integral-derivative (PID) controllers use a trial and error process where various methods are iteratively used to determine design variables of an acceptable system. PID controllers, according to Kozák (2016), are suitable for single-input-single-output (SISO) applications and have low accuracy in non-linear processes, or those with a large time delay. Constrained direct inversion (CDI) controllers are superior to conventional control strategies with feedback loop as they can handle disturbances, constraints, dead time and non-linearity. Optimal control, whether open-loop or closed-loop model predictive control (MPC), seeks to determine the control actions that will operate a process within physical and operational constraints while simultaneously minimizing or maximizing a given performance criteria online Kirk (2012). Though CDI uses an analytical approach while optimal control uses numeric optimization, they at times offer similar results. This makes CDI to be more computationally efficient than MPC hence suitable for simple control systems. Optimal controllers, on the other hand, perform better in reaching optimality as CDI's rule is not designed to be optimal Tóth et al. (2012). Optimal controllers can effectively, robustly and accurately handle multivariable dynamic systems through minimizing the cost function Mayne et al. (2000). The aim of this research, therefore, is to develop suitable optimal controllers that can reliably, autonomously and robustly control the water recycling system while considering energy-water nexus to maximize the benefits.

This paper introduces the first attempt to design novel, economical and advanced optimal controllers to operate the grey and rain water recycling system for residential areas. Open-loop optimal control and closed-loop MPC systems are designed to meet hourly potable and non-potable water demand for a house leading to water conservation and energy efficiency. The control strategies are designed to cater for different application requirements. Although open-loop control is more cost effective and easy to implement, it is suitable where water demand is known to be relatively uniform. However, in cases where it is difficult to accurately predict water demand and the system is susceptible to external disturbances that significantly affect the demand pattern, the closed-loop MPC should be adopted. It, however, requires installation of additional monitoring devices to the system such as level monitoring of the tanks thereby increasing the cost and complexity of the control system. The proposed system, if widely adopted, would reduce the demand for potable water, energy and sewage services from the utilities and municipalities, leading to lower cost of potable and waste water which corresponds to lower bills paid by the end user associated to both resources. However, it is important for subsidies and rebates to be offered by the government to lower the cost of implementing such systems in individual houses.

This paper is outlined as follows: Section 2 shows how the model for the proposed grey water recycling and rain water harvesting is developed. Section 3 discusses the design of the controllers, Section 4 provides information about the case used to test the control systems and other information necessary for implementing the proposed strategy. Section 5 discusses the results while Section 6 gives the conclusion and recommendations.

2. System development

2.1. Schematic layout

A typical grey and rain water water recycling system for a stand alone house is shown in Fig. 1.

Two scenarios motivated by the water situation in South Africa are considered. Firstly, the house is considered to have reliable municipal water supply such that pumping and storage is not required as water flows to various end uses in the house and through valve u_2 when necessary. Secondly, water supply is unreliable, either because of low water pressure or water rationing taking place in the area. Therefore, water pumping and storage is necessary to improve the reliability and convenience of potable water supply to the house occupants. A fixed speed potable water pump whose state is represented as u_1 pumps water to the potable storage water tank from where it flows by gravity to various end uses in the house. Some end uses such as shower and washing machine produce grey water that can easily be treated for non-potable end uses such as toilet and garden irrigation. Further, rain water can also be harvested from the roof and used for the same non-potable end uses. Recycling grey water and harvesting rain water would lead to water conservation as well as reducing demand for potable water and sewage services in both scenarios. Both grey and rain water pass through a filter to remove particles and then flows to the holding tank for temporary storage. This tank must be emptied, through the drainage valve represented as u_4, every 24-h to prevent formation of bacteria responsible for producing foul smell. Collected grey water is then pumped through an ultraviolet (UV) water purifier and stored in a rooftop grey water tank, from where it flows by gravity to non-potable end uses. In instances where grey water tank has insufficient treated water, potable water is allowed to flow through potable water valve u_2 to assist in meeting the demand. Finally, black water, which cannot easily be recycled, is allowed to flow to the drainage. Therefore, the aim is to control pumps and valves to ensure convenient and reliable water supply that ensures both energy-water efficiency and

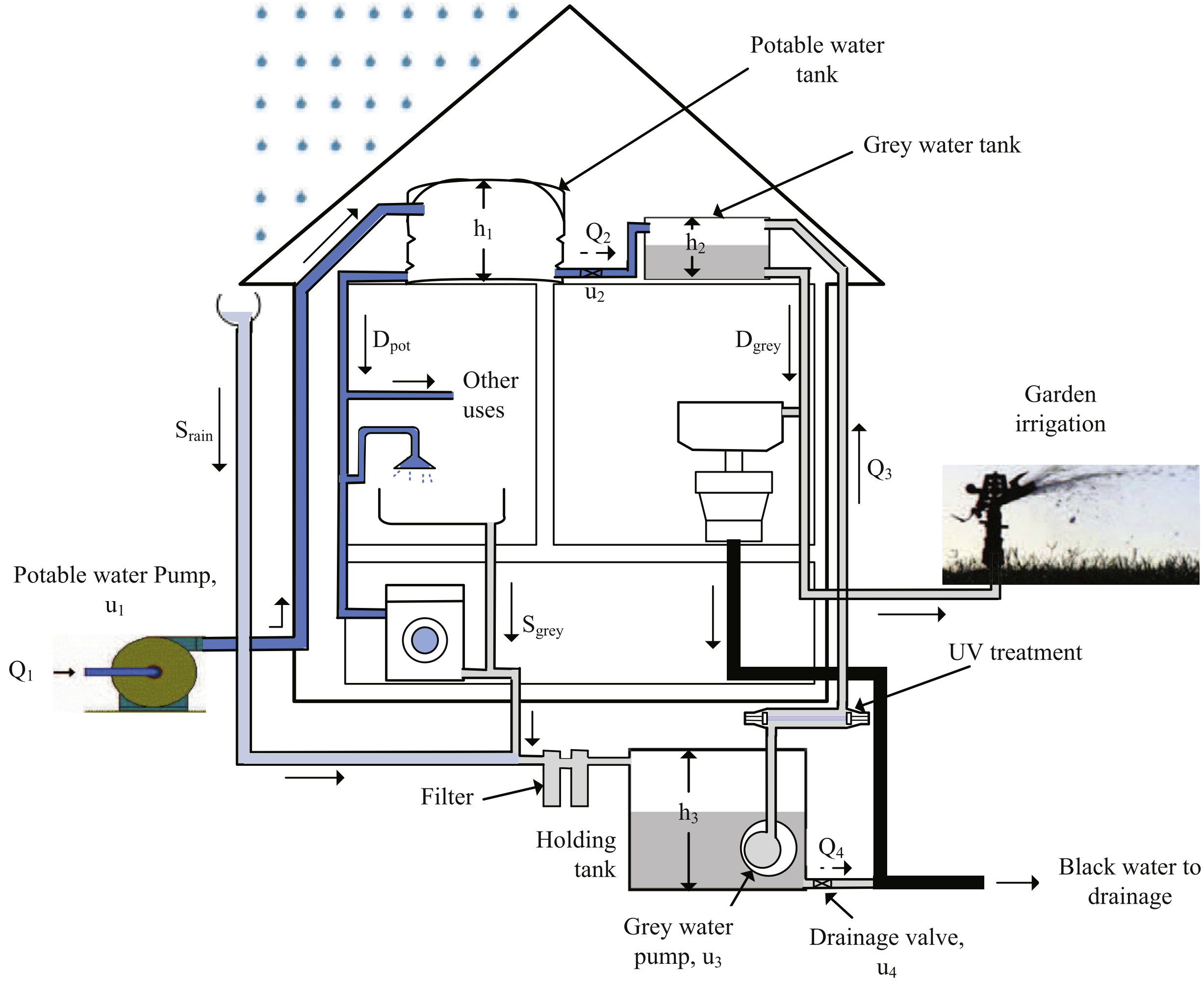


Fig. 1. Schematic of grey and rain water water recycling system.

conservation are achieved.

2.2. Potable water tank

In the scenario where municipal water supply is reliable, potable water tank is not required. On the contrary, if municipal water supply is unreliable, water has to be pumped by potable water pump into the potable water tank for storage, where it flows by gravity to various end uses. Assuming that all tanks in this study have uniform cross-sectional area, the volume of water in this tank, V_1 (m^3), can be modelled as,

$$\dot{V}_1 = A_1^t \dot{h}_1 = Q_1 u_1 - Q_2 u_2 - \dot{D}_{pot}, \tag{1}$$

where A_1^t is the cross-sectional area of the tank (m^2) while h_1 is the height of water in the tank (m). D_{pot} is the potable water demand (m^3) in the house while Q_1 and Q_2 are the flow rates (m^3/h) of potable water pump and solenoid valve respectively. Differential Equation (1) can be expressed in discrete-time domain by a first order difference equation as follows;

$$h_1(j+1) = h_1(j) + \frac{1}{A_1^t}\left[t_s Q_1 u_1(j) - t_s Q_2 u_2(j) - D_{pot}(j)\right], \tag{2}$$

where j the sampling interval and t_s is the sampling period during a full operating cycle of 24-h. Level sensors are economical and easy to use in measuring the volume of water in tanks with uniform cross-sectional area Lipták (2005). Therefore, Equation (2) can be modelled in terms of the water level in a sampling interval, $h_1(j)$, which the controller would use to convert to volume. Through recurrence manipulation, the equation becomes,

$$h_1(j) = h_1(0) + \frac{t_s}{A_1^t}\sum_{i=1}^{j}\left[Q_1 u_1(i) - Q_2 u_2(i)\right] - \frac{1}{A_1^t}\sum_{i=1}^{j} D_{pot}(i) \qquad 1 \le j \le N, \tag{3}$$

where N is the total number of cycles during the full 24-h operating cycle, obtained as $N = \frac{24}{t_s}$.

2.3. Grey water tank

Treated grey water is stored in this tank for future use by non-potable water end uses. If the tank has no water, potable water is allowed into this tank through valve u_2 and then flows to meet the required demand. Therefore, volume, V_2 (m^3), of water in this tank is,

$$\dot{V}_2 = A_2^t \dot{h}_2 = Q_2 u_2 + Q_3 u_3 - \dot{D}_{grey}, \tag{4}$$

where A_2^t is the cross-sectional area (m^2) of the tank, h_2 is the height of water (m) in the tank, D_{grey} is the grey water demand (m^3)

while Q_3 is the water flow rate (m^3/h) through the grey water pump. Expressing Equation (4) in discrete-time domain yields,

$$h_2(j+1) = h_2(j) + \frac{1}{A_2^t}\left[t_s Q_2 u_2(j) + t_s Q_3 u_3(j) - D_{grey}(j)\right], \quad (5)$$

which can further be expressed as

$$h_2(j) = h_2(0) + \frac{t_s}{A_2^t}\sum_{i=1}^{j}[Q_2 u_2(i) + Q_3 u_3(i)] - \frac{1}{A_2^t}\sum_{i=1}^{j} D_{grey}(i) \quad 1 \le j \le N. \quad (6)$$

2.4. Holding tank

Untreated grey and rain water flows through filters to remove physical impurities to temporary storage in the holding tank. Whenever treated grey water is required in the grey water tank, the collected water is pumped by grey water pump through the UV purifier. It is important to empty the holding tank every 24 h to prevent formation of bacteria that cause foul smell. Consequently, the volume, V_3 (m^3), of water in this tank can be modelled as,

$$\dot{V}_3 = A_3^t \dot{h}_3 = \dot{S}_{grey} + \dot{S}_{rain} - Q_3 u_3 - Q_4 u_4, \quad (7)$$

where A_3^t is the cross-sectional area (m^2) of the tank, h_3 is the height (m) of water in the tank. S_{grey} and S_{rain} are the volume (m^3) of water supplied from the recyclable potable water end uses and rain water harvesting respectively, while Q_4 is the flow rate (m^3/h) of untreated grey water through the drainage valve. Expressing Equation (7) in discrete-time domain leads to,

$$h_3(j+1) = h_3(j) + \frac{1}{A_3^t}\left[S_{grey}(j) + S_{rain}(j) - t_s Q_3 u_3(j) - t_s Q_4 u_4(j)\right], \quad (8)$$

which transforms to,

$$h_3(j) = h_3(0) + \frac{1}{A_3^t}\sum_{i=1}^{j}[S_{grey}(i) + S_{rain}(i)] - \frac{t_s}{A_3^t}\sum_{i=1}^{j}[Q_3 u_3(i) + Q_4 u_4(i)] \quad 1 \le j \le N. \quad (9)$$

Dynamic Equations (3), (6) And (9) are used in designing the two controllers that optimally operate the proposed grey and rain water recycling system.

3. Controller design

In this study, two model based controllers that use advanced optimal control concept are designed. The controllers seek to minimize cost of pumping energy of potable and collected grey water, minimize consumption of potable water in the house and finally maximize the life of these pumps through minimizing the maintenance cost normally represented as the number of times a pump is switched on and off during the operating cycle.

3.1. Open-loop optimal controller

The open-loop optimal controller uses the feed forward principle in that hourly water demand in the house is measured prior to running the controller. This demand pattern is used by the controller to predict the future behaviour of the system throughout the full operating cycle. As previously stated, the open-loop controller seeks to minimize the cost of pumping energy, consumption of potable water in the house and maximize the life of the pumps. These performance indicators can be modelled to form the following objective function,

$$J = \sum_{j=1}^{N}\left[\alpha_1 t_s p_e(j) P_1^m u_1(j) + \alpha_2 t_s Q_2 u_2(j) + \alpha_3 t_s p_e(j) P_3^m u_3(j)\right] + \alpha_4 \sum_{j=1}^{N}[s_1(j) + s_3(j)], \quad (10)$$

where P_1^m (kW) and P_3^m (kW) are potable and grey water power pump's power consumption respectively, while p_e and t_s are cost of electricity using the TOU tariff during the j^{th} sampling interval and the sampling time respectively. $s_1(j)$ and $s_3(j)$ are auxiliary variables used to minimize the switching frequency of potable and grey water pumps respectively. Mathaba et al. (2014) show that each auxiliary variable is represented by a value 1 whenever a pump's state changes from off to on. Weights α_1 to α_4 are used to tune the controller according to user's preference. First and third terms in Equation (10) minimize the cost of energy consumed by the pumps, the second term minimizes the consumption of potable water by grey water end uses while the fourth term is responsible for minimizing the switching frequency the two pumps.

Every system functions within certain physical and operational constraints for safe and reliable operation. Constraints present in this system are mathematically modelled as follows;

$$h_1^{min} \le h_1(j) \le h_1^{max}, \quad (11)$$

$$h_2^{min} \le h_2(j) \le h_2^{max}, \quad (12)$$

$$h_3^{min} \le h_3(j) \le h_3^{max}, \quad (13)$$

$$h_3(N) = h_3^f, \quad (14)$$

$$u_1(1) - s_1(1) \le 0, \quad (15)$$

$$u_1(j) - u_1(j-1) - s_1(j) \le 0, \quad (16)$$

$$u_3(1) - s_3(1) \le 0, \quad (17)$$

$$u_3(j) - u_3(j-1) - s_3(j) \le 0, \quad (18)$$

$$u_m(j) \in \{0,1\} \text{ where } m = 1,2,3,4, \quad (19)$$

$$s_1(j), s_3(j) \in \{0,1\}. \quad (20)$$

Various tank capacities are the physical constraints while emptying of the holding tank and switching frequency of the pumps are the main operational constraints affecting the system. Therefore, the tanks are modelled in inequalities (11), (12) and (13) to have the level of water maintained between set minimum and maximum levels. Potable and grey water tanks should never be emptied whereas the holding tank must be emptied within the 24-h operating cycle. This is given by Equation (14), where h_3^f is the final water level in the tank. Inequalities (15) and (17) initialize the auxiliary variables as the initial state of the respective u while inequalities (16) and (18) favour the control that involves less switching frequency of the respective pumps. Finally Equations (19)

And (20) are bounds for the control variables that is, the status of the pumps and switches as well as the auxiliary variables respectively.

3.1.1. Open-loop control algorithm

The objective function and constraints are solved using the canonical form presented by Numbi and Xia (2015);

$$\min f^T X \tag{21}$$

subject to

$$\begin{cases} AX \le b(\text{linear inequality constraint}), \\ A_{eq}X = b_{eq}(\text{linear equality constraint}), \\ L_B \le X \le U_B(\text{lower and upper bounds}). \end{cases} \tag{22}$$

Here, vector X consists of all the control variables in the optimization problem, that is,

$$X=[u_1(1),\dots,u_1(N),u_2(1)\dots,u_2(N),u_3(1),\dots,u_3(N),u_4(1),\dots,u_4(N), s_1(1),\dots,s_1(N),s_3(1),\dots,s_3(N)]^T_{6N\times1}, \tag{23}$$

while elements of vector f^T are obtained from objective function (10) as,

$$A_2 = \begin{bmatrix} 0 & \dots & 0 & -Q_2t_s & 0 & \dots & 0 & -Q_3t_s & 0 & \dots & 0 & 0 & \dots & \dots & 0 \\ 0 & \dots & 0 & -Q_2t_s & -Q_2t_s & \dots & 0 & -Q_3t_s & -Q_3t_s & \dots & 0 & 0 & \dots & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & -Q_2t_s & -Q_2t_s & \dots & -Q_2t_s & -Q_3t_s & -Q_3t_s & \dots & -Q_3t_s & 0 & \dots & \dots & 0 \end{bmatrix}, \tag{30}$$

$$f^T=[\alpha_1t_sP_1^mp_e(1),\dots,\alpha_1t_sP_1^mp_e(N),\alpha_2t_sQ_2\dots,\alpha_2t_sQ_2,\alpha_3t_sP_3^mp_e(1), \dots,\alpha_3t_sP_3^mp_e(N),0,\dots,0,\alpha_4,\dots,\alpha_4,\alpha_4,\dots,\alpha_4]_{1\times6N}. \tag{24}$$

Since there are several linear inequalities, each is modelled separately and later combined into the canonical linear inequality form ($AX\le b$). First, linear inequality constraint (11) is modelled to,

$$\begin{matrix} A_1X \le b_1, \\ -A_1X \le b_2, \end{matrix} \tag{25}$$

where

$$A_1 = \begin{bmatrix} -Q_1t_s & 0 & \dots & 0 & Q_2t_s & 0 & \dots & 0 & 0 & \dots & \dots & 0 \\ -Q_1t_s & -Q_1t_s & \dots & 0 & Q_2t_s & Q_2t_s & \dots & 0 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ -Q_1t_s & -Q_1t_s & \dots & -Q_1t_s & Q_2t_s & Q_2t_s & \dots & Q_2t_s & 0 & \dots & \dots & 0 \end{bmatrix}_{N\times6N}, \tag{26}$$

$$b_1 = \begin{bmatrix} -D_{pot}(1) - A_1^t\left\{h_1^{min} - h_1(0)\right\} \\ -\left\{D_{pot}(1) + D_{pot}(2)\right\} - A_1^t\left\{h_1^{min} - h_1(0)\right\} \\ \vdots \\ -\left\{D_{pot}(1) + \dots + D_{pot}(N)\right\} - A_1^t\left\{h_1^{min} - h_1(0)\right\} \end{bmatrix}_{N\times1} \tag{27}$$

and

$$b_2 = \begin{bmatrix} D_{pot}(1) + A_1^t\{h_1^{max} - h_1(0)\} \\ \{D_{pot}(1) + D_{pot}(2)\} + A_1^t\{h_1^{max} - h_1(0)\} \\ \vdots \\ \{D_{pot}(1) + \dots + D_{pot}(N)\} + A_1^t\{h_1^{max} - h_1(0)\} \end{bmatrix}_{N\times1}. \tag{28}$$

Then, inequality constraint (12) becomes,

$$\begin{matrix} A_2X \le b_3, \\ -A_2X \le b_4, \end{matrix} \tag{29}$$

where

$$b_3 = \begin{bmatrix} -D_{grey}(1) - A_2^t\left\{h_2^{min} - h_2(0)\right\} \\ -\left\{D_{grey}(1) + D_{grey}(2)\right\} - A_2^t\left\{h_2^{min} - h_2(0)\right\} \\ \vdots \\ -\left\{D_{grey}(1) + \dots + D_{grey}(N)\right\} - A_2^t\left\{h_2^{min} - h_2(0)\right\} \end{bmatrix} \tag{31}$$

and

$$b_4 = \begin{bmatrix} D_{grey}(1) + A_2^t\{h_2^{max} - h_2(0)\} \\ \{D_{grey}(1) + D_{grey}(2)\} + A_2^t\{h_2^{max} - h_2(0)\} \\ \vdots \\ \{D_{grey}(1) + \dots + D_{grey}(N)\} + A_2^t\{h_2^{max} - h_2(0)\} \end{bmatrix}, \tag{32}$$

while inequality (13) can be remodelled to,

$$\begin{matrix} A_3X \le b_5, \\ -A_3X \le b_6, \end{matrix} \tag{33}$$

where

$$A_3 = \begin{bmatrix} 0 & \dots & \dots & 0 & Q_3t_s & 0 & \dots & 0 & Q_4t_s & 0 & \dots & 0 & 0 & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 & Q_3t_s & Q_3t_s & \dots & 0 & Q_4t_s & Q_4t_s & \dots & 0 & 0 & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & 0 & Q_3t_s & Q_3t_s & \dots & Q_3t_s & Q_4t_s & Q_4t_s & \dots & Q_4t_s & 0 & \dots & \dots & 0 \end{bmatrix}, \tag{34}$$

$$b_5 = \begin{bmatrix} S_{grey}(1) + S_{rain}(1) - A_3^t\left\{h_3^{min} - h_3(0)\right\} \\ \{S_{grey}(1) + S_{grey}(2) + S_{rain}(1) + S_{rain}(2)\} - A_3^t\left\{h_3^{min} - h_3(0)\right\} \\ \vdots \\ \{S_{grey}(1) + \ldots + S_{grey}(N) + S_{rain}(1) + \ldots + S_{rain}(N)\} - A_3^t\left\{h_3^{min} - h_3(0)\right\} \end{bmatrix} \tag{35}$$

and

$$b_6 = \begin{bmatrix} -\{S_{grey}(1) + S_{rain}(1)\} + A_3^t\{h_3^{max} - h_3(0)\} \\ -\{S_{grey}(1) + S_{grey}(2) + S_{rain}(1) + S_{rain}(2)\} + A_3^t\{h_3^{max} - h_3(0)\} \\ \vdots \\ -\{S_{grey}(1) + \ldots + S_{grey}(N) + S_{rain}(1) + \ldots + S_{rain}(N)\} + A_3^t\{h_3^{max} - h_3(0)\} \end{bmatrix}. \tag{36}$$

Lastly, auxiliary variables in inequalities (15)–(18) are remodelled as

$$A_4X \le b_7, \tag{37}$$

where

$$A_4 = \left[\begin{array}{ccccc:ccc:ccccc:ccc:cccc:cccc} 1 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 1 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & -1 & 0 & \ldots & 0 & -1 & 0 & \ldots & 0 \\ -1 & 1 & \ldots & 0 & 0 & 0 & \ldots & 0 & -1 & 1 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & -1 & \ldots & 0 & 0 & -1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & -1 & 1 & 0 & \ldots & 0 & 0 & 0 & \ldots & -1 & 1 & 0 & \ldots & 0 & 0 & 0 & \ldots & -1 & 0 & 0 & \ldots & -1 \end{array}\right] \tag{38}$$

and

$$b_7 = [0 \quad \ldots \quad 0]^T. \tag{39}$$

Matrices A_1 to A_4 have $(N \times 6N)$ dimension while vectors b_1 to b_7 have a dimension of $(N \times 1)$. Therefore, linear inequality in the canonical form (22) becomes,

$$A = \begin{bmatrix} A_1 \\ -A_1 \\ A_2 \\ -A_2 \\ A_3 \\ -A_3 \\ A_4 \end{bmatrix}_{7N \times 6N} \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \end{bmatrix}_{7N \times 1}. \tag{40}$$

In the same degree, linear equality constraint (14) becomes,

$$A_{eq}X = b_{eq}, \tag{41}$$

where

$$A_{eq} = \left[\begin{array}{ccc:cccccc:ccc} 0 & \ldots & 0 & 0 & \ldots & 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & 0 & \ldots & 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & \ldots & 0 & Q_3t_s & \ldots & Q_3t_s & Q_4t_s & \ldots & Q_4t_s & 0 & \ldots & 0 \end{array}\right]_{N \times 6N} \tag{42}$$

and

$$b_{eq} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ A_3^t\left\{h_3(0) - h_3^f\right\} + \{S_{grey}(1) + \ldots + S_grey(N) + S_{rain}(1) + \ldots + S_{rain}(N)\} \end{bmatrix}_{N \times 1}, \tag{43}$$

while the bounds given in equations (19) and (20) become,

$$L_B = [0 \quad \ldots \quad 0]^T_{6N \times 1} \quad \text{and} \quad U_B = [1 \quad \ldots \quad 1]^T_{6N \times 1}. \tag{44}$$

This binary integer optimization problem is solved using the SCIP solver in OPTI toolbox, a free Matlab optimization toolbox. This solver is indicated as the fastest non-commercial optimization solver by Setlhaolo and Xia (2016).

3.2. Closed-loop MPC control

Closed-loop model predictive control (MPC) has the ability to predict the future behaviour of the system, cope with constraints in the design process and robustly deal with disturbances present in the system. MPC makes use of both feed forward and feed back measurements from the system to compute the control law on-line Mayne et al. (2000). As discussed by Wang (2009), it obtains the current control response by solving an open-loop optimal control optimization problem using the current state of the plant as the initial state in each sampling time. From the optimal sequence generated, only the first control is implemented. The state of the plant (water level in the tanks) is measured. During the next iteration, k + 1, objective function and constraints are updated while taking the previous state of the tanks (water level at sampling time k) as the initial state. The process of optimization is carried out in real time over the new control horizon ($N_c = N - k + 1$) to give the receding horizon control law. This process is repeated throughout the entire operating cycle Xia and Zhang (2015).

The objective function, J_{mpc}, can be derived from the open-loop objective (10) as,

$$J_{mpc} = \sum_{j=k}^{k+N_c-1} \left[\alpha_1 t_s p_e(j) P_1^m u_1(j|k) + \alpha_2 t_s Q_2 u_2(j|k) + \alpha_3 t_s p_e(j) P_3^m u_3(j|k)\right] + \alpha_4 \sum_{j=k}^{k+N_c-1} [s_1(j|k) + s_3(j|k)], \tag{45}$$

where N_c is the control horizon, $u_1(j|k)$, $u_2(j|k)$ and $u_3(j|k)$ are optimized control actions while $s_1(j|k)$ and $s_3(j|k)$ are auxiliary values at j^{th} sampling interval based on most recent measurements carried out at time k. Although MPC problems normally have predicting, N_p, and control, N_c, horizons, only the control horizon, N_c, is included in this optimization problem since none of the state variables (height of water in the tank) is present in the objective function. Therefore, the control horizon can be given as

$$N_c = N - k + 1. \tag{46}$$

State equations are modified from Equations (3), (6) And (9) to,

$$h_1(j|k) = h_1(k) + \frac{t_s}{A_1^t}\sum_{i=k}^{j}[Q_1u_1(i|k) - Q_2u_2(i|k)] - \frac{1}{A_1^t}\sum_{i=k}^{j}D_{pot}(i), \tag{47}$$

$$h_2(j|k) = h_2(k) + \frac{t_s}{A_2^t}\sum_{i=k}^{j}[Q_2u_2(i|k) + Q_3u_3(i|k)] - \frac{1}{A_2^t}\sum_{i=k}^{j}D_{grey}(i), \tag{48}$$

$$h_3(j|k) = h_3(k) + \frac{1}{A_3^t}\sum_{i=k}^{j}[S_{grey}(i) + S_{rain}(i)] - \frac{t_s}{A_3^t}\sum_{i=k}^{j}[Q_3u_3(i|k) + Q_4u_4(i|k)], k \le j \le k + N_c - 1, \tag{49}$$

where $h_1(j|k)$, $h_2(j|k)$, and $h_3(j|k)$ are predicted levels of water in the respective tanks at j^{th} sampling interval based on information available at time k. Moreover, the system experiences the same physical and operational constraints as the open-loop control system. With feedback of water level measurements in various tanks at every iteration, constraints and Equations (11)–(20), are modified to,

$$h_1^{min} \le h_1(j|k) \le h_1^{max}, \tag{50}$$

$$h_2^{min} \le h_2(j|k) \le h_2^{max}, \tag{51}$$

$$h_3^{min} \le h_3(j|k) \le h_3^{max}, \tag{52}$$

$$h_3(N) = h_3^f, \tag{53}$$

$$u_1(1|k) - s_1(1|k) \le 0, \tag{54}$$

$$u_1(j|k) - u_1(j-1|k) - s_1(j|k) \le 0, \tag{55}$$

$$u_3(1|k) - s_3(1|k) \le 0, \tag{56}$$

$$u_3(j|k) - u_3(j-1|k) - s_3(j|k) \le 0, \tag{57}$$

$$u_{m,c}(j|k) \in \{0,1\} \text{ where } m = 1,2,3,4, \tag{58}$$

$$s_1(j|k), s_3(j|k) \in \{0,1\}. \tag{59}$$

3.2.1. MPC algorithm

Similar to the open-loop control algorithm, control vector, represented as X_{mpc}, contains the control variables such that,

$$X_{mpc} = [u_1(k|k), u_1(k+1|k), \ldots, u_1(k+N_c-1|k)), u_2(k|k), u_2(k+1|k), \ldots, u_2(k+N_c-1|k)), u_3(k|k), u_3(k+1|k), \ldots, u_3(k+N_c-1|k)), u_4(k|k), u_4(k+1|k), \ldots, u_4(k+N_c-1|k)), s_1(k|k), s_1(k+1|k), \ldots, s_1(k+N_c-1|k), s_3(k|k), s_3(k+1|k), \ldots, s_3(k+N_c-1|k)]^T_{6N \times 1}. \tag{60}$$

Zhang and Xia (2011) describe the work flow of the MPC controller as follows;

1. For time, k, find the control horizon ($N_c(k)$) using Equation (46).
2. *Optimization:* Find the optimal solution within the control horizon;
   minimize objective function (45),
   subject to constraints (50)–(59).
3. From the optimal solution, implement $[u_1(k|k), u_2(k|k), u_3(k|k), u_4(k|k)]^T$ to the plant.
4. *Feed back:* Measure state variables $h_1(k+1)$, $h_2(k+1)$ and $h_3(k+1)$.
5. Set $k = k+1$ and update system states and inputs and outputs.
6. Repeat steps 1–5 until k reaches a predefined value.

This binary integer optimization problem solved using the SCIP solver in OPTI toolbox.

4. Pertinent information

4.1. Case study

A building in Pretoria, South Africa, has unreliable water supply forcing occupants to pump and store water in a rooftop storage tank, from where it flows by gravity to all end uses. Consumption of potable water and associated energy is used as the baseline for this study. The 0.8 kW fixed speed pump with a flow rate of 0.9 m^3/h is controlled by level switches that just detect empty and full levels. Whenever the tank is empty, water is pumped until the tank is full, regardless of TOU period. End uses in the house were identified and some had their hourly water use measured using digital flow meters and data loggers while others were estimated after interviewing the occupants. These end uses were categorized as those that must use potable water, those that could use treated grey water and those whose used water is suitable for recycling. The hourly water demand for a typical week day and a weekend in this house is shown in Fig. 2.

It can be seen from the curves that the grey water supply, S_{grey}, is always less than the potable water demand, D_{pot}, as some of this potable water qualifies to be recycled. On the contrary, the grey water demand, D_{grey}, doesn't necessarily follow the others, as this demand entirely depends on the human behaviour.

The current cylindrical potable water tank has the dimensions given in Table 1. In order to incorporate grey and rain water recycling, two tanks; grey and holding water tanks are required. Typical dimensions and capacity constraints of these tanks are given in Table 1.

Level sensors will be used monitor the water level between minimum and maximum levels given in Table 1. This is to ensure safe and reliable operation of the system by avoiding either running the tanks completely empty or spilling the water hence damaging the roof of the house. The grey water pump to be incorporated would be rated at 650 W with flow rate of 0.35 m^3/h.

To enable rain water harvesting, about 50 m^2 of the house's roof can easily have rain water directed to the holding tank, through the filters. The area's weather data, that includes the hourly amount of rainfall, is obtained from the Southern African Universities

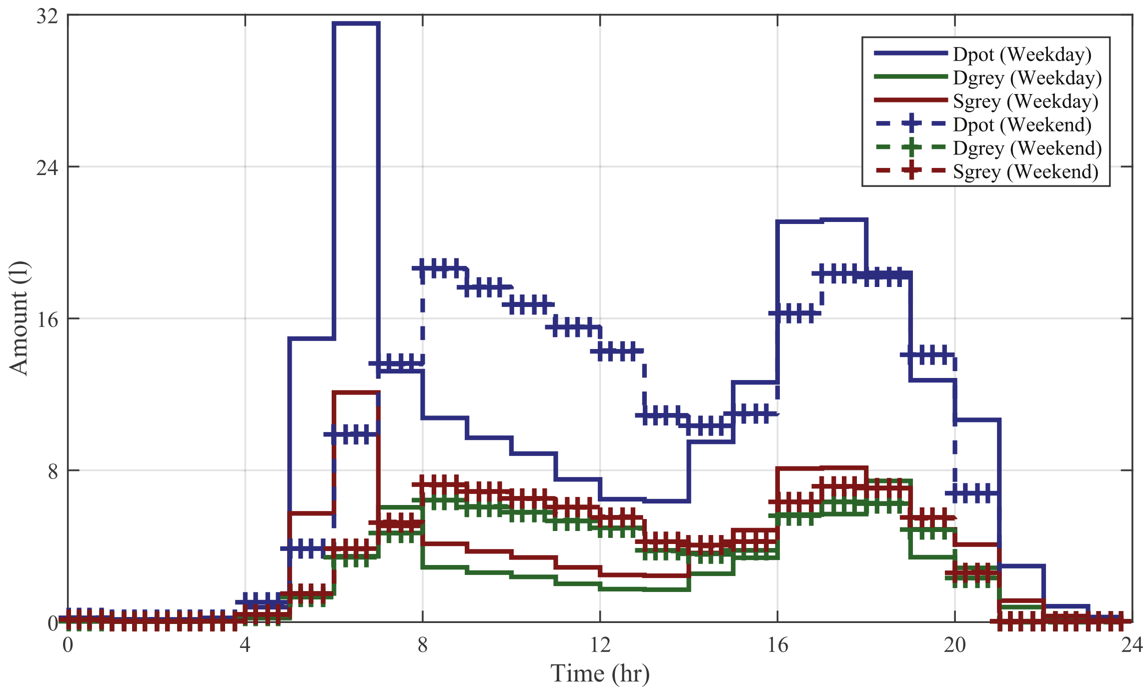


Fig. 2. Hourly water profile for a typical week day and weekend.

Table 1
Dimensions and capacity of the tanks.

Tank	Radius (m)	Height (m)	Min	Max
Potable	0.55	1.2	0.1	1.0
Grey	0.36	1.0	0.1	0.8
Holding	0.30	0.6	0	0.5

Table 2
City of Tshwane tariff for 2014/2015[3].

Water (m^3/month)	0–6	7–12	13–18	19–24	25–30	31–42	43–72	>72
Price (R/m^3)	6.81	9.72	12.77	14.77	16.89	18.25	19.53	20.91
Discharge (%/month)	98	90	75	60	52	10	1	1
Price (R/m^3)	5.06	6.83	8.81	8.81	8.81	8.81	8.81	8.81

Radiometric Network,[1] University of Pretoria's station.

4.2. Time-of-use electricity tariff

Time-of-use (TOU) tariff, commonly used across the world to encourage shifting of peak load, and Middelberg et al. (2009) show that it can vary by time of day, day of week and season. Eskom's TOU Homeflex[2] structure for residential consumers given below is used.

$$p_e(t) = \begin{cases} p_{off} & = 0.5510 \quad R/Kwh \quad \text{if } t \in [0,6] \cup [10,18] \cup [20,24], \\ p_{peak} & = 1.748 \quad R/Kwh \quad \text{if } t \in [7,10] \cup [18,20], \end{cases} \tag{61}$$

where p_{off} is the off peak price, p_{peak} is the peak time price, R is the South African currency, Rand, and t is the time of day in hours.

4.3. Potable and waste water tariffs

Table 2 shows the water and waste water tariffs for domestic consumers in the City of Tshwane.[3]

The amount of waste water discharged into the drainage system is calculated as a percentage of the amount of potable water consumed in a household per month. Since potable and waste water are charged through an incremental block tariff, it is important to carry out simulations over a month in order to obtain the cost incurred by the end user. It is assumed that the demand pattern repeats itself over the 24-h operating cycle, for weekdays and the two days of the week end. In this study, the weekday water demand profile, $D_{pot}(weekday)$, is assumed to be the same for all the 5 week days. Similarly, the weekend demand profile, $D_{pot}(weekend)$, is also taken to be the same for the 2 days of the weekend. Therefore, both open-loop and closed-loop control systems are run over the 24-h operating cycle and then repeated over a month. The month is taken to have 4 complete weeks with each week having 5 weekdays and 2 days of the weekend. Taking the first day of the month to be a Monday, the cumulative volume of potable water consumed up to a certain weekday, $D_{pot,wkdy}$, or a weekend, $D_{pot,wknd}$, is obtained as;

$$\begin{aligned} D_{pot,wkdy} &= (5q)D_{pot}(weekday) + (2q-2)D_{pot}(weekend), \\ D_{pot,wknd} &= (5q)D_{pot}(weekday) + (2q-1)D_{pot}(weekend), \end{aligned} \tag{62}$$

where q is the number of the week in the month ($q = 1, 2, 3, 4$). This amount is then used to compute the amount of waste water discharged and eventually the cost of potable and waste water in a month.

5. Analysis of optimal results

The two control strategies are verified using the case study in

[1] http://www.sauran.net.
[2] http://www.eskom.co.za/.
[3] www.tshwane.gov.za.

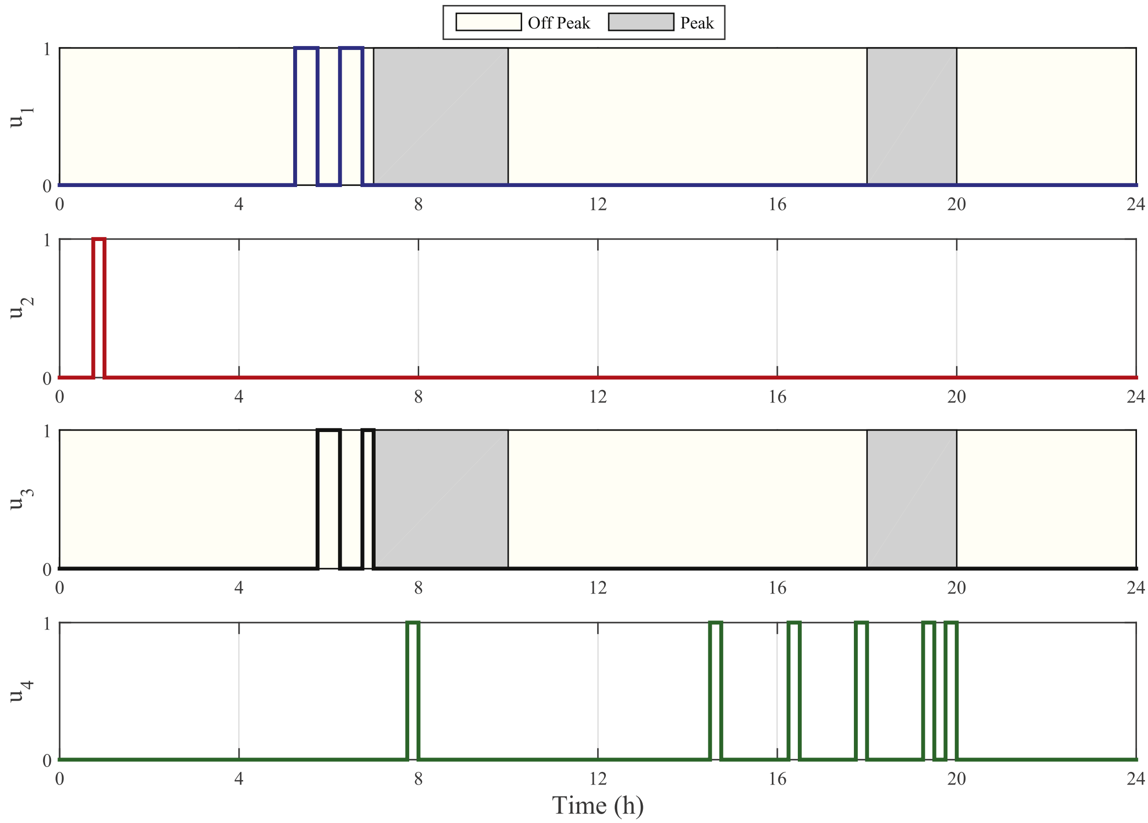


Fig. 3. Optimal operation of pumps and valves by open-loop controller.

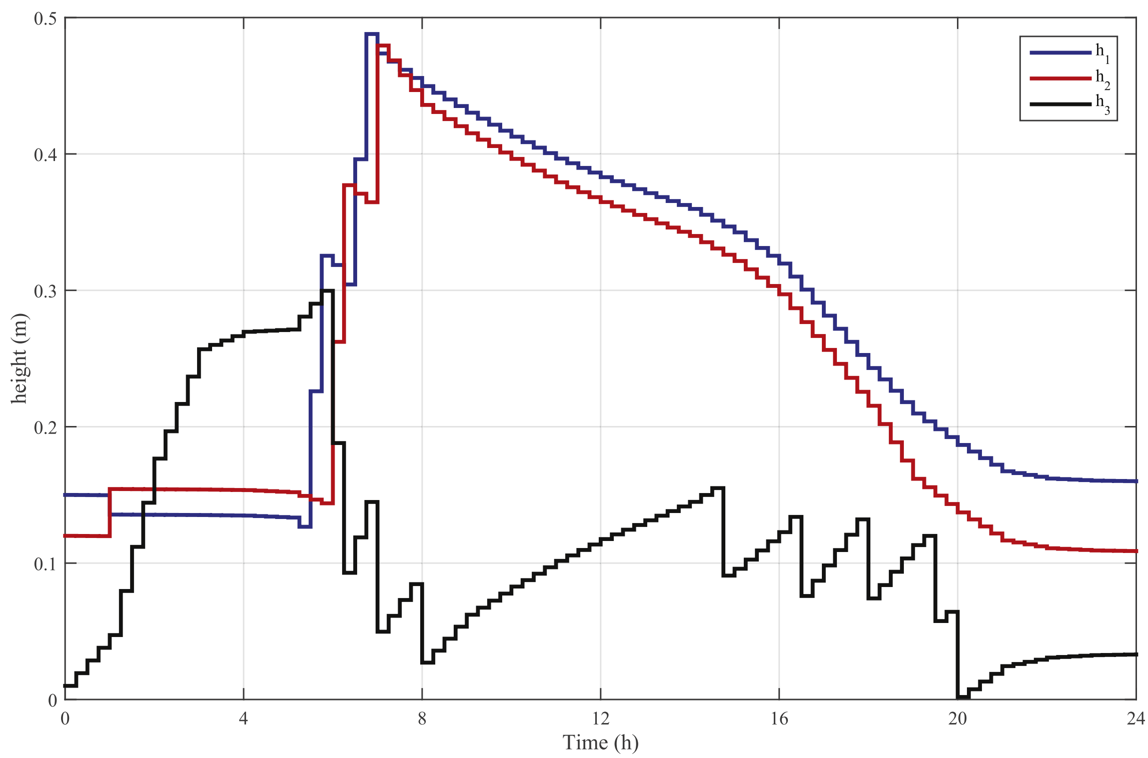


Fig. 4. Variation of water level in respective tanks with open-loop controller.

section 4.1. Simulations are carried out for an operating cycle of 24-h with a sampling period, $t_s = 15$ minutes. The legend showing peak and off peak periods of the TOU electricity tariff (section 4.2) is used throughout the paper. Moreover, only potable and grey water pumps, whose status are represented by u_1 and u_3 respectively, are considered to consume power hence subjected to the TOU tariff.

5.1. Open-loop optimal control strategy

The optimal operation of the pumps and valves in the proposed system by using open-loop optimal controller is shown in Fig. 3.

The controller operates both potable and grey water pumps during the off-peak period of the TOU tariff in meeting the household potable and grey water demand. This effectively shifts the electrical load to the period when the grid experiences less load, hence improving its stability. In addition, the controller switches both pumps only 2 times during the 24-h operating cycle in line with the objective seeking to minimize the maintenance cost of the pumps. This cost is attributed to frequent switching of a pump that causes wear and tear to its motor as it tries to overcome dead load (water) while changing from off to on status. In addition, the controller operates potable water valve once in early morning to supplement treated grey water. It also operates the drainage valve several times after predicting that collected water is no longer needed for treatment and pumping, and yet the holding tank has to be emptied within the operating cycle.

Optimal operation of the proposed system using the open-loop controller leads to variation of water level in various tanks as shown in Fig. 4.

The controller does not violate any constraints in operating the system throughout the 24-h operating cycle. After a 15 min draw by potable water valve, the controller predicts that potable tank does not have sufficient water to take it through the high morning water demand, which coincides with peak TOU period. It, therefore, opts to switch on the pump twice at 05:15 and 06:15 for 30 min each, raising the level, h_1, to 0.49 m, which is sufficient for the remaining period of the operating cycle. After this, water level h_1 keeps dropping while meeting potable water demand to a low of 0.16 m at the end of the operating cycle. At the onset, the holding tank is empty while the treated water level in the grey water tank is almost at the minimum allowable level. For this reason, the controller has to use potable water to meet grey water demand in the early morning leading to the potable water valve being switched on at 00:45 for 15 min. A simultaneous rise of water level in grey water tank and drop in potable water tank takes place during this time. As the day progresses, more water is collected hence there is no need for using potable water for non-potable uses. The controller predicts an increase in grey water demand in the morning hours, which again coincides with the peak TOU period. It consequently operates the grey water pump twice at 05:45 for 30 min and 06:45 for 15 min leading to a rise in level h_2 to 0.48 m, which is sufficient to meet the grey water demand for the rest of the operating cycle. In addition, the pumping leads to a simultaneous drop in level h_3 to 0.05 m. Since the controller predicts that the treated grey water is sufficient for the rest of the operating cycle, it then keeps draining the collected water to the drainage, and also ensures that the tank is emptied within the operating cycle to avoid formation bacteria responsible for foul smell.

5.2. Closed-loop MPC strategy

The closed-loop MPC strategy operates the proposed system by switching the pumps and valves as shown in Fig. 5.

Just like the open-loop controller, the closed-loop controller also operates both pumps during the cheaper off-peak TOU periods, in line with the utility's desire. Additionally, the closed-loop controller

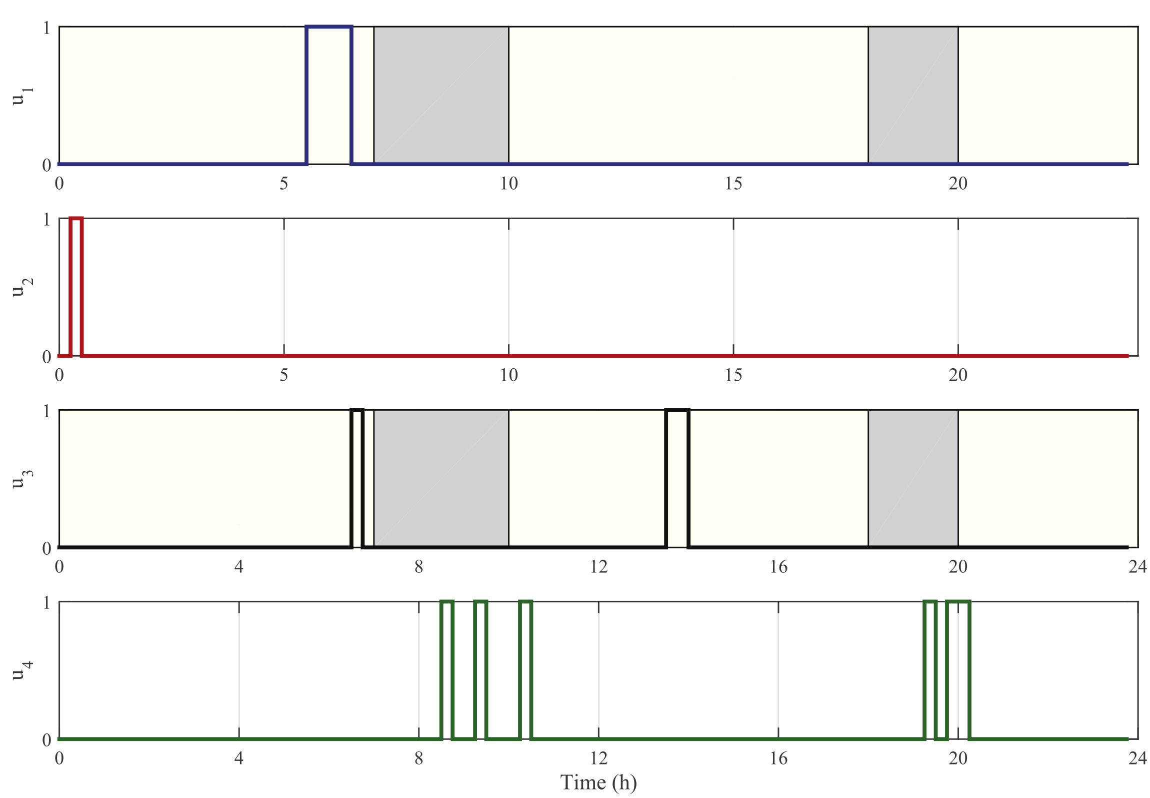


Fig. 5. Optimal operation of pumps and valves using MPC strategy.

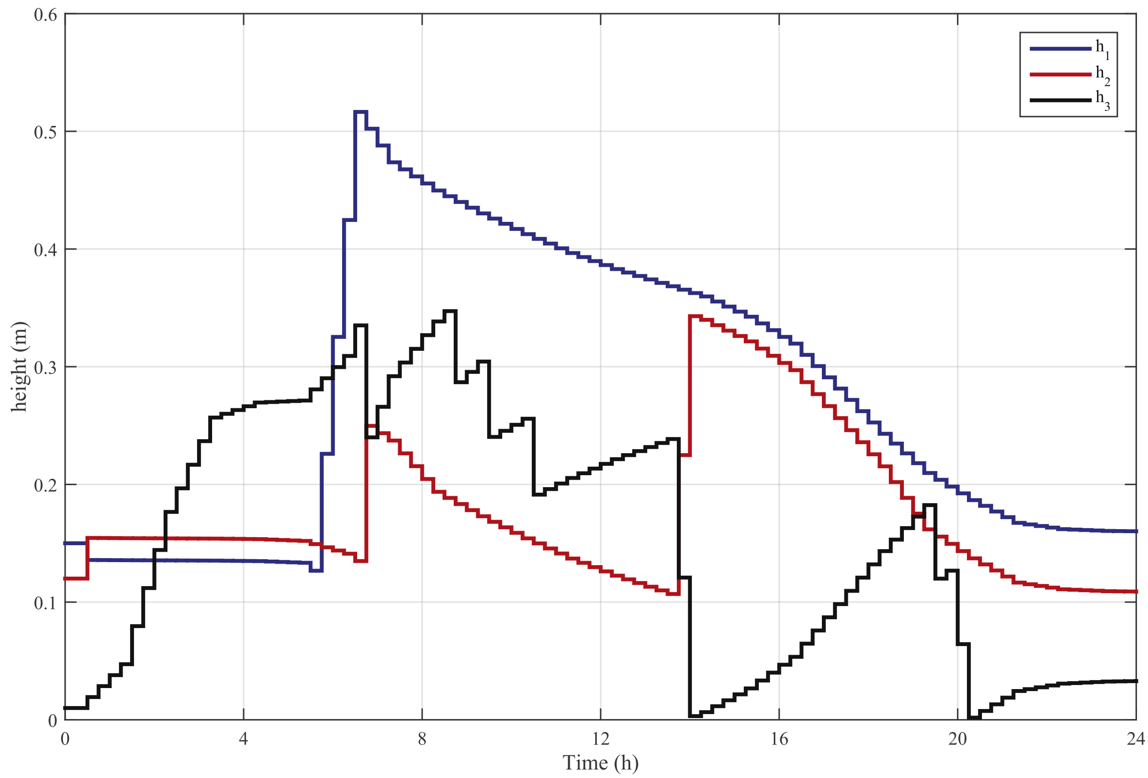


Fig. 6. Variation of water level in respective tanks with MPC.

ensures that both pumps are not switched on frequently in order to minimize the maintenance cost. In predicting increasing potable water demand in the same period as the peak TOU period, the closed-loop controller switches the potable water pump once at 05:30 for 1 h. This water is enough to meet potable water demand in the house for the remaining period of the operating cycle. In addition, the controller operates the grey water pump twice, first at 06:30 for 15 min and later at 13:30 for 30 min. However, the solenoid valves are switched on at any time since they use negligible amount of power. The controller switches the potable water valve early in the morning at 00:15 for 15 min when there is insufficient collected and treated grey water, and yet there is grey water demand to be met. It also switches the drainage valve frequently as more water is collected during the operating cycle to ensure the tank is emptied.

Optimal operation of the pumps and valves leads to variation of water levels in various tanks as shown in Fig. 6.

The closed-loop controller also ensures that none of the constraints is violated. Similar to the open-loop controller, the closed-loop controller predicts that the amount of stored potable water is not sufficient to meet the high potable water demand that coincides with the morning TOU peak. This situation is made worse, by insufficient collected and treated grey water in respective tanks forcing a 15 min draw of potable water to meet grey water demand in the early morning. This prompts the controller to operate the potable water pump at 05:30–06:30 raising the water level, h_1, to 0.52 m, which is enough to meet potable water demand for the remaining period of the operating cycle. To meet the early morning grey water demand, the controller has to operate the potable water valve, u_2 for 15 min leading to a rise in level h_2 to 0.15 m and a simultaneous drop of h_1 to 0.14 m. By morning hours, enough water has been collected in the morning even though the demand for grey water is increasing during the peak TOU period. Consequently, the controller pumps water from the holding tank at 06:30 raising the water level in the grey tank to 0.25 m while at the same time leading to a drop of water level in the holding tank to 0.24 m. The treated grey water helps in meeting the morning water demand but unfortunately, it is not enough for the rest of the operating cycle. Therefore more water is treated and pumped to the grey tank at 13:30 for 30 min raising water level, h_2, to 0.34 m while also

Table 3
Comparison of weekly water consumption.

Wk	Day	Baseline				Proposed strategies			
		Potable (m^3)	Cost (R/m^3)	Waste (m^3)	Cost (R/m^3)	Potable (m^3)	Cost (R/m^3)	Waste (m^3)	Cost (R/m^3)
1	Weekday	5.80	6.81	5.68	5.06	4.47	6.81	4.38	5.06
	Weekend	8.14	9.72	7.81	6.83	6.34	9.72	6.19	6.83
2	Weekday	13.94	12.77	12.89	8.81	10.81	9.72	10.21	6.83
	Weekend	16.29	12.77	14.65	8.81	12.68	12.77	11.79	8.81
3	Weekday	22.09	14.77	18.53	8.81	17.15	12.77	15.14	8.81
	Weekend	24.43	16.89	19.90	8.81	18.89	14.77	16.31	8.81
4	Weekday	30.23	18.25	22.82	8.81	23.31	14.77	18.97	8.81
	Weekend	31.61	18.25	25.98	8.81	24.18	16.89	19.47	8.81

Table 4
Weekly consumption by grey water end uses.

Wk	Day	Potable water (m^3)		Treated water (m^3)	
		Open-loop	MPC	Open-loop	MPC
1	Weekday	0.05	0.05	0.18	0.18
	Weekend	0.13	0.13	0.18	0.18
2	Weekday	0.05	0.05	0.18	0.18
	Weekend	0.13	0.13	0.18	0.18
3	Weekday	0.05	0.05	0.18	0.18
	Weekend	0	0	0.26	0.26
4	Weekday	0	0	0.26	0.26
	Weekend	0	0	0.26	0.26

emptying the holding tank, as desired. Thereafter, the closed-loop controller predicts that the water in both storage tanks is sufficient to meet the demand for the remaining period of the operating cycle, hence, no more pumping is required. It therefore keeps operating the drainage valve and empties the tank again in the evening, in line with ensuring that the tank remains healthy and bacteria forms.

5.3. Analysis and discussion

The performance of the two optimal controllers is compared with the baseline, where potable water is used to meet all end uses in the house, over a period of one month. Table 3 shows the weekly water consumption, waste water discharge and the associated cost in the baseline and the proposed water recycling system operated using either control strategies. The consumption of water presented in the table holds for both scenarios with reliable and unreliable municipal water supply while the cost of water and waste water discharge is obtained using tariffs provided in section 4.3. Baseline and proposed strategies columns show the cumulative amount of potable water consumed and waste water discharged from the house together with their respective unit price. The weekday or weekend cumulative water is the amount of either potable or treated grey water used in the house at the end of 5 week days or 2 days of the weekend respectively.

It is evident that more potable water is consumed in the baseline than when using the water recycling and harvesting system controlled by either control systems. Consequently, more waste water is discharged from the baseline than from the proposed system. As a result, the household ends up paying for potable water at a maximum unit cost 18.25 R/m^3 in the baseline and as opposed to 16.89 R/m^3 in the proposed strategies at the end of the month. Similarly, the household currently (baseline) pays for waste water at the maximum 8.81 R/m^3 in a month from the second week while the proposed water conservation interventions would lead to the same happening from the weekend of the second week. Therefore, end users, whether with reliable or unreliable water supply, will have the added benefit of lower cost of potable and waste water in addition to conserving it. Operation of the proposed system by either control strategies consumes about 0.05 m^3 and 0.13 m^3 of potable water for grey uses in a week day and weekend, respectively, during the first 2 weeks as shown in Table 4.

Thereafter, 0.05 m^3 is used during the week day of the third week. Up to this point, the cost of water has risen to 12.77 R/m^3. Nonetheless, during the weekend of the third week, when the unit price rises to 14.77 R/m^3, both controllers do not use potable water for grey end uses. This results from weight of the term responsible for minimizing the cost of water in objective functions (10) and (45) increasing significantly, making both controllers to give this term more preference as compared to the other terms. Further, the increasing weighting factor leads to an increase in the use of grey water from 0.18 m^3 to 0.26 m^3 for both controllers.

The monthly water and energy consumption, waste water discharge and the associated costs in the baseline and the two control strategies operating the water recycling and harvesting system are compared as shown in Table 5.

If municipal water supply is reliable, the baseline water consumption is about 31.61 m^3 leading to a discharge of about 25.98 m^3 in a month. Therefore, the cost of both water supply and waste water discharge is about 589.69 $R/month$, as there is no energy cost associated with pumping potable water for storage. Adoption of the proposed system in such a house would reduce the monthly potable water consumption and waste discharge by about 23.5% and 25.1% with the corresponding cost reduction of 32.3% and 29.5% respectively. Considering the proposed system would not pump potable water for storage, up to 30.8% cost savings can be achieved, irrespective of the added cost of energy incurred by the grey water pump. In the second scenario where potable water is pumped and stored, the baseline still uses 31.61 $m^3/month$ of potable water, discharges about 25.98 $m^3/month$ but at a higher cost of 604.02 $R/month$, resulting from cost of pumping potable water to the storage tank. Incorporation of the proposed system would still conserve about 23.5% potable water, reduce discharge by up to 25.1%, corresponding to monthly cost saving from water and waste water of 32.3% and 29.5% respectively. In addition, the two controllers can save the cost of energy by up to 35.7% through shifting the load to the cheaper off-peak periods of the TOU tariff. Eventually, the proposed water recycling system with optimal control would save up to 31.5% of the total operation cost. This shows that the proposed system would conserve water, reduce waste water discharge and lead to economic benefits in both water supply scenarios.

Previous studies have shown that the two control strategies are known to adapt differently. Luo et al. (2015) hold that the closed-loop MPC is stable and robust in dealing with disturbances, unlike the open-loop controller which, as discused by Wanjiru et al. (2016b), can only deal with disturbances that do not largely change the demand pattern. This, however, comes at a higher cost in terms of computation and financial as well as more complexity, since it would require extra components to enable feedback of the height of water in the tanks to take place Qi et al. (2015). Adoption of either controller depends on the nature of each application. The open-loop controller is suitable where the demand pattern does not change significantly, otherwise the closed-loop MPC is suitable.

Collection of grey water is dependent on human behaviour making it a more reliable alternative source of water for non-

Table 5
Proposed system's operating costs.

	Baseline	Open-loop	MPC
Potable water			
Amount ($m^3/month$)	31.61	24.18	24.18
Cost ($R/month$)	395.15	267.46	267.46
Waste water			
Amount($m^3/month$)	25.98	19.47	19.47
Cost ($R/month$)	194.54	137.18	137.18
Potable pump			
Energy ($kWh/month$)	13.04	8.00	8.00
Cost ($R/month$)	14.33	5.84	5.84
Grey pump			
Energy($kWh/month$)	0[a]	3.25	3.25
Cost ($R/month$)	0[a]	3.37	3.37
Total cost ($R/month$)[b]	604.02	413.85	413.85

[a] The household was only using potable water.
[b] Cost of pumping energy, potable and waste water.

Table 6
Life cycle cost analysis of the water recycling and harvesting system.

Year	Initial investment (R)		Salvage value (R)	Annual cost (R)		Revenue (R)	Total (R)	Discounting factor	Cash flows (R)	
				Operation	Maintenance				Discounted	Cumulative
0	Water tanks	(3500)	1800							
	Pumps	(280.00)	50							
	UV purifier and filters	(4500.00)	1700							
	Controller	(10 000.00)	500							
	Accessories	(6000.00)								
	Installation cost	(8000.00)								
	Total Capital	(28 230.00)						1.00	(28 230.00)	(28 230.00)
1				(4966.20)	(200.00)	2282.04	(2884.16)	0.94	(2706.61)	(30 936.61)
2				(4966.20)	(200.00)	2282.04	(2884.16)	0.88	(2539.98)	(33 476.59)
3				(4966.20)	(200.00)	2282.04	(2884.16)	0.83	(2383.62)	(35 860.21)
4				(4966.20)	(200.00)	2282.04	(2884.16)	0.78	(2236.88)	(38 097.09)
5				(4966.20)	(200.00)	2282.04	(2884.16)	0.73	(2099.17)	(40 196.26)
6				(4966.20)	(200.00)	2282.04	(2884.16)	0.68	(1969.94)	(42 166.21)
7				(4966.20)	(200.00)	2282.04	(2884.16)	0.64	(1848.67)	(44 014.88)
8				(4966.20)	(200.00)	2282.04	(2884.16)	0.60	(1734.86)	(45 749.74)
9				(4966.20)	(200.00)	2282.04	(2884.16)	0.56	(1628.06)	(47 377.81)
10				(4966.20)	(200.00)	2282.04	(2884.16)	0.53	(1527.84)	(48 905.64)
11				(4966.20)	(200.00)	2282.04	(2884.16)	0.50	(1433.78)	(50 339.43)
12				(4966.20)	(200.00)	2282.04	(2884.16)	0.47	(1345.52)	(51 684.94)
13				(4966.20)	(200.00)	2282.04	(2884.16)	0.44	(1262.68)	(52 947.63)
14				(4966.20)	(200.00)	2282.04	(2884.16)	0.41	(1184.95)	(54 132.58)
15				(4966.20)	(200.00)	2282.04	(2884.16)	0.39	(1112.00)	(55 244.58)
16				(4966.20)	(200.00)	2282.04	(2884.16)	0.36	(1043.55)	(56 288.13)
17				(4966.20)	(200.00)	2282.04	(2884.16)	0.34	(979.30)	(57 267.43)
18				(4966.20)	(200.00)	2282.04	(2884.16)	0.32	(919.02)	(58 186.45)
19				(4966.20)	(200.00)	2282.04	(2884.16)	0.30	(862.44)	(59 048.89)
20				(4966.20)	(200.00)	2282.04	(2884.16)	0.28	(809.35)	(59 858.24)

potable end uses. On the other hand, raining is a natural occurrence making rain water harvesting largely dependent on climatic conditions and weather patterns. Since both rain water and grey water are collected to the same holding tank, rain water would have greatest impact early in the morning when the holding tank is almost empty. This would enable the two controllers to use this other than using potable water, and would lower the cost of operation even further. In countries with low amount of annual rainfall, like South Africa, grey water is a more reliable alternative source of water, and its amount hugely determines the design of the system. However, for regions with higher annual rainfall, the system design could be modified to maximize the efficiency of rain water. The potential of rain water harvesting in a year is estimated as a product of local precipitations, roof's catchment area and a non-dimensional runoff coefficient. This coefficient is important in accounting for losses arising due to spillage, leakage, wetting the surface and evaporation. It is therefore useful in predicting the amount of water running off the roof's surface and is conveyed to the storage system. Generally, sloping roofs have a higher coefficient than flat roofs Farreny et al. (2011). The two control systems can easily be modified for different types of roofs so as to accurately predict the amount of rain water that can be collected. Wide adoption of the system would greatly and positively influence the environment. The savings out of water conservation, waste water reduction and energy efficiency would immensely benefit municipal companies and energy utilities over a long time due to reduced demand for the two resources and sewerage services.

5.4. Life cycle cost analysis

It is necessary to evaluate the feasibility of implementing any project, not only in terms of environmental benefits, but also based on economic effect. The cost effectiveness of implementing this water recycling and harvesting system is based on comprehensive consideration of various cost and revenue components. One effective method is the present worth method that discounts back all future elements of the financial analysis of a project to their present worth, apart from capital costs that are already given in present terms. Thereafter, the positive and negative elements of the cash flow are summed, and if the net present value (NPV) is positive, then Vanek et al. (2012) argue that the investment is financially attractive. Life cycle cost (LCC) involves carrying out such analysis over the entire life of the project, and therefore has the benefit of capturing all costs and revenue that would take place during operation of the project. In this analysis, it is assumed that interest rate, taken as the inflation, revenue and operation cost are constant throughout the life of the project. Based on Capehart et al. (2006), costs included in analysis of LCC include cost of acquisition, operation, maintenance and disposal. Therefore,

$$LCC = C_c + C_o + C_s \tag{63}$$

where C_c is the capital cost, C_o is the operation cost and C_s is the salvage cost at the end of life of the system. Capital cost includes total cost of acquiring and installing the system and labour. In the operation stage, the operation cost includes water, waste water, energy and maintenance cost incurred during the service life. Finally, salvage cost is the cost incurred at the end of system's life including the salvage value of the system, cost of removal and disposal Bull (2015). Equation (63) can be written in terms of the discounting factor, that is, the factor by which future cash flows must be multiplied with to get the present worth, as,

$$LCC = C_c + \sum_{n=1}^{m} \frac{C_o(n)}{(1+r)^n} + C_s, \tag{64}$$

where n and m are the number of years and project lifetime respectively while r is the discounting factor. The costs involved in this study are based on the South African market rates. Some assumptions are made while carrying out the life cycle cost analysis of the proposed system; the discounting factor is taken as South Africa's average inflation rate in 2016. The inflation rate, depreciation, operation and maintenance costs are assumed to be constant throughout the life of the system. The annualized cost and revenue are average from monthly values are obtained when the simulations are carried out over the four seasons in a year having varying demand for water.

Table 6 shows the life cycle cost analysis of the proposed system controlled using the MPC strategy. Expenses are indicated using negative values (brackets) while revenue is indicated as positive values. A discount factor of 6.56%, which is the average inflation rate of South Africa for 2016[4] is used to obtain the time value of money. In this analysis, all capital investment is taken to be done in the beginning of the project, and all components of the system will be operational for the 20 year life of the system. Further, cost of potable and waste water is assumed to remain constant for the entire life of the project. The discounted cash flows continuously increase the cumulative cash flows in each year, and the year which the cumulative cash flows becomes zero is an indicator of the break even point or the payback period. In this case, the proposed water recycling and harvesting system does not pay back in its 20 year life period, attributed to high capital cost coupled with low cost of water in South Africa, even though the country is semi arid. These findings are similar to a study done in two universities in South Africa by Ilemobade et al. (2013), as well as in other parts of the world such as Li et al. (2010) in Ireland, Fountoulakis et al. (2016) in Greece and Jabornig (2014) in Austria. A study by Adewumi et al. (2010) reveals that the current low water tariffs significantly influence end users' willingness to embrace water recycling. Government subsidies are therefore necessary in order to encourage the uptake of these technologies that will help in preventing water insecurity around the country and the region. Even though the proposed strategy currently looks economically infeasible, it is important to conduct a thorough analysis while considering full cost of water supply and waste water treatment.

South Africa is a semi-arid country that has constantly struggled to provide reliable water supply to the population. In the recent past, the situation has become worse forcing municipalities to implement water restriction in various parts of the nation.[5] In addition, since the demand for water and energy is expected to keep growing as the population increases, their will keep increasing and the proposed system could soon become economically feasible. The implication of water scarcity and increased pressure on existing infrastructure is evident in the City of Cape Town[6] where the municipality has opted to increase the cost of water and waste water in an effort to encourage efficient and sustainable use. The proposed water recycling and harvesting system is therefore an important intervention in ensuring reliable water supply, water conservation and energy efficiency are achieved.

6. Conclusion

Water and energy, two inseparable resources, are vital for sustainable economic development of any country. Supply of these resources is however unreliable in South Africa due to various factors such as climatic factors, population increase, improved standards of living and rapid urbanization. This has led in increased demand surpassing existing supply capacity and various demand management strategies are required. Grey water recycling and rain water harvesting are suitable for conserving water by providing alternative sources hence reducing the demand for water and waste water services from the municipalities. Two controllers are designed in this study to optimally operate the grey water and rain water recycling system for a house subject to the TOU electricity tariff in South Africa, where a case was considered. The two controllers perform the same, however, the open-loop controller is easier and more cost effective to implement while the closed-loop MPC is more robust and reliable in controlling the proposed system in domestic houses. The proposed system can potentially reduce potable water consumption by 23.5% and consequent waste water discharge by 25.1% as compared to the baseline. Optimal operation of the system using either controllers can reduce the cost of energy by 35.7% through load shifting. For the two scenarios considered in this study, that is reliable and unreliable municipal water supply, optimal operation of the proposed system can lead to a total operation cost saving of up to 30.8% and 31.5% respectively. Despite the proposed system having the benefit to conserve water and efficiently use energy, it does not pay back the cost within its lifetime. Importantly, studies in other parts of the world have shown similar results, predicting that the worsening water insecurity due to increased demand and climate change could make such systems financially feasible in the near future. Governments should encourage adoption of these systems in order to conserve water and environment at large. This can be done through policies, regulations, subsidies and incentives to encourage their uptake especially in cities with functional centralized water and waste water systems. Otherwise, there will be no motivation for building developers and owners to consider these important systems. In cities with dysfunctional or non-existent centralized water supply and waste water systems such as Nairobi in Kenya, Jakarta in Indonesia and Lima in Peru, building owners rely on exorbitantly expensive water vendors to augment water supply and septic tanks for sanitation. Considering these cities are rapidly expanding, the demand for housing requiring water and sanitation infrastructure will keep increasing. It is therefore prudent for authorities in such countries to develop proper and acceptable policies that would increase water and sanitation security. The decentralized system developed in this paper is necessary and comes as a huge relief in increasing reliability and security of water supply at lower cost. Better still, less water would go down the drain taking a longer period before the septic tank requires emptying. More studies could be carried out in these areas to determine the economic feasibility. The system can be incorporated in new building designs or retrofitted in existing ones, which would be more expensive. The complexity of retrofitting would, however, depend on the plumbing system and space available. Wide adoption of the system would have huge benefit to the environment and municipalities that would not require to rapidly expand their existing supply and drainage infrastructure. This study forms the basis for future research into optimal operation of grey water recycling and rain water harvesting systems in the built sector, whether domestic or commercial, to enhance resources conservation and efficiency.

Acknowledgements

The authors would like to thank the National Hub for Energy Efficiency and Demand Side Management (EEDSM) for supporting this research.

References

Adewumi, J., Ilemobade, A., Van Zyl, J., 2010. Treated wastewater reuse in South Africa: overview, potential and challenges. Resour. Conservat. Recycl. 55,

[4] www.inflation.eu/inflation-rates/south-africa/historic-inflation/cpi-inflation-south-africa-2016.aspx.

[5] www.droughtsa.org.za/about-the-drought/water-restrictions.html.

[6] https://www.westerncape.gov.za/general-publication/how-manage-water-restrictions-your-home.

221–231.
Al-Jayyousi, O.R., 2003. Greywater reuse: towards sustainable water management. Desalination 156, 181–192.
Badenhorst, W., Zhang, J., Xia, X., 2011. Optimal hoist scheduling of a deep level mine twin rock winder system for demand side management. Electr. Power Syst. Res. 81, 1088–1095.
Bohensky, E.L., 2006. A Social-ecological Systems Perspective on Water Management in South Africa. Ph.D. thesis. University of Pretoria.
Brandoni, C., Bošnjaković, B., 2017. HOMER analysis of the water and renewable energy nexus for water-stressed urban areas in Sub-Saharan Africa. J. Clean. Prod. 155, 105–118.
Bull, J.W., 2015. In: Illustrated (Ed.), Life Cycle Costing: for the Analysis, Management and Maintenance of Civil Engineering Infrastructure. Whittles Publishing.
Capehart, B.L., Turner, W.C., Kennedy, W.J., 2006. Guide to Energy Management. The Fairmont Press, Inc.
Chatterjee, A., Zhang, L., Xia, X., 2015. Optimization of mine ventilation fan speeds according to ventilation on demand and time of use tariff. Appl. Energy 146, 65–73.
Chou, M., Xia, X., 2007. Optimal cruise control of heavy-haul trains equipped with electronically controlled pneumatic brake systems. Control Eng. Pract. 15, 511–519.
Cobbinah, P.B., Erdiaw-Kwasie, M.O., Amoateng, P., 2015. Africas urbanisation: implications for sustainable development. Cities 47, 62–72.
Elaiw, A., Xia, X., Shehata, A., 2012. Application of model predictive control to optimal dynamic dispatch of generation with emission limitations. Electr. Power Syst. Res. 84, 31–44.
Fan, Y., Xia, X., 2017. A multi-objective optimization model for energy-efficiency building envelope retrofitting plan with rooftop pv system installation and maintenance. Appl. Energy 189, 327–335.
Farreny, R., Morales-Pinzón, T., Guisasola, A., Taya, C., Rieradevall, J., Gabarrell, X., 2011. Roof selection for rainwater harvesting: quantity and quality assessments in Spain. Water Res. 45, 3245–3254.
Fountoulakis, M., Markakis, N., Petousi, I., Manios, T., 2016. Single house on-site grey water treatment using a submerged membrane bioreactor for toilet flushing. Sci. Total Environ. 551, 706–711.
Ghisi, E., de Oliveira, S.M., 2007. Potential for potable water savings by combining the use of rainwater and greywater in houses in southern Brazil. Build. Environ. 42, 1731–1742.
Hoy, L.H., 2009. A Proactive Water Supply Shortage Response Plan Focusing on the Green Industry in the Rand Water Supply Area. Ph.D. thesis. University of South Africa.
Ilemobade, A., Olanrewaju, O., Griffioen, M., 2013. Greywater reuse for toilet flushing at a university academic and residential building. Water sa. 39, 351–360.
Jabornig, S., 2014. Overview and feasibility of advanced grey water treatment systems for single households. Urban Water J. 11, 361–369.
Kahinda, J.M., Taigbenu, A., 2011. Rainwater harvesting in South Africa: challenges and opportunities. Phys. Chem. Earth, Parts A/B/C 36, 968–976.
Kim, R.H., Lee, S., Jeong, J., Lee, J.H., Kim, Y.K., 2007. Reuse of greywater and rainwater using fiber filter media and metal membrane. Desalination 202, 326–332.
Kirk, D.E., 2012. Optimal Control Theory: an Introduction. Dover Publications Inc., New York, U.S.A.
Kozák, Š., 2016. From PID to MPC: control engineering methods development and applications, in: cybernetics Informatics (K& I), Levoča. Slovakia 1–7.
Lam, L., Kurisu, K., Hanaki, K., 2015. Comparative environmental impacts of source-separation systems for domestic wastewater management in rural China. J. Clean. Prod. 104, 185–198.
Li, Z., Boyle, F., Reynolds, A., 2010. Rainwater harvesting and greywater treatment systems for domestic application in Ireland. Desalination 260, 1–8.
Lipták, B.G., 2005. Instrument Engineers' Handbook: Process Control and Optimization, vol. 2. CRC Press.
Loux, J., Winer-Skonovd, R., Gellerman, E., 2012. Evaluation of combined rainwater and greywater systems for multiple development types in mediterranean climates. J. Water Sustain. 2, 55–77.
Luo, J., Huang, W., Zhang, S., 2015. Energy cost optimal operation of belt conveyors using model predictive control methodology. J. Clean. Prod. 105, 196–205.
Malatji, E.M., Zhang, J., Xia, X., 2013. A multiple objective optimisation model for building energy efficiency investment decision. Energy Build. 61, 81–87.
Mathaba, T., Xia, X., 2015. A parametric energy model for energy management of long belt conveyors. Energies 8, 13590–13608.
Mathaba, T., Xia, X., Zhang, J., 2014. Analysing the economic benefit of electricity price forecast in industrial load scheduling. Electr. Power Syst. Res. 116, 158–165.
Mayne, D.Q., Rawlings, J.B., Rao, C.V., Scokaert, P.O., 2000. Constrained model predictive control: stability and optimality. Automatica 36, 789–814.
Middelberg, A., Zhang, J., Xia, X., 2009. An optimal control model for load shifting—with application in the energy management of a colliery. Appl. Energy 86, 1266–1273.
Morales-Pinzón, T., Rieradevall, J., Gasol, C.M., Gabarrell, X., 2015. Modelling for economic cost and environmental analysis of rainwater harvesting systems. J. Clean. Prod. 87, 613–626.
Ntsaluba, S., Zhu, B., Xia, X., 2016. Optimal flow control of a forced circulation solar water heating system with energy storage units and connecting pipes. Renew. Energy 89, 108–124.
Numbi, B., Xia, X., 2015. Systems optimization model for energy management of a parallel HPGR crushing process. Appl. Energy 149, 133–147.
Numbi, B., Xia, X., 2016. Optimal energy control of a crushing process based on vertical shaft impactor. Appl. Energy 162, 1653–1661.
Odhiambo, N.M., 2009. Electricity consumption and economic growth in South Africa: a trivariate causality test. Energy Econ. 31, 635–640.
Qi, R., Mei, H., Chen, C., Qian, F., 2015. A fast MPC algorithm for reducing computation burden of MIMO. Chin. J. Chem. Eng. 23, 2087–2091.
Roy, S.S., Rouault, M., 2013. Spatial patterns of seasonal scale trends in extreme hourly precipitation in South Africa. Appl. Geogr. 39, 151–157.
Rozos, E., Makropoulos, C., Butler, D., 2009. Design robustness of local water-recycling schemes. J. Water Resour. Plan. Manag. 136, 531–538.
Setlhaolo, D., Xia, X., 2015. Optimal scheduling of household appliances with a battery storage system and coordination. Energy Build. 94, 61–70.
Setlhaolo, D., Xia, X., 2016. Combined residential demand side management strategies with coordination and economic analysis. Int. J. Electr. Power & Energy Syst. 79, 150–160.
Sichilalu, S., Mathaba, T., Xia, X., 2017. Optimal control of a wind-pv-hybrid powered heat pump water heater. Appl. Energy 185, 1173–1184.
Sichilalu, S.M., Xia, X., 2015. Optimal power dispatch of a grid tied-battery-photovoltaic system supplying heat pump water heaters. Energy Convers. Manag. 102, 81–91.
Stec, A., Kordana, S., Słyś, D., 2017. Analysing the financial efficiency of use of water and energy saving systems in single-family homes. J. Clean. Prod. 151, 193–205.
Stillwell, A.S., King, C.W., Webber, M.E., Duncan, I.J., Hardberger, A., 2010. The energy-water nexus in Texas. Ecol. Soc. 1–20.
Tazvinga, H., Xia, X., Zhang, J., 2013. Minimum cost solution of photovoltaic–diesel–battery hybrid power systems for remote consumers. Sol. Energy 96, 292–299.
Tazvinga, H., Zhu, B., Xia, X., 2014. Energy dispatch strategy for a photovoltaic–wind–diesel–battery hybrid power system. Sol. Energy 108, 412–420.
Tazvinga, H., Zhu, B., Xia, X., 2015. Optimal power flow management for distributed energy resources with batteries. Energy Convers. Manag. 102, 104–110.
Tóth, L.R., Nagy, L., Szeifert, F., 2012. Similarities of model predictive control and constrained direct inverse. Intelligent Control Automation 3, 278.
Vanek, F.M., Albright, L.D., Angenent, L.T., 2012. Energy systems engineering: evaluation and implementation. McGraw-Hill Prof. chapter Econ. Anal. Energy Proj. Syst. 120–190.
Wang, B., Xia, X., Zhang, J., 2014. A multi-objective optimization model for the life-cycle cost analysis and retrofitting planning of buildings. Energy Build. 77, 227–235.
Wang, L., 2009. Model Predictive Control System Design and Implementation Using MATLAB®. Springer Science & Business Media.
Wanjiru, E., Sichilalu, S., Xia, X., 2017a. Optimal operation of integrated heat pump-instant water heaters with renewable energy. Energy Procedia 105, 2151–2156.
Wanjiru, E., Xia, X., 2017. Optimal energy-water management in urban residential buildings through grey water recycling. Sustain. Cities Soc. 32, 654–668.
Wanjiru, E.M., Sichilalu, S.M., Xia, X., 2016a. Optimal integrated diesel grid-renewable energy system for hot water devices. Energy Procedia 103, 117–122.
Wanjiru, E.M., Sichilalu, S.M., Xia, X., 2017b. Model predictive control of heat pump water heater-instantaneous shower powered with integrated renewable-grid energy systems. Appl. Energy. https://doi.org/10.1016/j.apenergy.2017.05.033.
Wanjiru, E.M., Sichilalu, S.M., Xia, X., 2017c. Optimal control of heat pump water heater-instantaneous shower using integrated renewable-grid energy systems. Appl. Energy 201, 332–342.
Wanjiru, E.M., Xia, X., 2015. Energy-water optimization model incorporating rooftop water harvesting for lawn irrigation. Appl. Energy 160, 521–531.
Wanjiru, E.M., Zhang, L., Xia, X., 2016b. Model predictive control strategy of energy-water management in urban households. Appl. Energy 179, 821–831.
Xia, X., Zhang, J., 2015. Operation efficiency optimisation modelling and application of model predictive control. IEEE/CAA J. Automatica Sinica 2, 166–172.
Xia, X., Zhang, L., 2016. Industrial energy systems in view of energy efficiency and operation control. Annu. Rev. Control 42, 299–308.
Zhang, D., Gersberg, R.M., Wilhelm, C., Voigt, M., 2009. Decentralized water management: rainwater harvesting and greywater reuse in an urban area of Beijing, China. Urban Water J. 6, 375–385.
Zhang, H., Xia, X., Zhang, J., 2012. Optimal sizing and operation of pumping systems to achieve energy efficiency and load shifting. Electr. Power Syst. Res. 86, 41–50.
Zhang, J., Xia, X., 2011. A model predictive control approach to the periodic implementation of the solutions of the optimal dynamic resource allocation problem. Automatica 47, 358–362.
Zhang, L., Xia, X., Zhang, J., 2014. Improving energy efficiency of cyclone circuits in coal beneficiation plants by pump-storage systems. Appl. Energy 119, 306–313.
Zhuan, X., Xia, X., 2013. Optimal operation scheduling of a pumping station with multiple pumps. Appl. Energy 104, 250–257.

Renewable and Sustainable Energy Reviews 82 (2018) 2791–2805

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Measurement uncertainty in energy monitoring: Present state of the art

Herman Carstens[a,*], Xiaohua Xia[a], Sarma Yadavalli[b]

[a] Centre for New Energy Systems (CNES), Department of Electrical, Electronic, and Computer Engineering, University of Pretoria, South Africa
[b] Department of Industrial and Systems Engineering, University of Pretoria, South Africa

ARTICLE INFO

Keywords:
Measurement and verification
Energy metering
Project risk
Measurement error models
Calibration
Metrology
Mismeasurement
Errors-in-variables

ABSTRACT

Measurement uncertainty is a key component in the overall uncertainty calculation for Measurement and Verification (M & V) projects. However, in some cases, it is reduced to outlier detection or basic uncertainty propagation calculations. In other cases, funds are spent on determining uncertainties that have little effect on project decisions. Therefore a need exists for a fuller treatment of the subject in the light of literature from M & V and other fields. This paper surveys general M & V literature, as well as relevant research from metrology, electrical engineering, economics, decision analysis, and statistics. Electrical metering and sub-metering uncertainty is investigated, as well as often-overlooked considerations such as power quality and the cost of calibration. The effect of mismeasurement on energy models and practical techniques for mitigating such effects are assessed. Last, research on building simulation and project decisions in the light of measurement error is surveyed. Bayesian methods are found to be a recurring theme in much of the research being conducted on all of these aspects. Power quality and mismeasurement effects have also been found to make a material difference in project evaluation. The survey is concluded with recommendations for further research in the light of current trends in data analysis and energy evaluation.

1. Introduction

The International Performance Measurement and Verification Protocol (IPMVP) [1] notes that three forms of uncertainty arise in energy Measurement and Verification (M & V): measurement uncertainty, sampling uncertainty, and modelling uncertainty [1]. Although research on combining sampling and modelling uncertainty has been done by Ye et al. [2,3] and Carstens et al. [4] on lighting projects, and Sun on building energy performance [5], measurement uncertainty is usually assumed to be negligible. Nevertheless, the cost-effective allocation of measurement resources continues to be a pertinent question for decision makers. The aim of this survey is to introduce M & V professionals and researchers to the salient literature on various topics related to measurement uncertainty in energy monitoring.

While one usually associates measurement in M & V with electricity meters, instruments measuring with error also include surveys and questionnaires [6], tracking databases, non-intrusive load monitoring, and inspection reports [7]. These instruments may measure or record any number of variables such as occupancy [8], floor area, schedules, income, the proportion of Miscellaneous Electrical Loads (MELs) [9,10], etc. Sometimes data such as plug load energy use are used as a proxy to measure occupancy [11]. More about this in Section 3.5.

Are cheaper, smarter meters and the big data revolution not going to render measurement uncertainty concerns obsolete? Advanced Metering Infrastructure (AMI) is being rolled out in the United Kingdom (UK) and Europe, although state regulation is more fractured in the US [12]. Although these regions represent only 12.4% of the world population, they consume 66.2% of the world's electricity.[1] The nature of M & V in these regions is changing, with promising results for M & V 2.0 already being published [13]. On the other hand (or hemisphere), 17% of the world population still have no access to electricity, and 38% still cook using biomass [14]. Many of these live in sub-Saharan Africa, and

Abbreviation: AMI, Advanced Metering Infrastructure; ANSI, American National Standards Institute; ASHRAE, American Society of Heating, Refrigeration, and Air Conditioning Engineers; CCC, California Commissioning Collective; CDM, Clean Development Mechanism; CFL, Compact Fluorescent Lamp; DMM, Digital Multimeter; CT, Current Transformer; DAQ, Data Acquisition (board); ECM, Energy Conservation Measure; EPC, Energy Performance Contract; ESCO, Energy Services COmpany; EV, Expected Value; G14, ASHRAE Guideline 14-2002 and 14-2014; GP, Gaussian Process; GUM, Guideline for the expression of Uncertainty in Measurement; HVAC, Heating, Ventilation, and Air Conditioning; IPMVP, International Performance Measurement and Verification Protocol; IEC, International Electrotechnical Commission; IEEE, Institute for Electrical and Electronic Engineers; ISO, International Standard Organization; MC, Monte Carlo; MEL, Miscellaneous Electrical Load; MID, Measurement Instrument Directive; MCMC, Markov Chain Monte Carlo; M & V, Measurement and Verification; MEM, Measurement Error Model; MLE, Maximum Likelihood Estimation; OLS, Ordinary Least Squares; NREL, National Renewable Energy Laboratory; PDF, Probability Density Function; SEE Action, State and Local Energy Efficiency Action group; SEM, Stick-on Electricity Meter; TUR, Test Uncertainty Ratio; UMP, Uniform Methods Project; UUT, Unit Under Test

* Corresponding author.
E-mail address: hermancarstens@gmail.com (H. Carstens).

[1] World Bank Open Data Portal. URL: beta.data.worldbank.org.

http://dx.doi.org/10.1016/j.rser.2017.10.006
Received 6 September 2016; Received in revised form 16 August 2017; Accepted 3 October 2017
Available online 13 October 2017
1364-0321/

for the companies serving these billion people, the big data revolution is still some way off.

We should also note that AMI improves sampling rather than measurement uncertainties. Even so, investigations into big data in energy monitoring [13,15,16] are welcome, although bigger data are no remedy if it is still measured with error. Although the tools and methods are improving and becoming automated, measurement error will continue to be relevant to M & V professionals. However, it does not seem to be discussed directly in most M & V literature, and we hope that this work goes some way in addressing this gap.

This survey is structured around the following questions:

- What does current literature say about measurement uncertainty? How is it addressed in metrology?
- What are the sources of electrical metering uncertainty? What are the effects of mismeasurement, has it been documented in energy monitoring, and how can it be mitigated?
- How does measurement uncertainty affect project decisions?

2. Background

2.1. Measurement uncertainty in M & V literature

Measurement uncertainty is acknowledged in M & V literature, although firm guidance is seldom given. A summary of guideline characteristics in this respect can be found in Table 1. The American Society of Heating, Refrigeration and Air Conditioning Engineers (ASHRAE's) Guideline 14-2002 [17,18] (henceforth referred to as G14) is the foremost technical resource for M & V. It provides comprehensive guidance on instrumentation, data-handling, uncertainty calculations, as well as a catalogue of uncertainties for a wide variety of energy-related measurement instruments. It has recently been updated to a 2014 version [19], although the original remains useful. ‘G14’ will refer to both, unless stated otherwise. G14 and the California Commissioning Collective [20] (CCC) adopt Reddy and Claridge's alternative fractional-savings parametrisation of measurement uncertainty [21]. The IPMVP [1,22] provides general guidance on uncertainty but does not address measurement uncertainty in much detail. The National Renewable Energy Laboratory (NREL's) Uniform Methods Project (UMP) [23] establishes best practices for energy data collection and is the only guideline to discuss mismeasurement at all. ASHRAE Guideline RA96: *Engineering Analysis of Experimental Data* [24] also deserves mention. It is a general quantitative introduction to handling measurement uncertainty in engineering measurements and could be applied to some M & V cases. The State and Local Energy Efficiency (SEE) Action group's *Energy Efficiency Programme Impact Evaluation Guide* [25] (hereafter referred to as the SEE Action Guide) is also notable and does give practical guidance on uncertainty. Finally, some preliminary work on the relative contributions of measurement and sampling uncertainty in M & V has also been presented by Carstens, Xia, and Yadavalli [26], and a method for low cost calibration of energy meters proposed [27]. Recently, Ligier et al. [28] proposed a method for accounting for M & V uncertainty alongside building simulation, and did consider measurement uncertainty in the model.

Greenhouse Gas reduction programmes often require M & V. Vine et al. reported on different options considered for dealing with measurement uncertainty in such cases [29]. Although this was a work in progress in 2002, it is still relevant, since the debate around the advantages and disadvantages of different measurement approaches is explained well. Discount factors to compensate for the uncertainty of various methods are also listed. The scale of the United Nations Framework Convention for Climate Change's Clean Development Mechanism (UNFCCC CDM) methodology specifications dwarfs other M & V documentation. It contains over two hundred methodologies for different project scales and applications. Accuracy requirements vary, but the 90/10 criterion is most common, although Sonnenblick and Eto [30] have shown that this precision level is only necessary for projects where the savings to cost ratio to be verified is small. In many cases, 90/50 is adequate for identifying project cost-effectiveness, that is, whether or not a project saved energy.

Shishlov and Belassen [31] provided a useful review of how monitoring uncertainty is approached in the CDM. For example, CDM AM0046 requires Compact Fluorescent Lamp Retrofit programmes to be monitored very stringently at the insistence of regulators, even requiring custom-made meters. Michaelowa, Hayashi, and Marr [32] who developed the methodology noted that no projects were completed under AM0046 until the alternative AMS II.C [33] was adopted. Later AMS II.J [34] was also adopted. In it, every CFL is deemed to operate for 3.5 h/day, eliminating the need for measurement. Even so, they assert that there are still projects that would reduce emissions but are ineligible. These difficulties illustrate that measurement goals should always be construed in the larger project and social context. Achieving

Table 1
The treatment of measurement uncertainty in leading M & V guidelines.

Name	Year	Level of detail	Features	Reference
G14	2002, 2014	10	• Most comprehensive treatment of M & V uncertainty • Excellent methods • Instrument uncertainty database • Itemized measurement costs • Technology slightly dated in 2002 version	[18,19]
IPMVP	2012	5	• Introductory treatment • Sensitivity and Uncertainty Analysis worked examples	[1,22] [1]
CDM	2015	8	• Approach varies between methodologies • Emphasis on being conservative • Discount factors used for >95/5 measurement error • 95/10 assumed for unknown measurement error • Deemed Savings also used • MC recommended for complex cases	[41,31] [32] [35] [35] [34]
UMP	2014	6	• Varies with authors of chapters • Errors-in-variables discussed in Chapters 13, 23 • Metering error discussed in Chapter 9 • Survey error discussed in Chapter 11	[23] [43,44] [45] [46]
SEE Action Guide	2012	4	• Practical guidance • Discussion of uncertainty and project risk	[25]
CCC	2012	6	• Appendix on uncertainty analysis • Adopts and simplifies fractional savings approach	[20]

Abbreviations: CCC, California Commissioning Collective; CDM, Clean Development Mechanism; G14, ASHRAE Guideline 14-2002 and 14-2014; IPMVP, International Performance Measurement and Verification Protocol; SEE Action Guide: State and Local Energy Efficiency Programme Impact Evaluation Guide; UMP, Uniform Methods Project.

important individual statistical outcomes is never an end in itself. It may even hinder meeting overarching programme goals such as emissions reduction or development. Research on efficient sampling designs has been conducted to reduce the sampling burden as much as possible [2–4], although this is still much scope in this field. The CDM board is also working towards a stringency/cost trade-off system to replace the current system [35]. We discuss such approaches in Section 6.

2.2. Measurement uncertainty in metrology

Metrology is the science of measurement, and its guiding document is the ISO Guide to the Expression of Uncertainty in Measurement, also known as the GUM [36]. The GUM has standardised the expression of uncertainty across most quantitative scientific disciplines and is also applied to energy monitoring. Instructive tutorials have been written, most notably by the British [37,38] and European [39] accreditation agencies. ISO/IEC 17025 [40] *General requirements for the competence of testing and calibration laboratories* has contributed to the GUM's popularity by stipulating that complying laboratories apply a procedure to estimate uncertainty in measurement.

The GUM distinguishes between measurement uncertainty calculated by statistical methods from measured data (Type A), and those measured or stipulated from prior information or judgement (Type B). It also standardised the expression of uncertainty as a *coverage interval*, also known as an *expanded uncertainty*. This is the confidence/precision format of expressing uncertainty, which should be familiar to most M & V professionals and is used in the IPMVP [1], RA96 [24], and CDM [41] documents. For example, when a measurement is expressed as 10 ± 1, the precision range (or *semi-range*) is $p = 1$. We expect the interval from nine to eleven to correspond to the 95% confidence interval if no more information is given [37,24]. Since the standard score of the normal distribution $z_{95\%} = 1.96 \approx 2$, we say that the *coverage factor* is 2. The rectangular/uniform distribution is recommended rather than the Normal distribution for digital volt meters and instruments where uncertainties are not stated [37]. Although this is conservative, it is not a realistic assumption for M & V. Energy data are usually aggregated or integrated over a time interval such as 30 min, and such errors would then be normally distributed. If an M & V practitioner opts for the uniform distribution assumption, and later convolves it with a normal distribution for sampling error, for example, the resultant coverage interval will be a statement about uncertainties, not probability density intervals [42]. We recommend Monte Carlo (MC) convolution to obtain the probability distribution in such a case.

M & V professionals should also be aware of the concept of dominant uncertainty components. As a rule of thumb, if one uncertainty component is two to three times larger than the next highest one, it may be considered to be the sole contributor to the overall uncertainty [37, p.17]. This is because of the sum-of-squares approach to adding standard deviations together allows larger standard deviations to dominate the final result. Commenting on the efficient allocation of measurement resources between Type A and Type B measurements, Birch, therefore, remarks that the "quantification of uncertainties in testing normally involves a large element of estimation of…uncertainty components. Consequently, it is seldom justifiable to expend undue effort in attempting to be precise in the evaluation of uncertainty for testing" [37, p.15].

2.2.1. New directions in metrology

Although acknowledged as very helpful, the GUM has drawn criticism, most notably from Bayesian statisticians [42]. One point of contention relevant to M & V is that the propagation of errors calculation is defined as a first-order Taylor series approximation, which does not always hold.

Some physicists and statisticians are also uncomfortable with the frequentist approach to how confidence intervals are calculated in the GUM. It has been shown from first principles that this approach is invalid in many measurement cases [42]. The standard (frequentist) confidence interval, for example '90%', is a product of a process that produces an interval containing the true value 90% of the time [47]. It is *not* an expression of certainty or degree of belief, as 10% of the time the interval will not contain the true value at all. The Bayesian credible interval can claim this, however. For many cases the distinction is academic, as these intervals may agree [48], frequentists may borrow Bayesian language [42]. For other situations, however, standard confidence intervals are inappropriate for risk calculations, and credible intervals are recommended.

In reaction to the criticisms above, the GUM was updated, and a supplement describing a Monte Carlo (MC) alternative was published [49]. It is especially useful for non-linear cases, where any distribution other than the Gaussian or scaled-and-shifted T is used, or where the error propagation function is complex. It also delivers the final error estimation as a probability distribution rather than an uncertainty interval. Therefore it is all but recommended as the de facto method for uncertainty propagation calculation by the supplement. MC can be too computationally expensive for high-dimensional problems and approaches such as MC-Latin Hypercube Sampling or Sobol' Sequences [50]. Respected Bayesian metrologists such as Lira have advocated analytical calculus-based approaches over MC methods where possible [51]. However, we do not see this as a viable alternative in the energy M & V industry.

A second, useful approach is the Mellin Transform Moment Calculation (MTMC) method [52,53], which has a free online toolbox for calculation [54]. The method has been developed as an analytical alternative to MC and allows the moments of a distribution resulting from a polynomial function of constituent distributions to be expressed exactly. Once mean, variance, skewness, kurtosis, and higher order moments are obtained, these can be used to calculate the shape of the resultant distribution in a more computationally efficient and consistent manner than MC. This has been used in M & V [55] by fitting a Johnson S_B [56] distribution using Hill's algorithm [57]. Rajan et. al [58] provided more information on moment-based distribution fitting.

Regarding the Bayesian approach, the UK Accreditation Service (UKAS) noted that "Bayesian statistics is becoming recognised as being particularly useful in certain areas of testing" [38], and as of 2016 the GUM itself is also in the process of being extensively revised to accommodate the Bayesian paradigm [59]. This signals an interesting shift in metrology and the way in which uncertainty is viewed and calculated, and M & V professionals would do well to take notice. (Some already have, as will be seen in Section 6.) For those seeking an introduction, Estler [60] provides a comprehensive tutorial of Bayesian theory in the context of measurement and the GUM, while shorter theoretical Bayesian frameworks for metrology have also been written [61,62]. Although M & V practitioners should be cognizant of the underlying theory at the level presented in these papers, the specific mathematics in these sources are replaced by MC methods implemented in software. Rossi developed domain-specific MC software for calculating measurement error by Bayesian methods [63], although general-purpose software may be preferable by most M & V professionals, as discussed in Section 5.2.2.

In a recent study, Carstens et. al. [27] used a Simulation Extrapolation (SIMEX, see Section 5.2.2) [64] method enhanced by a Bayesian approach to calibrate energy meters in-situ while controlling for uncertainty.

3. Metering uncertainty

Metering uncertainty can be dominated by other uncertainties such as sampling or modelling [26], but can nonetheless be significant depending on the application. Below we will consider five cases: general energy metering uncertainty, sub-metering and its contribution to measurement uncertainty, how power quality affects metering uncertainty, virtual instrumentation, and the possibility of in-situ meter calibration.

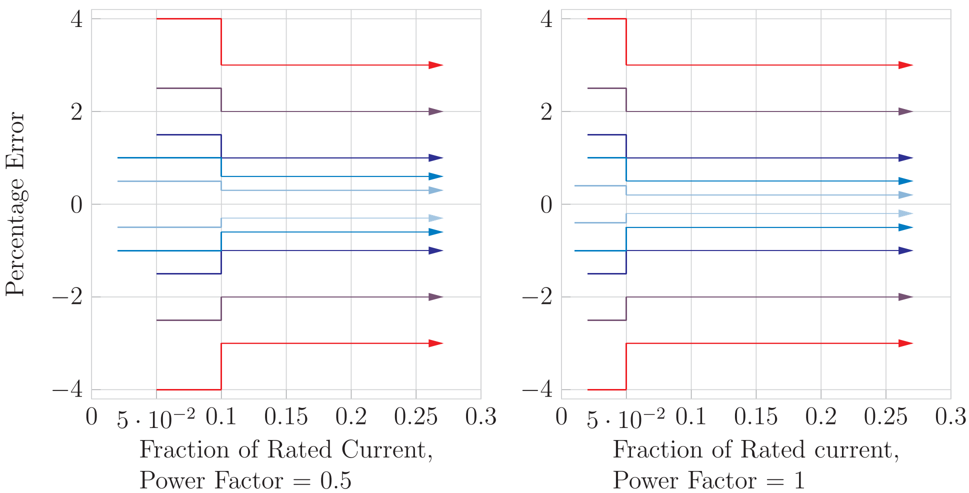


Fig. 1. Comparison of different IEC accuracy class meters [66–68] for transformer-connected single or polyphase meters with balanced loads under sinusoidal conditions.

Regarding metering uncertainty, static (solid-state) electrical energy meters used for reporting purposes have to be qualified to standards set by the International Electro-technical Commission (IEC), or its equivalents, such as ANSI C 12–20 [65] in the US. Metering classes indicate maximum allowable percentage errors over the majority of the measurement range, so that a Class 1 m is 1% accurate, for example. IEC 62053-21 [66] refers to Class 1 and 2 (active), 62053-22 [67] to class 0.2S and 0.5S (active), and 62053-23 [68] to class 2 and 3 (reactive) meters. A graphic illustration of the accuracy requirements is shown in Fig. 1. Close attention should be paid when acquiring meters, as accuracy class (mis)specification has also been abused as a marketing tool, as catalogued by Irwin [69]. M & V professionals should also note that influence quantities such as harmonics are tested for, but in a one-at-a-time fashion, with all other quantities held at default levels.

A distinction between calibrated and qualified meters should be drawn at this point. A qualified meter model range conforms to the IEC standards (called 'type conformity' by the European Measurement Instrument Directive (MID) [71]). Models of this type have undergone many different tests to prove that their results are stable within certain specified operating ranges for factors such as temperature, power factor, and humidity. Thus qualification is a matter of the *quality* of a given meter model. An individual meter, although qualified as a unit in a model range, may still give incorrect readings because it is not *calibrated*. This may occur when internal conversion factors have drifted over time, for example.

Even when meters are qualified to these standards, errors or bias can be introduced by environmental conditions. For example, even though temperatures in Saudi Arabia still fall within IEC specifications, systematic bias is introduced due to consistently abnormal values [72]. Even such small biases on revenue meters metering large installations can lead to significant billing errors.

The discussion above applies to the meter itself, but not to the current transformer (CT) often used to measure the current. In many cases, CT accuracies are lower than the meter accuracies. An example of CT accuracy specifications can be found in Fig. 2. They need to be considered separately from metering uncertainty and added using the sum-squared error method. In many situations, the accuracy class of the CT and meter, together with their rated currents will suffice to determine the overall accuracy of the measuring system.

Although accuracy influences meter prices, the communication protocol used by the meter is also significant, as shown by Ahmad et al. in their review of energy and related sensors [73].

3.1. Sub-metering

Sub-metering an installation often provides valuable insight into the main load drivers but can be expensive if revenue-accuracy meters are used. One can consider less accurate and costly options in these applications.

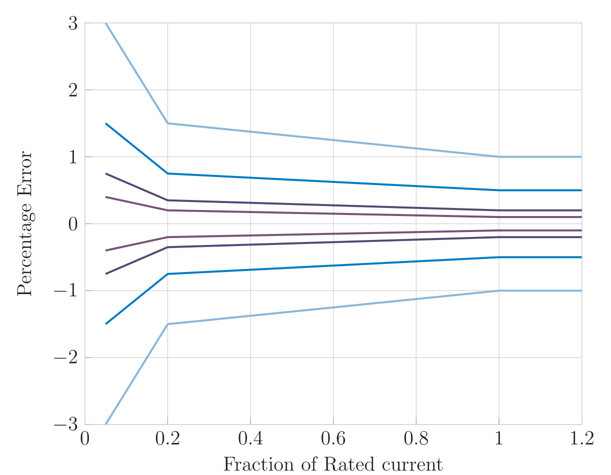


Fig. 2. Instrument Current Transformer accuracies according to IEC 60044-8 [70]. For Class 3 and Class 5, the limits are flat at 3% and 5% respectively.

Plug-through meters are popular for metering MELs. Polese et al. provided a comprehensive case study detailing the challenges in implementing such a solution at a large retailer, for an NREL study [74]. The study demonstrates the inaccuracy of such meters, as well as other factors that contribute to general measurement uncertainty. In this study, 41% of the meters had significant portions of the data series that were erroneous. Errors of 20% in the range 0–20 W were common, and 6% in the range 25–100 W. Given that 40% of the MELs operated below the 60 W level, these errors are significant.

Stick-on Electricity Meters (SEMs) represent an exciting new low-cost measurement or logging option [75]. These sensors are placed on the circuit breaker in the distribution board, and senses when current is drawn on the circuit. Tests indicate an accuracy of 5% or less. It is important to note that these do not work where relays are present.

Current-only meters are becoming a popular option for residential metering. They usually use split-core CTs and are much more affordable than revenue energy meters, but are not as accurate, or even qualified. In personal correspondence with a popular meter manufacturer based in the UK, the accuracy was quoted as 10% [76]. Given that they operate in a narrow environmental and electric range, this is usually not of great concern, provided that they can be verified in some way. However, they can not be recommended as the sole meters used for projects. The voltage may vary due to supply-side fluctuations, or due to facility-level demand factors. On the demand side, current-only meters multiply their readings by a nominal voltage. The resultant power measurement is in Volt-Ampéres: apparent power, not true power in Watts. The power factor is thus assumed to be unity. Inductive power electronic equipment found in most households will decrease the power factor to below one, biasing the measurement by this power factor. On the supply side, the utility voltage is seldom at the nominal level. It is regulated to be in a certain range [77]. In Europe, utility supply voltage is determined to be 230 V ± 10% [78], and in the United States, 120 V ± 5% [65]. However, certain asymmetrical tolerances may also hold. For ANSI C84.1 Range B [79], these tolerances are − 13% and + 6%. These asymmetrical tolerances may skew the calculation since under-voltages are higher and possibly more likely than over-voltages.

For the symmetrical tolerance case, it may be argued that unmeasured variations would cancel out over time. However, a constant voltage offset may also apply. The supply voltage at a facility such as a house varies with a number of factors. These include the distance of its distribution transformer from the substation on the primary feeder, the distance between this house and the transformer on the secondary feeder, the number of facilities on the secondary feeder, and the load on the feeders. The average incomer voltage at a house on the edge of a distribution network may be at the lower end of the specification interval, while a facility closer to a transformer may be at the upper end of the interval. Therefore the distribution of voltage for a single facility may not be symmetric around the country's nominal voltage, biasing the measurements for which a nominal voltage was specified.

3.2. Power quality

Power Quality is an important consideration in metering uncertainty calculation, although M & V does not discuss it very much. The IEC standards qualify meters only for sinusoidal conditions, but on networks with modern power electronic equipment, this assumption is usually invalid [80]. The harmonics which cause the non-sinusoidal condition may originate from some modern power electronics sources, such as Variable Speed Drives (VSDs), fluorescent lamps with electronics ballasts, switching power supplies, or controlled rectifiers [81]. These harmonics are generated by loads on the network but are observed as a supply quality problem when measured. For certain cases where the customer pollutes the power network with large harmonic power flows, the presence of harmonics may skew the reactive energy measurement to such an extent that a power factor greater than unity is indicated, even if this is not the case at all [81].

These conditions then lead to mismeasurement in static energy meters, especially when a non-unity power factor is present [82], and verification of meters for such cases have been proposed [83,84]. We note that this does not apply to older electromechanical induction meters, but only to solid-state (static) smart meters [85]. Berrisford provides an accessible and practical introduction to this problem [86]. Literature reviews of this field have been conducted [87] and updated [88], and readers are encouraged to consult them for more technical details, as we will focus on the M & V implications.

The problem with measuring non-sinusoidal loads is that reactive power is calculated and defined in numerous ways [84]. Although the different formulas give the same result under sinusoidal conditions, they differ when harmonics are present. Current magnitude and power factor are the main uncertainty drivers [81]. An example of this inaccuracy has been documented in the field [86]: an approved Canadian meter using Budeanu's power definition [89] was replaced by an approved Canadian meter using Fryze's power definition [90]. This resulted in a power factor penalty being added to the customer's bill when the meter was changed, even though the energy use did not change. Further investigation revealed non-sinusoidal conditions due to the harmonics generated by the client's VSDs, which the meters measured in different ways. We wonder whether some of the inaccuracy noted by Polese et al. [74] in their metering of a retailer with many MELs may not be due to such effects.

Because of these different definitions and different calculation methodologies among different meters, Cataliotti et al. [91,92] recommend that when calibrating a meter in-situ, a reference meter implementing the same metric as the Unit Under Test (UUT) should be selected, so as not to compound the errors. If the manufacturer does not state the metric used, methods for determining it experimentally have been devised. However, it was found that in such a case, the UUT only adheres to the accuracy limits set in the standard when compared with the reference meter adopting the same power definition, not with the true energy value.

There is, however, a course charted through the reactive power-definition confusion. The IEEE Standard 1459 (2010) [93] gives guidance on how reactive power should be defined and calculated. The consensus among most of the papers cited here is that this definition should be adopted. It is also endorsed by the IEC. Berrisford has demonstrated that reprogramming certain kinds of digital watt meters in minor ways can lead to calculation according to the IEEE 1459 definition [86]. Although utilities do not itemise harmonic distortion on the bill, preliminary work has been done to prepare the way for future considerations [94,95].

We recommended that M & V professionals use meters measuring so-called ‘fundamental’ quantities, from which to calculate the true reactive power according to the IEEE 1459. Meters with sampling rates adequate for including relevant harmonics should be selected, although increasing the sampling rate increases the price of the meter significantly in the range 0–80 μs [96].

3.3. Analog to Digital Conversion (ADC) and virtual instrument measurement uncertainties

Most modern static meters employ ADC (also known as Digital Signal Processing). ADC is also used in Virtual Instrumentation (VI), where a transducer is connected to a personal computer via a Data Acquisition (DAQ) board, for user-built DSP software to process [97]. Note that VIs can measure any analog signal on which to perform ADC and that the general uncertainty principles remain the same. This field shows great promise for lower cost calibration and measurement of electrical signals for M & V purposes.

ADC technology is useful in electrical measurements as it has the potential for measuring true reactive non-sinusoidal power accurately, as discussed in Section 3.2. However, various standards specifying different parameters for ADC exist. Spataro [98] notes that ADC uncertainty has been quantified by the ISO GUM uncertainty propagation law (through a Fast Fourier Transform) [99], random-fuzzy variables [100], and MC approaches [97]. Due to the difficulty of convolving different uncertainty distributions analytically, such numerical methods make sense. These require any number of different variables, depending on the standard and method employed. Spataro identifies that only offset (bias), gain, Total Harmonic Distortion (THD), spurious tones, and the Signal to Noise Ratio (SNR) are needed to quantify power quality. The details of such errors depend on the electronic components of the DAQ itself, but such systems can reach standard-level accuracies at a fraction of the cost [101]. They are thus expected to increase in popularity as they become commercialised [99]. In any event, the uncertainties introduced by ADC is usually much smaller than those of the transducers themselves [97]. The most recent results in this field comprise a detailed theoretical model with experimental results for a DAQ-based sampling watt meter, based on the definitions set out in IEEE 1459 [88].

3.4. In-situ meter calibration

Due to the MID ratified by the European parliament in 2004 [71], European meters (gas, water, electricity, etc.) need to be calibrated under actual conditions, interpreted as the actual meter installation location [102]. This has lead to various studies of how such a calibration may be achieved. Femine et al. [102] have devised a scheme for a field laboratory with a travelling standard. Power generated by the laboratory then allows a set of tests to be conducted at the facility. The directive has been viewed as impractical since not all plants can be shut down for such a procedure, metering cost increases drastically with a call-out for a portable metrology laboratory, and man-hours needed to test all Italian meters twice-yearly is unrealistic [103]. To offset this burden, Amicone et al. proposed a low cost, stable, 'add-on' calibrator that can be activated twice yearly to perform the necessary calibration [103]. Crenna et al. [104] considered the MID as a step toward the modernization of legal metrology. They considered water meters and proposed an MC approach based on statistical metrology and risk techniques, similar to Pendrill and Källgren's work on CO_2 meters [105] discussed in Section 6. This seems by far to be the simplest and most affordable proposal, although it relies on large quantities of manufacturer data and does not address all the concerns raised by the other authors. Meter ageing and water temperature are considered as influence factors similar to power factor and harmonic distortion for energy meters, although the analogy is not close enough to use the method as-is in electric applications.

Measurement accuracy and its place in the smart grid are being investigated [106] and was proposed in rudimentary form a decade ago [107]. As smart meters become more common and interconnected, network cross-calibration to relieve the burden of calibrating every single meter may become a possibility, and represents an opportunity for future research.

3.5. Measurement uncertainty for non-electrical parameters

Often, non-electrical variables are also included in the energy model. Table 2 details typical errors for such cases. This is especially common when whole-facility regression models are constructed using measurements of variables such as temperature [108], occupancy [11,8] or flow rate [104]. Besides the error in the meter itself, poor meter selection, placement, or misestimation of independent variables may also contribute to unquantifiable errors in this case [22]. For example, the flow rate and temperature in a duct vary between the edge and the centre and features such as elbows impact flow and heat transfer characteristics for a non-negligible downstream portion of the duct. Because of these complex interactions, it is useful to work with general error estimates such as those found in G14 [18]. However, even these values should be used with caution. For example, CO_2 sensor accuracy was investigated [109] and the authors found that only seven of the eighteen sensors had errors of less than 20% at standard CO_2 levels for classrooms - a much higher value than that specified by G14.

Occupancy is a key factor in building energy use but is notoriously difficult to measure and model. Combinations of reed switches and passive infra-red (PIR) sensors seem to work well for offices [110], but these are very simple environments with single occupants per room. For more complex situations, proxies such as blind, fan, light, thermostat, door, or other sensors are used, although these are imperfect [111,112]. We note that recently Wang et al. [8] have shown in a sophisticated study that occupancy was not a significant energy use factor for their case study building. However, the building in question used a centrally controlled independent HVAC system, and this result is to be expected.

Occupancy models usually compare forecasts to data measured with error. However, as long as the measured variable predicts energy use well, the measurement error or true occupancy is not significant for energy models, unless occupant behaviour is being investigated.

4. Meter uncertainty as a component of M & V uncertainty

In South Africa, measured and verified energy savings achieved by businesses are eligible for tax deductions according to the 12L tax incentive [116]. However, measurement devices used for such projects need to be calibrated by accredited laboratories. This is a sound principle and has been adopted by many other agencies as listed by Ahmad et al. [73]. However, it greatly increases measurement costs, which can make M & V be prohibitively expensive and reduce the number of feasible projects significantly, as in the CDM case [32,31]. Given the small contribution to overall uncertainty made by electrical meters, especially when sampling is done [26], such requirements may be counter-productive. Overall accuracy requirements could be better served by spending the funds on obtaining a larger or more detailed sample, or measuring independent variables more accurately.

DAQ-based meter calibration discussed in Section 3.3 presents an interesting opportunity in this regard. We recognise that calibration is about more than having access to an accurate reference instrument and that quality and traceability procedures as set out in ISO 17025 [40] should also be in place. However, even energy meters calibrated to lower accuracies than the current classes should be sufficient for most M & V applications, where uncertainties are dominated by other factors (cf. Section 2.2).

One should also use these techniques when one measures independent explanatory variables such as temperature or occupancy with error. We now turn our attention to this topic.

Table 2
Instrument uncertainties for M & V Applications. Note that many of these values come from ASHRAE Guideline 14-2002 Appendix A5.6 [18], and are quoted at the 68% confidence level for this source. Guideline 14-2014 values are unchanged unless otherwise noted. Furthermore, Guideline 14-2014 stipulates these as minimum requirements, rather than typical values, but also recommends that they be used if no other values are available (Section 4.2.11.2). The confidence level for the other sources is unspecified or complex, and readers are referred to the original documents for more complete descriptions. FS denotes a percentage of full-scale.

Quantity	Type	Guideline 14	Other Source
Temperature	Ambient outdoor portable electronic	2–5%	
	Domestic water portable electronic	2%	
	Air ducts	5%	
	Pipes and ducts	2–5%	
Air velocity	Indoor: non-mechanical or blower door	5%	2–5% [73]
	Handheld anemometer	10%	
	Recording anemometer	5%	
	Meteorological grade anemometer	2%	
	Air ducts: array	2–5%	
Pressure	Gauge	0.25–2%	
	Ducts	1–5%	
	Pressurization/ depressurization	3–5%	
Energy	Electrical Energy meter	1%	0.2–0.5% [66–68]
	Current Transformer	2–3%	0.2–3% [70]
	Portable Watt meter	1–5%	
	Current: low cost home energy		>10% [76]
	Stick-on Meter		5% [75]
	Plug-through meter		20% [74]
	Relative humidity	2–5%	4.5% [73]
	Energy meter (gas)	1%	
Flow rate	Bucket and stopwatch, portable meter/probe	5%	< 1–5% [1]
	Domestic, accumulating	1–2%	
	HVAC inline or insertion meters	2%	< 1% [1]
	Ultrasonic, flare		2.5–5% [113]
	Smokestack gas		5–20% [114]
Run-time	Permanent	1–5%	
	Portable	2–5%	
Light	Sensor / logger		8–10% [73]
Other	Pyranometer	2–5%	>10% [115]
	Door position	2%	
	RPM	1%	
	CO_2		> 20% [109], 4% FS [73]
	Combustion	2%	~ 0.5% [105]

5. Mismeasurement

The measurement errors discussed thus far are mostly harmless. If random, mismeasurement of the *dependent* variable (usually energy) widens the confidence interval around the estimate but does not add bias to the parameter estimates. However, this is not the case when these noisy measurements are used as *independent* variables in a regression analysis. This *errors-in-variables* effect is seen in energy regression models when a covariate such as temperature or occupancy is measured with error, and may also occur when one calibrates an instrument against a standard with some error. In such cases, the random variation is no longer in $\mathbf{y}$, but in $\mathbf{x}$. Random errors in $\mathbf{x}$ have two effects. First, all the regression parameters become biased due to the "flattening out" of the data points as they spread out on the *x*-axis. This is called *attenuation*. Second, the confidence intervals on these estimates are narrower than they should be, giving misleadingly high confidence in biased values, also manifesting as a loss of statistical power [64]. This is because as the measurement error (variance) increases, it becomes increasingly difficult to distinguish it from the process variance. This lack of power may then be misinterpreted as a lack of effect when pre- and post-retrofit measurements are compared [64]. To regain this power, much larger sample sizes are then required. Table 3 summarises the effect of mismeasurement on various statistics, but we should note that effects vary with error type and regression model type.

To illustrate attenuation, consider attempting to use one unbiased meter to calibrate another when the reference meter reading contains random error. Let the reference meter be $\mathbf{x}$, and the UUT be $\mathbf{y}$. If both the reference and the UUT are perfectly accurate, a regression line with a gradient of one should be drawn on the *xy* plane:

$$\mathbf{y} = a\mathbf{x} + b, \tag{1}$$

where $a = 1$ and $b = 0$.

If only the UUT has an error (thus an error in the response or dependent variable measurement), the dependent variable $\mathbf{y}^* = \mathbf{y} + \epsilon$ will be measured by the UUT, where the $\mathbf{y}^*$ indicates the *surrogate* reading and ϵ the error. We thus observe $\mathbf{y}^*$ in lieu of $\mathbf{y}$, where:

$$\mathbf{y}^* \sim Normal(\mathbf{y}, \quad \tau\mathbf{y}) \tag{2}$$

The error will add noise, but will not bias the result, as illustrated in the left-hand graphs of Fig. 3. These are Ordinary Least Squares (OLS) regression estimates for increasing values of the standard deviation multiplier τ. We observe that increasing error does not bias the estimates. However, this does not hold for errors in $\mathbf{x}$ of the form

$$\mathbf{x}^* \sim Normal(\mathbf{x}, \ \tau\mathbf{x}), \tag{3}$$

As can be seen on the right-hand side of Fig. 3. For a further graphical illustration, see the UMP Chapter 13 [43], Section 3.2.

We note that mismeasurement is less of a problem for prediction, which is often the goal of M & V models. If you infer some function $\mathbf{y}^* = \theta^*\mathbf{x}^*$ based on measurements of $\mathbf{x}$ made with random error, that relationship defined by θ^* will continue to hold as long as you forecast and measure using $\mathbf{x}^*$ in lieu of $\mathbf{x}$. In such a case a Measurement Error Model (MEM) is unnecessary. This is part of the reason that measurement error is not a greater problem in M & V: often the baseline and reporting period measurements are made with the same instruments, and so the attenuation effect may 'cancel out', as long as inference about the physical meaning of the parameters (e.g. kWh/Heating Degree Day) is not attempted. Consider the 'time-of-week and temperature' M & V regression model [117] in a situation where the temperature is measured with error because the weather station is in a different microclimate to the facility [118]. The relationship between energy use and temperature would be attenuated. This would cause certain elements of the time-of-week parameter vector to seem more influential than they actually are. But this may not be a problem. Suppose that HVAC-related Energy Conservation Measure (ECM) is installed and the model is used for M & V. The forecast (adjusted baseline) energy use in the post-retrofit period will have the same attenuation as the baseline. It would, therefore, be accurate, assuming a calibrated model and same temperature data source. Therefore the total savings estimation will have a similar Normalised Mean Bias Error (NMBE) to the case with no measurement error, although the added noise may lead to a higher Coefficient of Variation on the Mean Squared Error (CVRMSE) on the training set. This being said, one cannot regress energy use against temperature to infer the effectiveness of the ECM, nor can such a regression be transported for project decisions in other places. Furthermore, the confidence interval around the reported savings will also be too narrow.

Table 3
Spurious effect of mismeasurement in $\mathbf{x}$ on various statistics assuming classical additive errors, summarised from Carroll et al. [64], Gustafson [119], and Ree et al. [120].

Statistic	Effect
Mean	None
Variance	Increases
Covariance	None
Regression, single predictor, slope	Decreases
Regression, single predictor, intercept	Increases
Regression, multiple predictors	Complex
Confidence on regression coefficients	Increases
Statistical power for detecting relationships	Decreases
Correlation	Decreases
Partial correlation	Increases
Non-linear features (such as $\mathbf{y} = sin\mathbf{x}$)	Masked

5.1. Mismeasurement in M & V literature

Although attenuation bias due to mismeasurement has been documented in M & V, the effect is not well-known. Except for the UMP Chapters 13 and 23 [43,44], all M & V guidelines discussed so far, as well as M & V regression guides [121] do not mention attenuation, even when measurement errors are discussed. The UMP Chapters 11 and 12 (Sample and Survey Design) [46,6] state that random measurement error does *not* lead to bias, even though survey measurement error is one of the most common MEM test cases [122]. G14-2014 stipulates that the total span of the extra uncertainty created by errors in independent variables shall be determined by biasing the variables to their maximum and minimum values [19]. Attenuation is unaccounted for.

Regarding literature, an MC analysis was done by Sonnenblick and Eto from Lawrence Berkeley in 1995. They found this bias effect for measurement precision of energy programmes [30], Fig. EX-2, and identified it as the errors in variables effect. The measurement of operating hours was considered to be the most sensitive to this effect.

Ridge [123] presented an informative paper on mismeasurement in M & V in 1997. He relates how the Californian utility Pacific Gas and Electric's 1992–1993 Commercial New Construction Program and the 1994 Commercial HVAC program realisation rate estimates were unreasonably low. The realisation rate is the ratio of expected to actual savings. He traced the problem back to random errors in independent (explanatory) variables that led to attenuated estimates. This was corrected for in subsequent studies by the use of dummy variables.

A more recent example of mismeasurement is found in the case where Canadian economists Rivers and Jaccard published a study which found that Demand Side Management (DSM) interventions made no statistically significant impact on energy demand when viewed at a national level [124]. This generated some controversy. Rivers and Jaccard proposed that measurement error in the independent variable (DSM spending proportion vs. EE spending proportion) may have played a role in attenuating the DSM-effect parameter estimate. However, although Violette et al. [125] also acknowledged this errors-in-variables possibility, they proposed that other features of the original Rivers and Jaccard model were more influential.

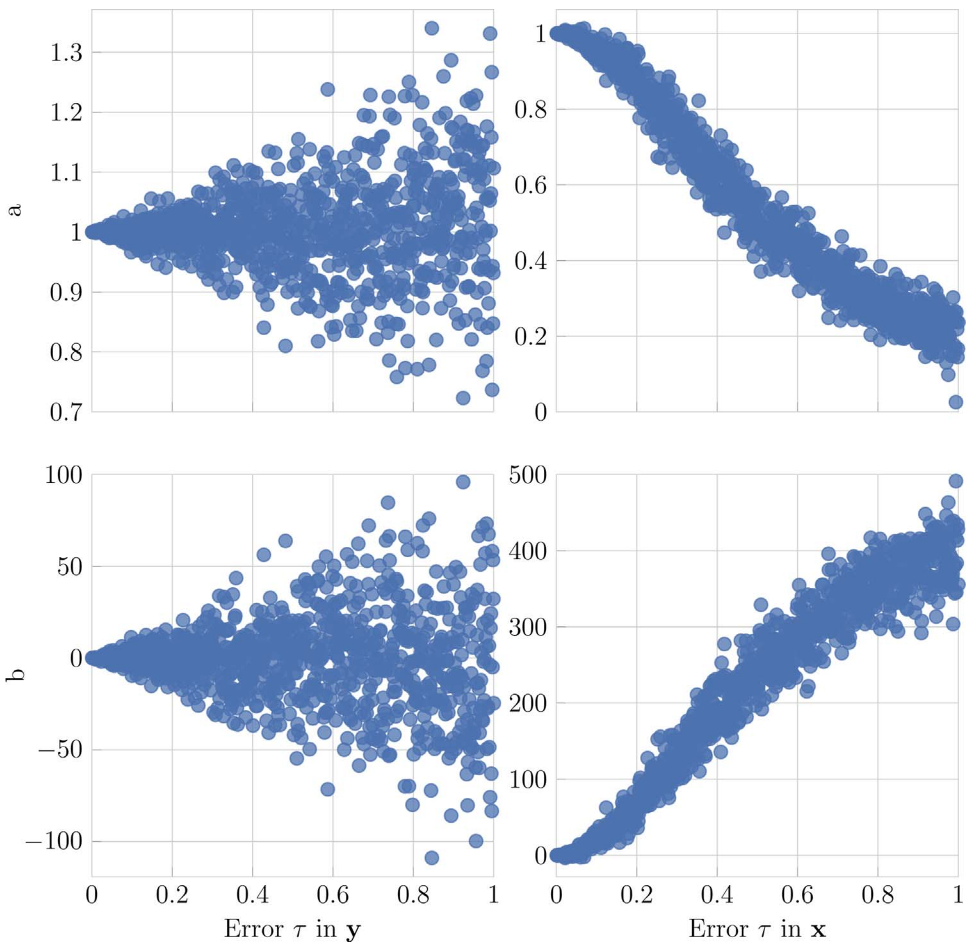


Fig. 3. OLS parameter estimates for $\mathbf{y} = a\mathbf{x} + b$, where $a = 1$ and $b = 0$, given measurement error τ in the form (2) and (3).

5.2. MEM and calibration techniques

There are two main bodies of research addressing measurement errors relevant to energy models. First, commercial electrical metrological techniques have been honed over the last half century. These methods usually employ Test Uncertainty Ratios (TURs), which is the ratio of the precision of the calibrator to that of the UUT. They have had to be revised recently as the accuracy of calibrators and digital multimeters (DMMs) has converged to 8.5 digits (one part in 10^8). Second, trans-disciplinary academic investigations have been conducted using a variety of approaches. These have advanced significantly in response to the stringent and complex requirements of medical fields such as epidemiology, coupled with the relatively poor accuracy of the instruments measuring certain human epidemiological variables.

5.2.1. Electrical calibration techniques

These techniques are applicable mainly to calibration. They are commercial techniques usually using indirect, empirical, conservative methods, and cannot be classified as true MEMs. A TUR of 4:1 is generally required. This means that an instrument accurate to p% may be used to calibrate an instrument accurate to $4p$% (called the Unit Under Test, UUT). This may reflect the other rule of thumb proposed in Section 2.2. However, since DMMs such as the 8.5-digit Fluke 8508A do not allow for a TUR >4 between the UUT and the calibrator, other techniques had to be developed. The simplest and most accurate is to characterize the long-term drift of the instrument by plotting the change in measurement errors over time, and then drawing a regression line through the successive measurement points [126,127]. This regression line has been shown to be more accurate than the individual calibrations [128]. Within limits, and with a large enough calibration history, this technique may be used to accurately quantify an instrument's error without recent calibration. This technique has also been proposed for characterising the stability of a calibrator that may not meet the TUR >4 nominally, but does meet it practically. This is possible as the calibrator's stability specifications are usually lower than what an individual instrument's stability may be, when measured with a more accurate DMM.

On the other hand, if one wants to test an instrument with no history, and one can not achieve the required TURs, alternative methods also exist [129]. For true calibration, the only option is “disciplining” the calibrator by using an additional, more accurate DMM to measure the calibrator output in real time [127].

In cases where an accept/reject decision has to be made rather than full calibration, there are three options: lower the confidence level of the test, invest in a more accurate standard, or analyse and document the measurement points for which inadequate TURs exist. The first option (lowering the confidence level) is called *guard banding*, and is popular in metrology [130–132]. A guard band is a test limit stricter than the instrument specification limit [133]. In other words, by employing guard bands, we can use a calibrator with a TUR of 2 instead of 4. The price we pay is that the UUT may still be rejected, even if the test result falls between the Lower Confidence Limit and the Upper Confidence Limit of the calibrator. This is because to compensate for our lower TUR, the test limits are narrower than the instrument specification limits. Thus guard banding keeps the consumer's risk constant even though a less accurate calibrator is used, but increases the producer's risk for such a case. When considering this approach, one must remember that at a certain level, testing becomes uneconomical. For example, for a TUR of 2 and specification limit of 2 σ, the consumer's risk is as large as it would be if no testing at all took place, and the consumer simply accepted the probability of the unit being outside of specification (probability = 1.2%) [129]. In such scenarios the expected value of the test, or the cost/benefit trade-off between testing and not testing, should be considered.

Rossi and Crenna [134] provided a good example of setting test limits lower than specification limits for in-house testing at the producer side to minimise risk, which they applied to water meters [104]. To this end, they have developed a software package called UNCERT - essentially an automated MC approach. Researchers from the US National Institute for Standards and Technology (NIST) have also shown that a Bayesian approach to the accept/reject decision rule of ISO 14253-1 (inspection of work pieces) [135] delivers superior results in cases where it is applicable [136].

5.2.2. Transdisciplinary techniques

Not all uncertainty analysis models (also known as uncertainty quantification models) considering measurement error are MEMs. On the other hand, some probabilistic models using MC methods could well be incorporated into MEMs, although their function in most literature is exploratory what-if analysis, sensitivity analysis, or forecasting (see Section 6). Other methods are simply robust: insensitive to outliers.

There is a notable amount of literature on MEMs, although much of it is too technical to be useful to the M & V practitioner without a strong background in statistics. For linear problems Fuller [137] is popular, and his method-of-moments is straightforward and recommended for OLS regression with additive measurement errors (cf. Carroll et al. [64]). The non-linear case presents a greater challenge, but may also be more relevant to M & V and instrument calibrations as shown by Carobbi et al. [138]. The most appropriate (and readable) treatments are by Carroll et al. [64], and Gustafson [119].

MEMs can be divided into functional and structural approaches. Functional approaches make no assumptions about underlying distributions (thus avoiding model misspecification) and include Regression Calibration and simulated extrapolation (SIMEX). Structural approaches make assumptions about the underlying distributions and relations governing the measurement system and include Maximum Likelihood Estimation (MLE) and Bayesian Markov Chain Monte Carlo (MCMC) techniques. All four of these techniques are powerful and can yield useful results if applied well. The choice of method depends on its appropriateness to the data and ease of implementation.

The **SIMEX** concept is simple and powerful. Suppose we know that our variance $\mathrm{VAR}(\mathbf{x}^*|\mathbf{x}) = \tau$. We also know our current parameter estimate $\theta^*|\mathbf{x}^*$, that is, $\theta^*|\tau_0^2$. We want to know our true parameters $\theta|\mathbf{x}$. If we now *increase* the error τ in the dataset, the parameter estimates will start drifting away from their true values due to attenuation. In this way, we can obtain values for $\theta^*|\tau_1^2$, $\theta^*|\tau_2^2$, $\theta^*|\tau_3^2$, ... We will observe a trend, and can fit a curve to these points. Extrapolating backwards will then yield $\theta^*|(\tau = 0)$, which is $\theta|\mathbf{x}$. The disadvantage is that SIMEX is difficult for cases where there are combined multiplicative and additive errors and that it can be expensive for non-linear higher dimensional models. It has also been found that in certain cases MLE methods yield considerable smaller variances [139], although for most applications SIMEX is simple and effective.

Regression Calibration methods essentially trade an exposure model for a validation (calibration) sample: a sub-sample measured without error, using a 'gold standard'. From the information gleaned from the sub sample, values for $\mathbf{x}$ are imputed instead of the $\mathbf{x}^*$ values measured. Repeated measurements may also be used. It is not susceptible to bias due to model misspecification since the exposure models do not need to be specified. Regression Calibration is useful for trials where extensive, precise, or repeated testing is only feasible for a small sub-sample.

One potential weakness of the Regression Calibration method is that it maps $\mathbf{x}^*$ onto $\mathbf{x}$ in a one-to-one fashion, where methods such as Bayes-MCMC consider all reasonable values for $\mathbf{x}$ given the data. Therefore the uncertainty is specified as fully as possible. This avoids the effect of not considering the uncertainty contribution of imputing $\mathbf{x}$ values for the first step of the Regression Calibration procedure.

Maximum Likelihood Estimation has become a very powerful structural approach in many areas of statistics. MLE techniques have the potential of producing better estimates than functional approaches if the model is well specified, although this is often difficult [64].

Kennedy and O'Hagan presented a seminal paper on which much of the current Gaussian Process (GP) energy MLE research is based [140]. The short discussion below will focus on this method, which may be classified as Bayesian or quasi-MLE, depending on your preference. Purer MLE MEMs are also used [64]. GPs are popular because they are a generic, convenient and accurate. In a GP, every data point is assumed to be normally distributed, with the dataset then assumed to have a multivariate normal distribution. The GP kernel is a function that describes how the covariance matrix between the data points behaves, and the parameters of the kernel function are determined using an MLE technique with a two-step Expectation Maximisation algorithm. In the E-step the algorithm averages over the unknown explanatory variable $\mathbf{x}$ based on the observations of the response $\mathbf{y}$ to $\mathbf{x}^*$, and updates the expected log-likelihood. It uses numerical integration as the expressions may not be closed-form. The M-step maximises the log-likelihood of $\mathbf{x}$, after which the algorithm returns to the E-step and iterates until maxima are found. Recently Burkhart et al. have applied this successfully in the energy monitoring and evaluation field [141]. They found that adding MC Expectation Maximisation to a Gaussian Process to account for uncertainty in input data makes parameter estimates more robust, and requires fewer data. They then propose trading GUM Type A uncertainties for Type B uncertainties to minimise cost.

Methods such as GP regression present advantages over full Bayesian methods in that model misspecification and computational expense becomes less of a concern. However, MLE methods are advanced empirical Bayesian methods. Full Bayesian methods provide some advantage since the models are easily specified and solved, no approximations are necessary, and standard errors on the estimates are more easily calculated [119]. Stopping or convergence criteria are a concern for both approaches [141]. Gelman [142] also notes that EM algorithms with multivariate normal approximations are not ideal for small data sets as convergence is only asymptotic, and the normal distribution not ideal for describing such cases.

Much literature on the technical merits and application of **Bayesian methods** exists, as it is the natural structural MEM approach [64]. It is more than a machine learning algorithm: it is rather a branch of statistics derived from conditional probability logic. Very briefly, Bayesianism can be explained as follows. The unknown parameters θ are viewed as random variables defined by 'prior' probability distributions. With the data $\mathbf{D}$, they are solved for as $\pi(\theta|\mathbf{D})$. Bayesianism is different to frequentism, which sees the parameters as fixed and the data as random realisations which will even out to the parameters in the long run. This distinction is often quoted, but remains obscure to someone without Bayesian modelling experience. As an explanatory example, consider the $\mathbf{y} = a\mathbf{x} + b$ linear regression case discussed in Section 5. We define a and b as

$$a, b \sim Normal(\mu = 0, \sigma = 10^5). \tag{4}$$

These are the priors: they define the information we have about the system that is not present in the data itself. In the case above, the priors are vague because we presume to know little about the system. Bayes theorem states that

$$\pi(\theta|\mathbf{D}) = \frac{\pi(\mathbf{D}|\theta)\pi(\theta)}{\pi(\mathbf{D})}, \tag{5}$$

and allows us to invert our priors $\pi(\theta)$ and data $\pi(\mathbf{D}|\theta)$ to find what we are interested in: the probability distributions of the unknown parameters, given the data: $\pi(\theta|\mathbf{D})$. This usually requires intractable integration and the specification of the probability of our data $\pi(\mathbf{D})$. However, the MCMC numerical algorithm circumvents this difficulty by generating a Markov process whose stationary distribution is the posterior $\pi(\theta|\mathbf{D})$. By sampling in Monte Carlo fashion from this distribution, parameter distributions are found numerically.

Bayesian approaches with non-informative priors provide MLE estimates of data [142]. However, they are more flexible since they do not require ad hoc techniques dealing with special cases, as with most frequentist statistics. This allows rapid model development and less time spent on building complex, realistic models. Mathieu et al. also recommend this approach for error analysis of energy measurement and verification, especially for cases where errors are financially significant [143]. For the reader unfamiliar with Bayesian techniques, Kruschke [144] and Gelman et al. [142] are recommended; Kruschke being more practically oriented and Gelman et al. more advanced.

The disadvantages of the Bayesian-MCMC techniques are that they can be computationally expensive, susceptible to model misspecification, and requires more thinking on the part of the practitioner. The computational expense becomes a problem when many variables (or data points) have uncertainties in them which need to be modelled using MCMC. The model then suffers from the curse of dimensionality. Thus, for problems such as the real-time calibration of thermal network parameters is needed, Bayesian techniques have been found to be too computationally expensive even though they are more robust than lightweight 'gray-box' techniques [145]. Variational inference may alleviate this concern, and although the technique is relatively new it has been implemented in popular software [146]. Model misspecification arises when the true error structure is different from the one specified in the model. Investigating the robustness or sensitivity of the model to such assumptions becomes necessary. Last, there are few simple 'recipes' in Bayesian statistics. There is no *t*-test or *F*-test blanket equivalent, although Kruschke provides alternatives [144]. Generally, however, Bayesian solutions are more problem-specific than popular frequentist tests.

Several non-technical reasons for the application of Bayesian approaches to M & V should be noted. First, a Bayesian MEM is similar to a standard, well-specified Bayesian model. The model's ability to deal with measurement errors follows from the nature of the Bayesian mathematics itself. Second, the development of Markov Chain Monte Carlo (MCMC) techniques has allowed for the previously intractable integration involved in most non-trivial Bayesian calculations to be done efficiently and accurately. The numerical MCMC model converges reliably on the analytical solution [147]. Third, as noted in the GUM Supplement [49], the MC approach is not distribution dependent and is, therefore, more flexible. Fourth, intuitive and powerful open-source software libraries have become available by which Bayesian models specified and solved. Scaling to more complex models is straightforward. Although BUGS and JAGS have been the mainstay software packages in the past, Stan [148] probably leads at the moment. It can be implemented in various languages such as Python, R, Matlab, Julia, or C++. PyMC3 [149] is also worth mentioning. It is written in and for the Python environment and is gaining popularity due to its simple interface, discrete variable and missing value support, and ease of integration into the popular scientific Python environment. Both packages are being developed actively.

6. Project decisions under measurement uncertainty

Pendrill [105] rightly observed that measurements are seldom made for their own sake, but rather in support of a financial decision. Indeed, decision maker uncertainty about cost-effectiveness is the most frequently-cited barrier to the commissioning of energy projects [150]. However, the contribution of technical uncertainty in the performance of the ECM is usually smaller than economic uncertainty contributions, as noted by Rysanek en Choudhary [151] and Friege and Chappin [152].

Regarding the M & V literature on the subject, project risk associated with measurement uncertainty has been identified by both researchers [143,153] and practitioners [154], but little M & V literature addresses this topic directly. Ligier et al.'s recent contribution [28] on decision support explicitly in the context of building simulation and M & V comes very close, and Boxer et al.'s method for self-benchmarking can also be viewed as an M & V and decision support tool [155]. We will consider four aspects below. First, M & V guides on risk or its components namely cost and uncertainty. Second, M & V research related to the aforementioned topics. Third, financial energy project decision support literature. Fourth, metrological decision support literature. Since building energy simulation is a subject on its own, that will be dealt with in Section 6.1.

Sonnenblick and Eto [30] investigated expected monitoring project value as a function of measurement precision in 1995 already. In that case, it was applied to overall DSM project cost-effectiveness: levelized project cost vs. levelized savings. Probably the most notable measurement/cost treatment is ASHRAE Guideline 14–2002 [18], which supplies elaborate tables for determining measurement costs for different instruments in various project scenarios. However, it does not calculate risk adequately [156]. The SEE Action Guide [25] also provides an introductory overview of measuring budgets in the context of project risk.

Regarding research, a foundational mathematical description of M & V has been compiled [157], and a useful theoretical summary of different uncertainty approaches in power systems given [158]. M & V sampling, metering have been traded off to minimise project cost [2–4], and modelling uncertainty was added later [159], although risk was not treated explicitly. These designs were extended to a Bayesian framework where risk could be incorporated [55,160], although the research did not focus on risk. An insightful cost-benefit trade-off for chilled-water system design in the context of uncertainty [161] influenced the G14 [18] approach. Preliminary work on decisions in Energy Performance Contracts (EPCs) under measurement uncertainty has also been presented [26]. It is noted that attempts have been made to quantify the risk due to energy meter measurement uncertainty [69,162]. However, this calculation was much too simplistic, and was presented by a marketing manager of a meter manufacturer calling for even-more-stringent standards to which the latest meters could be qualified. This standard is unnecessary since the current Class 0.2S energy meters are the smallest uncertainty sources in almost any conceivable project, and their uncertainties can already be neglected for risk calculation purposes in many cases [26].

Research on financial decision support related to EPC, project uncertainty and risk have been conducted from an economic perspective using MC analysis [163] and other techniques [164]. The US Department of Energy's *EnergyPlus* software is usually used [165]. Deng et al. [166] provided a useful summary of the design of EPCs under uncertainty and presented a relatively sophisticated EPC decision model [167]. Measurement uncertainty is not considered explicitly in these cases, although it can be incorporated without much extension.

Focusing now on measurement, relevant research on this topic has also been conducted from a legal metrological perspective. Here measurement uncertainty and cost are traded off in a decision support framework. Crenna [104] and Pendrill [168,105] used an MC method, while Fearn [169] used a more cumbersome analytical approach. However, the focus of these studies is accept/reject decisions based on a standard, rather than the verification of individual measurements. Risk was viewed from a government perspective as a function of the cost of emissions to society. Sonnenblick and Eto also used this cost function in their report on the cost-effectiveness estimates of energy projects in the context of measurement precision [30], and Rysanek and Choudhary [151] used the marginal abatement cost: the ratio of net present value to GHG units saved. These metrics seem more rational than short-term financial risk measures when one considers the broader goals of energy research.

6.1. Measurement uncertainty in building simulation

Research into uncertainty in building energy modelling (BEM) has increased dramatically in the last ten years. This is because it has been recognised that considering model input uncertainty is essential to

Table 4
Recent and notable energy project decision support literature considering input uncertainty. Note that literature on BEM calibration with goodness-of-fit as an outcome is not included. These usually employ CVRMSE and NMBE.

Author	Year	Application	Sensitivity Analysis	Uncertainty Analysis Method	Decision Analysis Metric	Reference
Sonnenblick, Eto	1995	Owner, ESCO profitability	–	MC	CE	[30]
Kammerud, Gillespie, Hydeman	1999	Chilled water system design	(Taylor Series Expansion)	Quadrature	Discounted cash flow CE	[161]
de Wit, Augenbroe	2002	Thermal comfort	–	Bayes	Expected Utility	[187]
Pendrill, Källgren	2006	Exhaust gas analysers	–	Analytical	Cost to society	[168]
Crenna, Rossi, Bovio	2009	Water meters	–	MC	Non-conformance cost	[104]
Jackson	2010	EE Investments	–	MC	Value at risk	[163]
Heo (PhD Thesis)	2011	BEM	Morris, MC-LHS	Bayes-GP	EV, CE with payback time, Guaranteed Savings, Savings Curve Score	[156]
Tian, Choudhary	2012	BEM	MC-LHS, SRC, MARS	LP, Bayes	EUI	[188]
Booth, Choudhary	2013	BEM, Decision Support	Factorial Sampling	Bayesian regression	NPV-PDF, Multi-criteria decision utility, CE, CEAC	[189]
Burhenne, Tsvetkova, Jacob, Henze, Wagner	2013	BEM	Sobol' Sequence	MC Filtering	NPV-CE	[164]
Lam, Yik, Chan	2013	EPC	Differential: Influence Coefficient	MC-LHS	Savings shortfall	[165]
Rysanek, Choudhary	2013	BEM	–	Non-probabilistic	CE; MAC: NPV vs. GHG emissions saved, Discounted payback period vs. required capital; Wald's Minimax, Hurcwiz's Maximin, Savage's Regret	[151]
Sun, Gu, Wu, Augenbroe	2014	HVAC sizing	MC-LHS, LASSO, ANOVA	MC	Unmet peak load percentage	[190]
Sun (PhD Thesis)	2014	BEM, HVAC	MC-LHS, LASSO, ANOVA	MC	CRPS, PIT	[5]
Heo, Augenbroe, Graziano, Meuhleisen, Guzowski	2015	BEM	Morris, MC-LHS	Bayes	% savings, EV(savings), fifth quantile savings predictions	[184]
Lee, Lam, Lee, Chan	2016	Air-cooled chiller replacement EPC	Correlation analysis	MC-LHS	EPC compliance PDF	[191]

Abbreviations: BEM, Building Energy Model; CEAC, Cost Effectiveness Acceptability Curve; CE; Cost Effectiveness; CRPS, Continuously Ranked Probability Score; CVRMSE: Coefficient of Variance on the Root Mean Square Error; EPC, Energy Performance Contract; ESCO, Energy Services COmpany; EUI, Energy Use Intensity; EV, Expected Value; GHG, Greenhouse Gas; GP, Gaussian Process; LASSO, Least Absolute Shrinkage and Selection Operator; LP, Linear Programming; MAC, Marginal Abatement Cost; MARS, Multivariate Adaptive Regression Spline; MC, Monte Carlo; MC-LHS, Monte Carlo Latin Hypercube Sampling; NMBE, Normalised Mean Bias Error; NPV, Net Present Value; PIT, Probability Integral Transform; PDF, Probability Density Function; SRC, Standardized Regression Coefficient.

identifying which ECMs should be implemented.

A full review of building simulation calibration literature is beyond the scope of this survey, and we will focus on cases where measurement uncertainty could be considered. For a broader view, a useful starting point is Reddy et al.'s research series forming part of ASHRAE's investigation of calibrated simulation in RP-1051 [170–173], and Coakley, Raftery, and Keane's more up-to-date review, considering uncertainty in detail as well [174]. Heo's PhD thesis also provided an in-depth discussion and case study of one approach [156].

Databases of parameter uncertainties have been compiled [175], and these, or results from the literature, are used for uncertainty analysis or quantification. The key problem, however, is that doing an MC simulation considering all parameters simultaneously is infeasible due to the curse of dimensionality. Sensitivity analysis methods are thus needed to reduce the number of parameters to a feasible figure. Sun et al. provided one of the better discussions on this topic [118], and Tian also wrote an informative review [176]. Several excellent examples of this process have been published, and are summarised in Table 4.

Most building simulation research accounts for varying input parameters through uncertainty and sensitivity analysis. However, much of this research concerns itself with how varying the input parameters changes the output, but not how *variance in* the input parameters affects the output. In other words, it does not ask how noisy input may attenuate the output, but how biased input will bias the output. It is possible that this is accounted for in GPs, although it is uncertain.

Two related studies deserve mention. To alleviate the burden of MC computation for building simulation studies with large uncertainties and many options and combinations, Rysanek and Choudhary proposed a lightweight non-probabilistic decision approach [151]. These scenarios apply more to simulation (modelling) uncertainty rather than measurement uncertainty. On the other side of the spectrum, Sanyal et al. reported a machine learning and supercomputer-based method to alleviate the modelling burden by pre-tuning simulation inputs to extant data for standard US buildings [177]. This speeds up model building significantly.

In what seems to be a recurring theme, the Bayesian approach is becoming increasingly popular because of its uncertainty quantification features. Riddle and Muehleisen provided a useful introduction to building calibration with such models [178], and Heo has recently presented an overview of building simulation models under uncertainty, as well as an introduction to the Bayesian approach [179]. Note that in a Bayesian framework measurement, sampling, and modelling errors are considered simultaneously, although they remain distinct [180].

Heo and Augenbroe have built up a noteworthy body of work on building simulation covariate calibration and uncertainty analysis using (Bayesian) Gaussian Process methods [181,182]. Quantitative risk analysis for decision support in retrofit project planning was then explored with a focus on the accuracy of the simulation rather than metering decision making [183]. Their latest research incorporates this into a scalable methodology whereby more optimal retrofit decisions can be made, given uncertainty in input parameters [184]. Along similar lines, a lightweight and reasonably accurate alternative to the GP has been proposed [185]. Another notable contribution has been made by Tian et al. who used sophisticated data analysis and Bayesian methods to show the relative importance of different data on building calibration, and the robustness of the Bayesian method to missing input data [186]. Bayesian methods have therefore been demonstrated to deliver very good estimates, but Heo notes that even if this were not the case, they could still be superior to deterministic models since they quantify model prediction uncertainty distributions [181].

7. Recommendations

In the light of the literature on measurement uncertainty and M & V, several recommendations can be made. Regarding M & V reporting,

1. The effect of power quality on M & V studies should be noted in M & V reports. Stating the meter type and meter calculation method should be standard.
2. The sensitivity to mismeasurement should at least be investigated for M & V regression models. In some cases it may be necessary to use MEMs to compensate for measurement error effects such as bias and unrealistically high statistical power.

Regarding further research,

1. Input uncertainty quantification is a now firmly established in the building simulation field. However, it is unclear whether the effect of mismeasurement on building energy simulation calibration is accounted for. Attenuation bias may produce incorrect results in the parameter screening phase by lowering the apparent influence coefficients of certain mismeasured, influential variables. A study on this phenomenon is therefore warranted.
2. The in-situ calibration of smart meters through the smart grid is an interesting and potentially revolutionary possibility. Instead of calibrating meters in a laboratory using reference instruments, other techniques could be used. For example, by cross-referencing meters in a network, or utilising smart devices acting as loads one could reduce calibration costs significantly.
3. Although risk-conscious capital expenditure decisions in energy projects have been investigated, the same depth of treatment has not been given to energy monitoring. By utilising metrics such as those found in Table 4, monitoring costs may be optimised, leading to risk-optimal measurement and sampling designs.

8. Conclusion

Measurement uncertainty remains an important consideration in energy M & V. Not only does this apply to electrical meter measurements, but also to the quantification of uncertainty in covariate specification. Even unbiased random error in covariate measurement may lead to biased parameter estimates. However, the contribution of individual measurement uncertainties, and the cost and effort expended to quantify or mitigate them should be considered carefully to allocate resources efficiently. In some cases, more accurate quantification or calibration of instruments may make little difference to the project decisions.

Many techniques are used for uncertainty quantification, but Bayesian methods are notable for their support in almost all related fields, from general metrology to Measurement Error Methods and decision support. However, these techniques are still new and represent a growing field in energy research.

9. Acknowledgements

The financial assistance of the National Research Foundation (NRF) (grant number 95122) and the National Hub for the Postgraduate Programme in Energy Efficiency and Demand Side Management towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the authors, and are not necessarily to be attributed to the NRF.

References

[1] Efficiency Valuation Organization, International Performance Measurement and Verification Protocol Vol. 1 (January 2012).
[2] Ye X, Xia X, Zhang J. Optimal sampling plan for clean development mechanism lighting projects with lamp population decay. Appl Energy 2014;136:1184–92. http://dx.doi.org/10.1016/j.apenergy.2014.07.056.
[3] Ye X, Xia X. Optimal metering plan for measurement and verification on a lighting case study. Energy 2016;95:580–92.
[4] Carstens H, Xia X, Ye X. Improvements to longitudinal Clean Development Mechanism sampling designs for lighting retrofit projects. Appl Energy 2014;126:256–65. http://dx.doi.org/10.1016/j.apenergy.2014.03.049.

[5] Sun Y. Closing the building energy performance gap by improving our predictions. Atlanta, Georgia: Georgia Institute of Technology; 2014.
[6] Baumgartner R. The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures, National Renewable Energy Laboratory, 2013, Ch. 12: Survey Design and Implementation Cross-Cutting Protocols for Estimating Gross Savings.
[7] Conant A, Liu H, Proctor J, Wilcox B, Sources of error in home energy use calculations: Evaluation in real-world laboratory homes. In: Proceedings of the International Energy Program Evaluation Conference, International Energy Programme Evaluation Conference, Long Beach, California; 2015.
[8] Wang Q, Augenbroe G, Kim J-H, Gu L. Meta-modeling of occupancy variables and analysis of their impact on energy outcomes of office buildings. Appl Energy 2016;174:166–80.
[9] Lanzisera S, Dawson-Haggerty S, Cheung HI, Taneja J, Culler D, Brown R. Methods for detailed energy data collection of miscellaneous and electronic loads in a commercial office building. Berkeley, California: Lawrence Berkeley National Laboratory; 2014. Tech. Rep. LBNL-6384E.
[10] Kamilaris A, Kalluri B, Kondepudi S, Wai TK. A literature survey on measuring energy usage for miscellaneous electric loads in offices and commercial buildings. Renew Sustain Energy Rev 2014;34:536–50.
[11] Ward R, Choudhary R, Heo Y, Rysanek A. Exploring the impact of different parameterisations of occupant-related internal loads in building energy simulation. Energy Build 2016;123:92–105.
[12] Institute for Electrical Innovation, Utility-scale smart meter deployments, Tech. rep., The Edison Foundation; September 2014.
[13] Granderson J, Touzani S, Custodio C, Sohn MD, Jump D, Fernandes S. Accuracy of automated measurement and verification (M & V) techniques for energy savings in commercial buildings. Appl Energy 2016;173:296–308. http://dx.doi.org/10.1016/j.apenergy.2016.04.049.
[14] World energy outlook 2015 factsheet, Tech. rep., International Energy Agency; 2015. URL 〈http://www.worldenergyoutlook.org/media/weowebsite/2015/WEO2015_Factsheets.pdf〉.
[15] Bartsch A, Burns C, Murphy B, Sanford-Long L. The promised land: Making use of data from smart meters. In: Proceedings of the International Energy Program Evaluation Conference, Long Beach, California; 2015.
[16] Stratton C, McGaraghan M, Kisch T, Zakarian A, Borocz P, Rubin E, Richter C, Anderson M. Leveraging big data to develop next generation energy efficiency programs and energy regulations. In: Proceedings of the International Energy Program Evaluation Conference, Long Beach, California, 2015.
[17] Haberl J, Culp C, Claridge D. Ashrae's guideline 14-2002 for measurement of energy and demand savings: How to determine what was really saved by a retrofit. In: Proceedings of the Fifth International Conference for Enhanced Building Operations, no. ESL-IC-05-10-50, Energy Systems Laboratory, 2005.
[18] American Society of Heating, Refrigeration and Air-Conditioning Engineers, Inc., Guideline 14-2002, Measurement of Energy and Demand Savings; June 2002.
[19] American Society of Heating, Refrigeration and Air-Conditioning Engineers, Inc., Guideline 14-2014, Measurement of Energy, Demand, and Water Savings; December 2014.
[20] California Commissioning Collaborative, Guidelines for Verifying Savings from Commissioning Existing Buildings; 2012.
[21] Reddy T, Claridge D. Uncertainty of measured energy savings from statistical baseline models. HVACR Res 2000;6(1):3–20. http://dx.doi.org/10.1080/10789669.2000.10391247.
[22] Efficiency Valuation Organization, International Performance Measurement and Verification Protocol: Statistics and Uncertainty for IPMVP; June 2014.
[23] National Renewable Energy Laboratory, Uniform Methods Project. URL 〈http://energy.gov/eere/about-us/ump-home〉.
[24] American Society for Heating, Refrigeration, and Air Conditioning Engineers, Engineering analysis of experimental data, Tech. rep., American Society for Heating, Refrigeration, and Air Conditioning Engineers; 1986.
[25] State and Local Energy Efficiency Action Network, Energy Efficiency Program Impact Evaluation Guide, US Department of Energy, DOE/EE-0829; December 2012.
[26] Carstens H, Xia X, Yadavalli S. Measurement uncertainty and risk in measurement and verification projects. In: Proceedings of the International Energy Programme Evaluation Conference, Long Beach, California; 2015.
[27] Carstens H, Xia X, Yadavalli S. Low-cost energy meter calibration method for measurement and verification. Appl Energy 2017;188:563–75. http://dx.doi.org/10.1016/j.apenergy.2016.12.028.
[28] Ligier S, Robillart M, Schalbart P, Peuportier B. Energy performance contracting methodology based upon simulation and measurement. in: Building Simulation 2017; 2017.
[29] Vine E, Kats G, Sathaye J, Joshi H. International greenhouse gas trading programs: a discussion of measurement and accounting issues. Energy Policy 2003;31(3):211–24.
[30] Sonnenblick R, Eto J. A framework for improving the cost-effectiveness of DSM program evaluations, Tech. Rep. LBL37158, Lawrence Berkeley Laboratory, Berkeley, California; 1995.
[31] Shishlov I, Belassen V. Review of monitoring uncertainty requirements in the CDM, Tech. Rep. Working Paper No. 2014-16, CDC Climat Research; October 2014.
[32] Michaelowa A, Hayashi D, Marr M. Challenges for energy efficiency improvement under the CDM - the case of energy-efficient lighting. Energy Effic 2009;2(4):353–67.
[33] United Nations Framework Convention for Climate Change, Approved Small-Scale Methodology AMS II.C, Demand-Side Activities for Specific Technologies.
[34] United Nations Framework Convention for Climate Change, Approved Small-Scale Methodology AMS II.J, Demand-Side Activities for Efficient Lighting Technologies.
[35] Concept note: Uncertainty of measurements in large-scale CDM methodologies, no. CDM-EB73-AA-A04, UNFCCC CDM; 2013.
[36] International Standard Organization, Guide 98-3 (2008) Uncertainty of measurement part 3: Guide to the expression of uncertainty in measurement (GUM: 1995).
[37] Birch K, Measurement good practice guide no. 36: Estimated uncertainties in testing, Tech. rep., British Measurement and Testing Association, Teddington, Middlesex, United Kingdom; March 2003.
[38] United Kingdom Accreditation Service, UKAS M3003: The expression of uncertainty and confidence in measurement; November 2012.
[39] European Accreditation, EA-4/02 M: 2013: Evaluation of the uncertainty of measurement in calibration.
[40] International Standards Organization, ISO 17025: General requirements for the competence of testing and calibration laboratories; 2005.
[41] United Nations Framework Convention for Climate Change, Clean Development Mechanism Methodology Booklet; November 2015.
[42] Kacker R, Jones A. On use of Bayesian statistics to make the Guide to the expression of Uncertainty in Measurement consistent. Metrologia 2003;40(5):235. http://dx.doi.org/10.1088/0026-1394/40/5/305.
[43] Violette D. The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures, National Renewable Energy Laboratory, 2013, Ch. 13: Assessing Persistence and Other Evaluation Issues in Cross-Cutting Protocols.
[44] Violette D, Rathbun P. The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures, National Renewable Energy Laboratory, 2014, Ch. 23 Estimating Net Savings: Common Practices.
[45] Mort D. The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures, National Renewable Energy Laboratory, 2013, Ch. 9: Metering Cross-Cutting Protocols.
[46] Khawaja MS, Rushton J, Keeling J. The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures, National Renewable Energy Laboratory, 2013, Ch. 11: Sample Design Cross-Cutting Protocols.
[47] Neyman J. Outline of a theory of statistical estimation based on the classical theory of probability. Philos Trans R Soc Lond Ser A, Math Phys Sci 1937;236(767):333–80.
[48] Lira I, Wöger W. Comparison between the conventional and Bayesian approaches to evaluate measurement data. Metrologia 2006;43(4):S249.
[49] Joint Committee for Guides in Metrology, OIML G1-101 Evaluation of measurement data - Supplement 1 to the Guide to the expression of uncertainty in measurement - Propagation of distributions using a Monte Carlo method, International Organization of Legal Metrology, Paris, France; 2008.
[50] Burhenne S, Jacob D, Henze GP. Sampling based on sobol' sequences for monte carlo techniques applied to building simulations. In: Proceedings of the International Conference on Building Simulation, 2011, pp. 1816–1823.
[51] Lira I, Grientschnig D. Bayesian assessment of uncertainty in metrology: a tutorial. Metrologia 2010;47:R1–14.
[52] Kuang YC, Rajan A, Ooi MP-L, Ong TC. Standard uncertainty evaluation of multivariate polynomial. Measurement 2014;58:483–94. http://dx.doi.org/10.1016/j.measurement.2014.09.022.
[53] Rajan A, Ooi MP-L, Kuang YC, Demidenko SN. Analytical standard uncertainty evaluation using Mellin transform. IEEE Access 2015;3:209–22. http://dx.doi.org/10.1109/ACCESS.2015.2415592.
[54] Mellin transform-based moment calculator for multivariate polynomials, [accessed 20 December 2016]. URL 〈http://tc32.ieee-ims.org/content/mellin-transform-based-moment-calculator-multivariate-polynomials〉.
[55] Carstens H, Xia X, Yadavalli S. Efficient longitudinal population survival survey sampling for the measurement and verification of building retrofit projects, Energy and Buildings 150 163–176. http://dx.doi.org/10.1016/j.enbuild.2017.04.084.
[56] Johnson NL. Systems of frequency curves generated by methods of translation. Biometrika 1949;36(1/2):149–76.
[57] Hill I, Hill R, Holder R. Algorithm AS 99: fitting Johnson curves by moments. J R Stat Soc Ser C (Appl Stat) 1976;25(2):180–9. http://dx.doi.org/10.2307/2346692.
[58] Rajan A, Kuang YC, Ooi MP-L, Demidenko SN. Benchmark test distributions for expanded uncertainty evaluation algorithms. IEEE Trans Instrum Meas 2016;65(5):1022–34. http://dx.doi.org/10.1109/TIM.2015.2507418.
[59] Lira I. The gum revision: the Bayesian view toward the expression of measurement uncertainty. Eur J Phys 2016;37(2):025803.
[60] Estler W. Measurement as inference: fundamental ideas. CIRP Ann - Manuf Technol 1999;48(2):611–31.
[61] Rossi GB. A probabilistic model for measurement processes. Measurement 2003;34(2):85–99.
[62] Cox MG, Rossi GB, Harris PM, Forbes A. A probabilistic approach to the analysis of measurement processes. Metrologia 2008;45(5):493.
[63] Rossi GB, Crenna F, Codda M. Measurement results by direct calculation of probability distributions. Adv Math Comput Tools Metrol VI 2004;66:106.
[64] Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. Measurement error in nonlinear models: a modern perspective. Boca Raton, Florida: CRC Press; 2006.
[65] ANSI C12.20 American National Standard for Electricity Meters - 0.2 and 0.5 Accuracy Classes.
[66] IEC 62053-21: Electricity metering equipment (a.c.) - particular requirements - part 21: Static meters for active energy (classes 1 and 2).
[67] IEC 62053-22 electricity metering equipment (a.c.) - particular requirements -part 22: Static meters for active energy (classes 0,2 s and 0,5 s).
[68] IEC 62053-23:2003 electricity metering equipment (a.c.) - particular requirements - part 23: Static meters for reactive energy (classes 2 and 3).

[69] L. A. Irwin, A high accuracy standard for electricity meters, in: 2010 IEEE Power and Energy Society Transmission and Distribution Conference and Exposition; 2010, pp. 1–3.
[70] IEC 60044-8: Instrument transformers part 8: Electronic current transformers.
[71] Directive 2004/22/EC of the European Parlaiment and of the council of 31 march 2004 on measuring instruments; 2004.
[72] Abbas M, Al-Hadhrami LM, Vaqar SA. Perspectives in electric energy metrology. In: Proceedings of the 2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), IEEE; 2015, pp. 1243–1248.
[73] Ahmad MW, Mourshed M, Mundow D, Sisinni M, Rezgui Y. Building energy metering and environmental monitoring-a state-of-the-art review and directions for future research. Energy Build 2016;120:85–102.
[74] Polese LG, Frank S, Sheppy M, Lobato C, Rader E, Smith J, Long N. Monitoring and characterization of miscellaneous electrical loads in a large retail environment, Tech. Rep. NREL/TP-5500-60668, National Renewable Energy Laboratory; February 2014.
[75] Lorek MC, Chraim F, Pister KS, Lanzisera S. COTS-based stick-on electricity meters for building submetering. IEEE Sens J 2014;14(10):3482–9.
[76] Personal Correspondence with primary author, low Cost metering manufacturer; July 2015.
[77] 5th CEER Benchmarking Report on the Quality of Electricity Supply 2011, Tech. rep., Council for European Energy Regulators; 2011.
[78] CENELEC Harmonisation Document HD 472 S1:1988.
[79] ANSI, C84.1-1982: American national standard standard for electric power system and equipment - voltage ratings (60hz).
[80] Sharma K, Saini LM. Performance analysis of smart metering for smart grid: an overview. Renew Sustain Energy Rev 2015;49:720–35.
[81] Gallo D, Landi C, Pasquino N, Polese N. A new methodological approach to quality assurance of energy meters under nonsinusoidal conditions. IEEE Trans Instrum Meas 2007;56(5):1694–702.
[82] Cataliotti A, Cosentino V, Nuccio S. The measurement of reactive energy in polluted distribution power systems: an analysis of the performance of commercial static meters. IEEE Trans Power Deliv 2008;23(3):1296–301.
[83] Cataliotti A, Cosentino V, Lipari A, Nuccio S. Metrological characterization and operating principle identification of static meters for reactive energy: an experimental approach under nonsinusoidal test conditions. IEEE Trans Instrum Meas 2009;58(5):1427–35.
[84] Gallo D, Landi C, Luiso M. Power meter verification issue: Reactive power measurement in non sinusoidal conditions. In: Proceedings of the 2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), IEEE; 2015, pp. 1255–1260.
[85] Barbaro PV, Cataliotti A, Cosentino V, Nuccio S. Behaviour of reactive energy meters in polluted power systems. In: XVIII IMEKO World Congress, Metrology for a Sustainable Development, Rio de Janeiro, Brazil, Vol. 172; 2006.
[86] Berrisford AJ. Smart meters should be smarter. In: 2012 IEEE Power and Energy Society General Meeting; 2012, pp. 1–6.
[87] Ferrero A. Measuring electric power quality: problems and perspectives. Measurement 2008;41(2):121–9.
[88] Cataliotti A, Cosentino V, Di Cara D, Lipari A, Nuccio S. A DAQ-based sampling wattmeter for IEEE Std. 1459–2010 powers measurements. uncertainty evaluation in nonsinusoidal conditions. Measurement 2015;61:27–38.
[89] Budeanu C. Reactive and fictitious powers. Rom Natl Inst 1927;2.
[90] Fryze S. Active, reactive and apparent powers in circuits with distorted voltages and currents. Elektro Z 1932;53(25–27):596–9.
[91] Cataliotti A, Cosentino V, Nuccio S. Static meters for the reactive energy in the presence of harmonics: an experimental metrological characterization. IEEE Trans Instrum Meas 2009;58(8):2574–9.
[92] Cataliotti A, Cosentino V, Lipari A, Nuccio S. On the calibration of reactive energy meters under non sinusoidal conditions. XIX IMECO Kongr Fundam Appl Metrol 2009:719–23.
[93] IEEE Std. 1459–2010: IEEE standard definitions for the measurement of electric power quantities under sinusoidal, Nonsinusoidal, Balanced or Unbalanced Conditions.
[94] Arseneau R, So E. Calibration services in support of smart grid applications. in: 2012 IEEE Power and Energy Society General Meeting, IEEE; 2012, pp. 1–4.
[95] Barbaro PV, Cataliotti A, Cosentino V, Nuccio S. A novel approach based on nonactive power for the identification of disturbing loads in power systems. IEEE Trans Power Deliv 2007;22(3):1782–9.
[96] O'Driscoll E, O'Donnell GE. Industrial power and energy metering-a state-of-the-art review. J Clean Prod 2013;41:53–64.
[97] Nuccio S, Spataro C, Monte Carlo A. method for the auto-evaluation of the uncertainties in analog-to-digital conversion-based measurements. COMPEL: Int J Comput Math Electr Electron Eng 2004;23(1):148–58.
[98] Spataro C. ADC based measurements: a common basis for the uncertainty estimation. In: Proceedings of the 17th Symposium IMEKO TC 4, 3rd Symposium IMEKO TC 19 and 15th IWADC Workshop: Instrumentation for the ICT Era, Kosice, Slovakia, 2010.
[99] Korczynski MJ, Hetman A. A calculation of uncertainties in virtual instrument. In: Proceedings of the IEEE Instrumentation and Measurement Technology Conference, 2005. IMTC 2005, vol. 3, IEEE, 2005, pp. 1697–1701.
[100] Ferrero A, Gamba R, Salicone S. A method based on random-fuzzy variables for online estimation of the measurement uncertainty of dsp-based instruments. IEEE Trans Instrum Meas 2004;53(5):1362–9.
[101] Cataliotti A, Cosentino V, Di Cara D, Lipari A, Nuccio S, Spataro C. Uncertainty evaluation in power measurements with commercial data acquisition boards. In: Proceedings of the International Workshop on ADC Modelling, Testing and Data Converter Analysis and Design and IEEE 2011 ADC Forum; 2011.
[102] Femine AD, Gallo D, Landi C, Luiso M. Advanced instrument for field calibration of electrical energy meters. IEEE Trans Instrum Meas 2009;58(3):618–25.
[103] Amicone D, Bernieri A, Ferrigno L, Laracca M. A smart add-on device for the remote calibration of electrical energy meters. In: Proceedings of the IEEE Instrumentation and Measurement Technology Conference, 2009. I2MTC'09, IEEE; 2009, pp. 1599–1604.
[104] Crenna F, Rossi GB, Bovio L. Probabilistic measurement evaluation for the implementation of the measuring instrument directive. Measurement 2009;42(10):1522–31.
[105] Pendrill L. Optimised measurement uncertainty and decision-making when sampling by variables or by attribute. Measurement 2006;39:829–40.
[106] Calamaro N, Elkonin V, Beck Y, Shmilovitz D. Electric energy metering developments at the smart grid: technology, accuracy, standardization, and verification, Tech. rep., Society for Electrical and Electronics Engineers in Israel; 2013. URL ⟨http://www.seeei.org/el2013/article/1030-art.pdf⟩.
[107] D'Apice B, Gallo D, Landi C, Rignano N. Distributed laboratory for metrological confirmation of power quality instruments. In: Proceedings of the 2005 IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2005. VECIMS 2005; 2005, pp. 13 – 15.
[108] Sun Y, Augenbroe G. Urban heat island effect on energy application studies of office buildings. Energy Build 2014;77:171–9.
[109] Fisk WJ, Faulkner D, Sullivan DP. Accuracy of co2 sensors, Tech. Rep. LBNL-1095E, Lawrence Berkeley National Laboratory; 2008.
[110] Agarwal Y, Balaji B, Gupta R, Lyles J, Wei M, Weng T. Occupancy-driven energy management for smart building automation. In: Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building, ACM; 2010, pp. 1–6.
[111] Yan D, O'Brien W, Hong T, Feng X, Gunay HB, Tahmasebi F, et al. Occupant behavior modeling for building performance simulation: current state and future challenges. Energy Build 2015;107:264–78.
[112] Hong T, Taylor-Lange S, D'Oca S, Yan D, Corgnati SP. Advances in research and applications of energy-related occupant behavior in buildings. Energy Build 2015;116:102–694.
[113] TUV NEL, Flare gas measurement using ultrasonic transit time: Good practice guide, Tech. rep., National Measurement System.
[114] Johnson A, et al., Measurement challenges and metrology for monitoring CO2 emissions from smokestacks - workshop summary, NIST Special Publication 1201, National Institute for Standards and Measurement; December 2015.
[115] Gueymard CA, Myers DR. Evaluation of conventional and high-performance routine solar radiation measurements for improved solar resource, climatological trends, and radiative modeling. Sol Energy 2009;83(2):171–85.
[116] Taxation Laws Amendment Act No. 25 of 2015, South African Government Gazette no. 39588; 2015.
[117] Mathieu JL, Price PN, Sila K, Piette MA. Quantifying changes in building electricity use, with application to demand response. IEEE Trans Smart Grid 2009;41(4):374–81.
[118] Sun Y, Heo Y, Tan M, Xie H, Wu CJ, Augenbroe G. Uncertainty quantification of microclimate variables in building energy models. J Build Perform Simul 2014;7(1):17–32. http://dx.doi.org/10.1080/19401493.2012.757368.
[119] Gustafson P. Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments. Boca Raton, Florida: CRC Press; 2003.
[120] Ree MJ, Carretta TR. The role of measurement error in familiar statistics. Organ Res Methods 2006;9(1):99–112.
[121] Regression for M & V: Reference guide, Tech. rep., Bonneville Power Administration; May 2012.
[122] Carroll RJ, Ruppert D. Transformation and Weighting in Regression, Vol. 30, CRC Press; 1988.
[123] Ridge R. Errors in variables: a close encounter of the third kind. In: Proceedings of the Energy Evaluation Conference, Chicago; 1997, pp. 479–487.
[124] Rivers N, Jaccard M, et al. Electric utility demand side management in canada. Energy J-Clevel 2011;32(4):93.
[125] Violette DM, Provencher B, Sulyma I. Assessing bottom-up and top-down approaches for assessing DSM programs and efforts. in: International Program Evaluation Conference, Rome; 2012.
[126] Huntley L. Establishing traceability for a high performance AC/DC transfer standard. ISA Trans 1990;29(4):41–7.
[127] Somppi J. A case study in characterizing & disciplining electrical calibrator instrumentation to improve test accuracies & measurement uncertainties. In: Proceedings of the NCSL International Workshop and Symposium; 2007.
[128] Schumacher RB. Systematic measurement errors. J Qual Technol 1981;13(1):10–24.
[129] Deaver D, Everett W., How to maintain your confidence. In: Proceedings 1993 NCSL Workshop and Symposium; 1993.
[130] Wong W. What TUR do you really need? Putting statistical theory into practice, Tech. rep., Fluke Corporation.
[131] Deaver D. Managing calibration confidence in the real world. In: NCSL Workshop and Symposium; 1995.
[132] Deaver D. Guardbanding in the world of ISO Guide 25. In: NCSL Workshop and Symposium; 1998.
[133] Eagle AR. A method for handling errors in testing and measuring. Ind Qual Control 1954;10(3):10–5.
[134] Rossi GB, Crenna F. A probabilistic approach to measurement-based decisions. Measurement 2006;39(2):101–19.
[135] ISO 14253-1, 1998: Inspection by measurement of workspieces and measuring equipment; part 1: Decision rules for proving conformance or nonconformance

with specifications,.
[136] Phillips SD, Estler WT, Levenson MS, Eberhardt KR. Calculation of measurement uncertainty using prior information. J Res Natl Inst Stand Technol 1998;103(6):625–32.
[137] Fuller WA. Measurement error models. Wiley-Interscience; 2006.
[138] Carobbi C, Pellicci G, Vieri S. Error modeling of static energy meters. In: XIX IMEKO World Congress, Fundamental and Applied Metrology; 2009, pp. 820–824.
[139] Küchenhoff H, Carroll R. Segmented regression with errors in predictors: semi-parametric and parametric methods. Stat Med 1997;16(2):169–88.
[140] Kennedy MC, O'Hagan A. Bayesian calibration of computer models. J R Stat Soc. Ser. B (Methodol.) 2001;63(Part 3):425–646.
[141] Burkhart MC, Heo Y, Zavala VM. Measurement and verification of building systems under uncertain data: a Gaussian process modeling approach. Energy Build 2014;75:189–98.
[142] Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. 2. Boca Raton, Florida: Taylor & Francis; 2014.
[143] Mathieu JL, Callaway DS, Kiliccote S. Variability in automated responses of commercial buildings and industrial facilities to dynamic electricity prices. Energy Build 2011;43(12):3322–30.
[144] Kruschke J. Doing Bayesian data analysis: a tutorial with R, JAGS, and Stan. 2nd edition New York City, New York: Academic Press; 2015.
[145] Pavlak GS, Florita AR, Henze GP, Rajagopalan B. Comparison of traditional and Bayesian calibration techniques for gray-box modeling. J Archit Eng 2013;20(2). [04013011-2-16].
[146] Kucukelbir A, Tran D, Ranganath R, Gelman A, Blei DM. Automatic differentiation variational inference, arXiv:1603.00788.
[147] Robert C, Casella G. Monte Carlo statistical methods. New York City, New York: Springer Science & Business Media; 2013.
[148] Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, Brubaker MA, Guo J, Li P, Riddell A. Stan: A probabilistic programming language, Journal of Statistical Software; In Press.
[149] Salvatier J, Wiecki TV, Fonnesbeck C. Probabilistic programming in Python using PyMC3, PeerJ Computer Science 2. http://dx.doi.org/10.7717/peerj-cs.55.
[150] Mills E, Friedman H, Powell T, Bourassa N, Claridge D, Haasl T, Piette MA. The cost-effectiveness of commercial-buildings commissioning, Tech. Rep. LBNL-56637 (Rev.), Lawrence Berkeley National Laboratory; December 2004.
[151] Rysanek A, Choudhary R. Optimum building energy retrofits under technical and economic uncertainty. Energy Build 2013;57:324–37. http://dx.doi.org/10.1016/j.enbuild.2012.10.027.
[152] Friege J, Chappin E. Modelling decisions on energy-efficient renovations: a review. Renew Sustain Energy Rev 2014;39:196–208.
[153] Lee P, Lam P, Lee W. Risks in Energy Performance Contracting (EPC) projects. Energy Build 2015;92:116–27.
[154] Mills E, Kromer S, Weiss G, Mathew PA. From volatility to value: analysing and managing financial and performance risk in energy savings projects. Energy Policy 2006;34(2):188–99.
[155] Boxer E, Henze GP, Hirsch AI. A model-based decision support tool for building portfolios under uncertainty. Autom Constr 2017;78:34–50.
[156] Heo Y. Bayesian calibration of building energy models for energy retrofit decision-making under uncertainty. Atlanta, Georgia: Georgia Institute of Technology; 2011. Ph.D. Thesis.
[157] Xia X, Zhang J. Mathematical description for the measurement and verification of energy efficiency improvement. Appl Energy 2013;111:247–56.
[158] Soroudi A, Amraee T. Decision making under uncertainty in energy systems: state of the art. Renew Sustain Energy Rev 2013;28:376–84.
[159] Olinga Z, Xia X, Ye X. A cost-effective approach to handle measurement and verification sampling and modelling uncertainties, submitted to Applied Energy.
[160] Carstens H, Xia X, Yadavalli S. Efficient combined metering and survey sampling longitudinal measurement and verification designs for lighting retrofit projects, submitted for review to Energy and Buildings on 13 April 2017.
[161] Kammerud R, Gillespie KL, Hydeman M. Economic uncertainties in chilled water system design. ASHRAE Trans 1999;105:1075–85. (SE-99-16-3).
[162] Irwin LA. White paper: A high accuracy standard for electricity meters, Tech. rep., Schneider Electric; 2011.
[163] Jackson J. Promoting energy efficiency investments with risk management decision tools. Energy Policy 2010;38:3865–73.
[164] Burhenne S, Tsvetkova O, Jacob D, Henze GP, Wagner A. Uncertainty quantification for combined building performance and cost-benefit analyses. Build Environ 2013;62:143–54.
[165] Lee P, Lam P, Yik FW, Chan EH. Probabilistic risk assessment of the energy saving shortfall in energy performance contracting projects-a case study. Energy Build 2013;66:353–63.
[166] Deng Q, Jiang X, Cui Q, Zhang L. Strategic design of cost savings guarantee in energy performance contracting under uncertainty. Appl Energy 2015;139:68–80.
[167] Deng Q, Jiang X, Zhang L, Cui Q. Making optimal investment decisions for energy service companies under uncertainty: a case study. Energy 2015;88:234–43.
[168] Pendrill L, Källgren H. Exhaust gas analysers and optimised sampling, uncertainties and costs. Accrédit Qual Assur 2006;11(10):496–505.
[169] Fearn T, Fisher SA, Thompson M, Ellison SL. A decision theory approach to fitness for purpose in analytical measurement. Analyst 2002;127(6):818–24.
[170] Reddy T. Literature review on calibration of building energy simulation programs: uses, problems, procedures, uncertainty, and tools. ASHRAE Trans 2006;112:226–40. (Part 1).
[171] Sun J, Reddy TA. Calibration of building energy simulation programs using the analytic optimization approach (rp-1051). HVACR Res 2006;12(1):177–96.
[172] Reddy TA, Maor I, Panjapornpon C. Calibrating detailed building energy simulation programs with measured data - Part I: general methodology (RP-1051). HVAC & R Res 2007;13(2):221–41.
[173] Reddy TA, Maor I, Panjapornpon C. Calibrating detailed building energy simulation programs with measured data - part II: application to three case study office buildings. HVACR Res 2007;13(2):243–65.
[174] Coakley D, Raftery P, Keane M. A review of methods to match building energy simulation models to measured data. Renew Sustain Energy Rev 2014;37:123–41.
[175] Lee BD, Sun Y, Augenbroe G, Paredis CJ. Toward better prediction of building performance: a workbench to analyze uncertainty in building simulation. In: BS2013: Proceedings of the 13th Conference of the International Building Performance Simulation Association, Chambery, France; 2013.
[176] Tian W. A review of sensitivity analysis methods in building energy analysis. Renew Sustain Energy Rev 2013;20:411–9.
[177] Sanyal J, New J, Edwards RE, Parker L. Calibrating building energy models using supercomputer trained machine learning agents. Concurr Comput: Pract Exp 2014;26(13):2122–33.
[178] Riddle M, Muehleisen RT. A guide to Bayesian calibration of building energy models. In: Proceedings of the Building Simulation Conference; 2014.
[179] Heo Y, Graziano DJ, Guzowski L, Muehleisen RT. Evaluation of calibration efficacy under different levels of uncertainty. J Build Perform Simul 2015;8(3):135–44.
[180] Wilkinson RD. Large-Scale Inverse Problems and Quantification of Uncertainty, John Wiley & Sons Ltd., Chichester, UK, 2010, Ch. Bayesian Calibration of Expensive Multivariate Computer Systems.
[181] Heo Y, Choudhary R, Augenbroe G. Calibration of building energy models for retrofit analysis under uncertainty. Energy Build 2012;47:550–60.
[182] Heo Y, Zavala VM. Gaussian process modeling for measurement and verification of building energy savings. Energy Build 2012;53:7–18.
[183] Heo Y, Augenbroe G, Choudhary R. Quantitative risk management for energy retrofit projects. J Build Perform Simul 2013;6(4):257–68.
[184] Heo Y, Augenbroe G, Graziano D, Muehleisen RT, Guzowski L. Scalable methodology for large scale building energy improvement: relevance of calibration in model-based retrofit analysis. Build Environ 2015;87:342–50.
[185] Li Q, Augenbroe G, Brown J. Assessment of linear emulators in lightweight Bayesian calibration of dynamic building energy models for parameter estimation and performance prediction. Energy Build 2016;124:194–202.
[186] Tian W, Yang S, Li Z, Wei S, Pan W, Li Y. Identifying informative energy data in Bayesian calibration of building energy models. Energy Build 2016;119:363–76.
[187] De Wit S, Augenbroe G. Analysis of uncertainty in building design evaluations and its implications. Energy Build 2002;34(9):951–8.
[188] Tian W, Choudhary R. A probabilistic energy model for non-domestic building sectors applied to analysis of school buildings in greater london. Energy Build 2012;54:1–11.
[189] Booth A, Choudhary R. Decision making under uncertainty in the retrofit analysis of the UK housing stock: implications for the Green Deal. Energy Build 2013;64:292–308.
[190] Sun Y, Gu L, Wu CF Jeff, Augenbroe G. Exploring HVAC system sizing under uncertainty. Energy Build 2014;81:243–52.
[191] Lee P, Lam P, Lee W, Chan E. Analysis of an air-cooled chiller replacement project using a probabilistic approach for energy performance contracts. Appl Energy 2016;171:415–28.

energies

MDPI

Article

Bayesian Energy Measurement and Verification Analysis

Herman Carstens [1,*], Xiaohua Xia [1] and Sarma Yadavalli [2]

[1] Centre for New Energy Systems, University of Pretoria, Pretoria 0002, South Africa; xxia@up.ac.za
[2] Department of Industrial and Systems Engineering, University of Pretoria, Pretoria 0002, South Africa; sarma.yadavalli@up.ac.za
* Correspondence: herman.carstens@up.ac.za; Tel.:+27-12-420-6470

Received: 5 December 2017; Accepted: 18 January 2018; Published: 6 February 2018

Abstract: Energy Measurement and Verification (M&V) aims to make inferences about the savings achieved in energy projects, given the data and other information at hand. Traditionally, a frequentist approach has been used to quantify these savings and their associated uncertainties. We demonstrate that the Bayesian paradigm is an intuitive, coherent, and powerful alternative framework within which M&V can be done. Its advantages and limitations are discussed, and two examples from the industry-standard International Performance Measurement and Verification Protocol (IPMVP) are solved using the framework. Bayesian analysis is shown to describe the problem more thoroughly and yield richer information and uncertainty quantification results than the standard methods while not sacrificing model simplicity. We also show that Bayesian methods can be more robust to outliers. Bayesian alternatives to standard M&V methods are listed, and examples from literature are cited.

Keywords: statistics; uncertainty; regression; sampling; outlier; probabilistic

---

1. Introduction

This study argues for the adoption of the Bayesian paradigm in energy Measurement and Verification (M&V) analysis. As such, no new Bayesian methods will be developed. Instead, the advantages, limitations, and application of the Bayesian approach to M&V will be explored. Since the focus is on application, a full explanation of the underlying theory of the Bayesian paradigm will not be given. Readers are referred to Sivia and Skilling [1] or Kruschke [2] for a basic introduction, or von der Linden et al. [3] or Gelman et al. [4] for more complete treatments.

The argument made below is not that current methods are completely wrong or that the Bayesian paradigm is the only viable option, but that the field can benefit from a increased adoption of Bayesian thinking because of its ease of implementation and accuracy of the results.

This paper is arranged as follows. After discussing the background of current M&V analysis methods and the opportunities for improvement in Section 1.1, the Bayesian paradigm is introduced and its practical benefits and some caveats are discussed in Section 2. Section 3 offers two well-known examples and their Bayesian solutions. We also discuss robustness and hierarchical modelling. Section 4 gives a reference list of Bayesian solutions to common M&V cases.

1.1. Background

M&V is the discipline in which the savings from energy efficiency, demand response, and demand-side management projects are quantified [5], based on measurements and energy models. A large proportion of such M&V studies quantify savings for building projects, both residential and commercial. The process usually involves taking measurements or sampling a population to create a baseline, after which an intervention is done. The results are also measured, and the savings are inferred as the difference between the actual post-intervention energy use, and what it would have been, had no intervention taken place. These savings are expressed in probabilistic terms following the International Standards Organization (ISO) Guide to the Expression of Uncertainty in Measurement (GUM) [6]. M&V study results often form the basis of payment decisions in energy performance contracts, and the risk-implications of such studies are therefore of interest to decision makers.

The Bayesian option will not affect the foundational M&V methodologies such as retrofit isolation or whole facility measurement, but only the way the data are analysed once one of these methods has been decided upon.

M&V guidelines such as the International Performance Measurement and Verification Protocol (IPMVP) [5], the American Society of Heating, Refrigeration, and Air Conditioning Engineers (ASHRAE)'s Guideline 14 on Measurement of Energy, Demand, and Water Savings [7], or the United States Department of Energy's Uniform Methods Project (UMP) [8], as well as most practitioners, use frequentist (or classical) statistics for analysis. Because of its popularity in the twentieth century, most practitioners are unaware that this is only one statistical paradigm and that its assumptions can be limiting. The term 'frequentist' derives from the method that equates probability with long-run frequency. For coin flips or samples from a production line, this assumption may be valid. However, for many events, equating probability with frequency seems strained because a large, hypothetical long-run population needs to be imagined for the probability-as-frequency-view to hold. Kruschke [2] gives an example where a coin is flipped twenty times and seven heads are observed. The question is then: what is the probability of the coin being fair? The frequentist answer will depend on the imagined population from which the data were obtained. This population could be obtained by "stopping after 20 flips", but it could also be "stopping after seven heads" or "stopping after two minutes of flipping" or "to compare it to another coin that was flipped twenty times". In each case, the probability that it is a fair coin changes, even though the data did not—termed *incoherence* [9]. In fact, the probabilities are dependent on the analyst's *intention*. By changing his intention, he can alter the probabilities. This problem becomes even more severe in real-world energy savings inference problems with many more factors. The hypothetical larger population from which the energy use at a specific time on a specific day for a specific facility was sampled is difficult to imagine. That is not to say that a frequentist statistical analysis cannot be done, or be useful. However, it often does not answer the question that the analyst is asking, committing an "error of the third kind". Analysts have become used to these 'statistical' answers (e.g., "not able to reject the null hypothesis"), and have accepted such confusion as part of statistics. For example, consider two mainstays of frequentist M&V: confidence intervals (CIs) and p-values. CIs are widely used in M&V to quantify uncertainty. According to Neyman, who devised these intervals, they do not convey a degree of belief, or confidence, as is often thought. Frequentist confidence intervals are produced by a method that yields an interval that contains the true value only in a specified percentage (say 90%) of cases [10]. This may seem like practically the same thing, but an explanation from most frequentist statistics textbooks will then seem very confusing. Consider Montgomery and Runger's *Applied Statistics and Probability for Engineers* [11], under "Interpreting a Confidence Interval" (CI). They explain that, with frequentist CIs, one cannot say that the interval contains the true number with a probability of e.g., 90%. The interval either contains the value, or it does not. Therefore, the probability is either zero or one, but the analyst does not know which. Therefore, the interval cannot be associated with a probability. Furthermore, it is a *random interval* (emphasis theirs) because the upper and lower bounds of the interval are random variables.

Consider now the p-value. Because of the confusion surrounding this statistic, the American Statistical Association issued a statement regarding its use [12], in which they state that p-values neither signify probabilities of the hypothesis being true or false, nor are they probabilities that the result arose by chance. They go on to say that business (or policy) decisions should not be based on p-value thresholds. p-values do not measure effect sizes or result importances, and by themselves are not adequate measures of evidence.

Such statements by professional statisticians leave most M&V practitioners justifiably confused. It is not that these methods are invalid, but that they have been co-opted to answer different *kinds* of questions to what they actually answer. The reason for their popularity in the 20th century has more to do with their computational ease, compared to the more formal and mathematical Bayesian methods, than with their appropriateness. The Bayesian conditional-probability paradigm is much older than the frequentist one but used to be impractical for computational reasons. However, with the rise in computing power and new numeric methods for solving Bayesian models, this is no longer a consideration.

2. The Bayesian Paradigm

Instead of approaching uncertainty in terms of long-run frequency, the Bayesian paradigm views uncertainty as a state of knowledge or a degree of belief, the sense most often meant by people when thinking about uncertainty. These uncertainties are calculated using conditional-probability logic and calculus, proceeding from first principles. For example, consider two conditions M and S. Let $\Pr()$ denote a probability and $|$ "conditional on" or "given". Furthermore, let I be the background information about the problem. Bayes' theorem states that:

$$\Pr(S|M, I) = \frac{\Pr(M|S, I)\Pr(S|I)}{\Pr(M|I)}. \tag{1}$$

Now, as stated previously, M&V is about verifying the savings achieved, based on some measurements and an energy model, and quantifying the uncertainty in this figure. If we let S be the savings, and M the measurements, Bayes' theorem as stated above answers that question exactly: it supplies a probability of the savings given the measurements and any background information that might be available; $\Pr(S|M)$. Bayes' theorem is, therefore, the natural expression of the M&V aim:

$$\text{Verification}|\text{Measurement} \equiv \Pr(S|M).$$

Whereas the frequentist paradigm views the data as random realisations of a process with fixed parameters, the Bayesian paradigm views the data (measurements) as fixed, and the underlying parameters as uncertain (thereby avoid the incoherence of the coin flip example [9]). This seems like a trivial distinction at first but is significant: the frequentist only solves for $\Pr(M|S)$: the probability of observing that data, given the underlying savings value. However, that is not the question M&V seeks to answer. In the frequentist paradigm, the analyst does not invert this as Bayes' theorem does to find the probability distribution on the savings, given the data. Therefore, in the frequentist case, the wrong question is being answered, as alluded to above (Technical note: to be fair, we note that, for constant priors, the likelihood may be equivalent to the posterior. When it is the case, the frequentist likelihood may borrow from Bayesian theory and be interpreted as a probability).

It is this inversion process that has often been intractable in higher dimensions until the advent of Markov Chain Monte Carlo (MCMC) techniques and increased computing power (Technical note: other Monte Carlo-based inversion techniques such as rejection or importance sampling are only efficient enough to be practical in low-dimensional settings. Note that we use Monte Carlo here in the sense of a straightforward sense of generating random numbers according to standard distributions [13]). MCMC software has allowed users to specify a model (e.g., a linear regression model), supply the observations or data (measurements), and infer the values on the model parameters *probabilistically*. This is called probabilistic programming. Probabilistic programming is compelling because, instead of working with point estimates on all unknown parameters (e.g., slope and intercept in a straight-line regression model), one describes the system in terms of probability distributions. Working with probability distributions rather than point estimates is preferable, since it is well known that doing calculations with point estimates can lead to erroneous conclusions [14]. When doing forward-calculations as illustrated in Figure 1, it is therefore desirable to use distributions on unknown variables and then apply a Monte Carlo simulation or Mellin Transform Moment Calculation method [15,16] to obtain a probability distribution on the result. MCMC allows one to do the inverse: inferring parameter distributions from given data and a model. Therefore, MCMC is to regression what Monte Carlo simulation is to deterministic computation. The adoption of the Bayesian paradigm therefore allows the analyst to move from deterministic to probabilistic M&V, as shown in Figure 1.

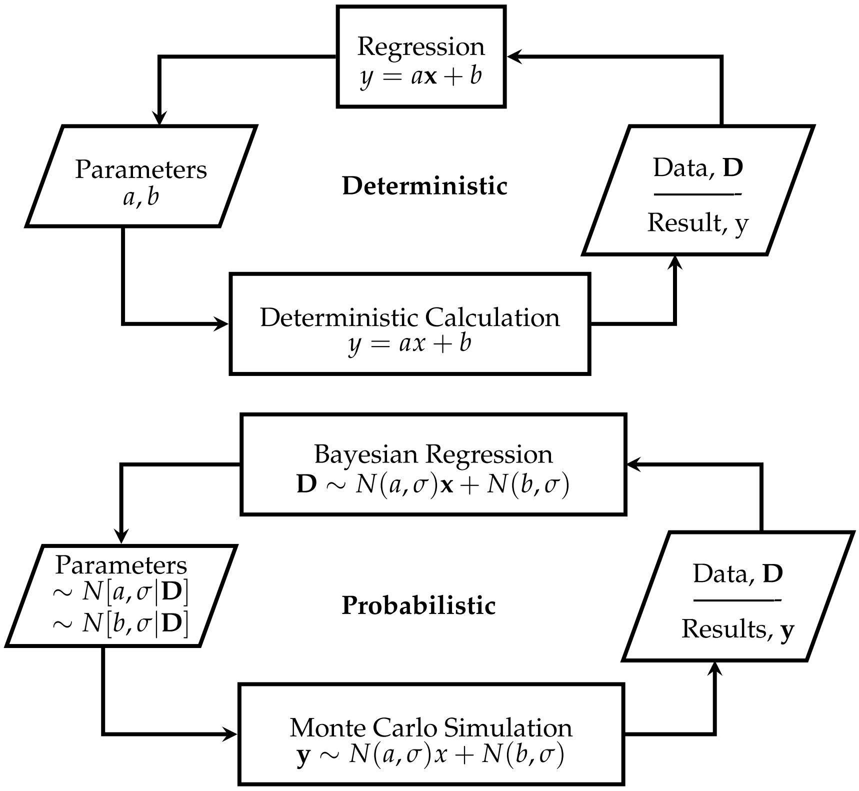


Figure 1. Deterministic and probabilistic calculation, simulation, and inverse modelling. The notation $\sim N[\cdot]$ denotes a normal distribution as a convenient substitute for any distribution. Note that this figure does *not* illustrate or recommend a cyclic work flow; usually, only one of the for processes is of interest for a particular problem. Indeed, continually updating, or "fiddling", a Bayesian prior based on the posterior (i.e., treating the illustration as a cycle) is poor modelling practice. We recommend that M&V analysts set, state, and defend their prior, and not change it to achieve a different outcome.

For the inversion described above to work, the $\Pr(S|I)$ term, called the prior, needs to be specified. Although the prior can be used to incorporate information into the model, which is not available through the data alone, it is, in essence, merely a mathematical device allowing inversion. The prior is often specified as "non-informative"—a flat probability distribution over the region of interest, allowing the data to "speak for itself" through the likelihood term. This will be discussed in more detail below. The other term, $\Pr(M|I)$, need not be specified in numeric MCMC models—it is a normalising factor ensuring that the right-hand side of the equation can integrate to unity, making it a proper probability density function (Technical note: this term becomes important in more sophisticated Bayesian analyses where model selection or experimental design is done [1]). The left-hand side of the equation is called the posterior distribution and is proportional, therefore, to the product of the prior and the likelihood.

Advanced Bayesian models may be nuanced, but the fundamental mechanics as described above stay the same for all Bayesian analyses: specify priors, describe the likelihood, and solve to find the posterior on the parameters of interest.

2.1. Practical Benefits

Besides the theoretical attractiveness discussed above, the Bayesian paradigm also offers many practical benefits for energy M&V:

1. Because Bayesian models are probabilistic, uncertainty is automatically and exactly quantified.
2. Uncertainty calculations in the Bayesian approach can be much less conservative than standard approaches. Shonder and Im [17] show a 40% reduction in uncertainty in one case. Since project payment is often dependent on savings uncertainties being within certain bounds, using the Bayesian approach can increase project feasibility.
3. By making the priors and energy model explicit, the Bayesian approach ensures greater transparency—one of the five key principles of M&V [5].
4. The Bayesian approach is widely used and is rapidly gaining popularity in other scientific fields. Lira [18] relates that even the GUM (adopted by many societies of physics, chemistry, electrotechnics, etc.) is being rewritten to be more consistent with this approach. Since M&V reports uncertainty according to the GUM, Bayesian calculations would be useful.
5. Bayesian models are more universal and flexible than standard methods. Bayesian modelling can be highly sophisticated, but the core of probabilistic thinking is consistent throughout. This is different to frequentist statistics where knowledge of one or even many tests will not necessarily aid the analyst in understanding a new metric, or approach to a problem not seen before. Many frequentist tests are ad hoc and apply only to specific situations. For example, *t*-tests have little to do with regression in frequentism, but, in Bayesian thinking, they are expressions of the same idea.
6. Being modular, Bayesian modelling is more flexible. Ordinary least squares (OLS) linear regression assumes residuals are normally distributed and that the variance is constant for all points. In a probabilistic Bayesian model, the parameters can be distributed according to any distribution, but the posterior for each will be determined by the data (if the prior is appropriately chosen). Models are also modular and can be designed to suit the problem. For example, it is no different to create terms for serial correlation, or heteroscedasticity (non-constant variance) than it is to specify an ordinary linear model. This also allows for easy specification of non-routine adjustments, the handling of missing values, and the incorporation of unmeasured yet important quantities such as measurement error, often problematic for energy models. For the retrofit isolation with a key parameter measurement approach, the unmeasured parameters (the estimates) can also be incorporated in this way.
7. Bayesian models can account for model-selection uncertainty. There are often multiple reasonable energy models which could describe a specific case—for example: time and dry-bulb temperature; occupancy and dry-bulb temperature; temperature, humidity, and occupancy, etc. The analyst usually chooses one model, discards the rest, and reports the uncertainty produced in that specific model. However, this uncertainty does not account for model selection. In other words, there is an uncertainty associated with choosing that specific model above another reasonable one. Bayesian model averaging allows many models to be specified simultaneously, and averages their results by automatically weighting each model's influence on the final result by that model's explanatory power. This gives a far more realistic uncertainty value [4].
8. Because uncertainty is automatically quantified, CIs can be interpreted in the way most people understand them: degrees of belief about the value of the parameter.
9. The Bayesian approach is well-suited to "small data" problems. This seems like a minor point in developed countries where questions surrounding big data are more pressing. However, big (energy) data is a decidedly "first-world problem". In developing countries, a lack of meters makes M&V expensive, and it is useful to have a method that is consistent on smaller data sets as well.
10. Bayesian approaches allow real-time or online updating of estimates [19–21]. For many other machine learning techniques, the data need to be split into testing and training sets, the model

trained on the training set, and then used to predict the testing set period. As new data becomes available, the model needs to be retrained in many cases (Technical note: Artificial Neural Networks (ANNs), stochastic gradient descent and passive-aggressive algorithms, as well as Dynamic Linear Models can also be updated online), making it computationally expensive to keep a model updated. In a Bayesian paradigm, previous data can be summarised by the prior so that the model need not be retrained.

11. The Bayesian approach allows for the incorporation of prior information where appropriate. The danger in this will be discussed in Section 2.2. However, in cases where it is warranted, known values or ranges for certain coefficients can be specified in the prior. This has been done successfully for energy projects [22–25]. Prior information is also useful in longitudinal studies, where measurements or samples from previous years can be taken into account [20,21].
12. When the savings need to be calculated for "normalised conditions", for example, a 'typical meteorological year', rather than the conditions during the post-retrofit monitoring period, it is not possible to quantify uncertainty using current methods. However, Shonder and Im [17] have shown that it can be naturally and easily quantified using the Bayesian approach.

2.2. Caveats

The Bayesian approach also comes with certain caveats that M&V practitioners and policy makers should bear in mind.

1. Modelling is non-generic. In point 5 above, it was stated that the Bayesian approach is more universal. This is true in the sense that the same basic approach is used for many different kinds of problems. However, the inherent modularity of the method as described in point 6 means that there is not a one-size-fits-all generic template in Bayesian modelling, the way there usually is in frequentist modelling. This necessitates more thinking from the analyst. However, we believe this to be an advantage: frequentist approaches make it easier to think less, but as a consequence, also to build poor models, which has led to the current replication crisis seen in research [26] and a general mistrust of statistical results [27]. High quality models require some thought and care, in any paradigm.
2. As with any method, it is not immune to abuse. The most popular criticism is that, by having a prior distribution on the savings, the posterior may be biased in a way not warranted by the data, making the result subjective. This is certainly possible. However, having a prior in an M&V analysis is actually an advantage.

    (a) As stated above, it allows for greater modelling transparency. The Bayesian form forces the analyst to be explicit about his or her modelling assumptions, and to defend them. Such assumptions cannot be imported by (accidentally or purposefully) choosing one test over another, as in the frequentist case.
    (b) It is sometimes necessary to include priors to *avoid* bias. Ioannidis [28] and Button [29] have shown that many medical studies contain false conclusions due to biased results. The bias that was introduced was to consider positive and negative outcomes from a clinical trial equally likely. However, the prior odds of an experimental treatment working is much lower than the odds of that treatment not working. Ignoring these prior odds leads to a high false-positive rate, since many of the positive results are actually false and due to noise. In M&V, the situation is reversed: the prior odds of energy projects saving energy are high. Having a neutral prior would therefore bias a result towards conservatism (Technical note: conservatism is one of the key principles of M&V [5], but we do not hereby advocate for neutral priors in all cases). Nevertheless, Button's study is an excellent illustration of why priors are an important part of probability calculus.
    (c) Because the assumptions and distributions used are clearly stated, it precludes hedging the M&V result with phrases such as "however, from previous studies/experience, we know

that this is a conservative figure . . . ”. Because the prior was stated and defended at the outset, the final result should already incorporate it and should not be hedged.

(d) The thorough analyst will test the effect of different priors on the posterior, demonstrating the bias introduced through his modelling assumptions, and justifying its use.

3. Bayesian methods can be computationally expensive for large datasets and complex models. It is true that numerical solvers are becoming more efficient and computational power is increasing. However, in comparison with matrix inversion techniques used for linear regression, for example, Bayesian methods are much slower and may be inappropriate for real-time applications [30].
4. The forecasting accuracy of other machine learning (ML) methods can be higher than regression in some cases [31,32], although regression-based approaches such as time-of-week-and-temperature [33] still perform very well [32,34] and may be preferred for simplicity. Note that this is a limitation of *regression*, not the overall Bayesian paradigm, although regression is the way most M&V analysts would use Bayesian methods. Many ML techniques also have Bayesian approaches, for example Bayesian tree-based ensemble methods [35] or Bayesian Artificial Neural Networks [36,37]. It also depends on the problem: it is not possible to know beforehand which model will work the best [38]. ML algorithms without Bayesian implementations also still only produce point estimates. Therefore, they cannot be compared to the full probabilistic approach, which provides much richer information and is not just a forecasting technique, but a full inference paradigm.
5. The parametric from of the model needs to be specified. Parametric Bayesian models as described in most of this study can only be correct in so far as their functional form describes the underlying physical process. Functional form misspecification is a real possibility. This is different to the machine learning methods described in the previous paragraph, which do not rely on a functional form being specified. Non-parametric models have their own benefits and limitations: for cases where the underlying physical process is well-understood, a parametric model can be more accurate. However, non-parametric methods such as Gaussian Processes (GPs) [22,39] or Gaussian Mixture Models [40] still require some model specification at a higher level (hyperparameters). GP models, for example, rely on an appropriate covariance function for valid inference. For more information on GPs for machine learning, see Rasmussen and Williams [41].

3. Bayesian M&V Examples

To demystify the Bayesian approach, two basic M&V calculations will be demonstrated. The reader will notice the recurring theme of expressing all variables as (conditional) probability distributions.

3.1. Sampling Estimation

Consider the following example from the IPMVP 2012 [5] (Appendix B-1). Twelve readings are taken by a meter. These are reported as monthly readings, but are assumed to be uncorrelated with any independent variables or other readings, and are therefore construed to be random samples. The values are:

$$\mathbf{D} = [950, 1090, 850, 920, 1120, 820, 760, 1210, 1040, 930, 1110, 1200]. \tag{2}$$

The units are not reported and the results below are therefore left dimensionless, although kWh would be a reasonable assumption. These data were carefully chosen, and have a mean $\mu = 1000$, sample standard deviation $s_s = 150$.

3.1.1. IPMVP Solution

The standard error is $SE = 43$. The confidence interval on the mean is calculated as:

$$CI = \mu \pm t \times SE. \tag{3}$$

Since $t_{90\%,11} = 1.80$, the 90% confidence interval on the mean was calculated as $1000 \pm 1.80 \times 43 = (933, 1077)$, or a 7.7% precision. Metering uncertainty is not considered in this calculation.

3.1.2. Bayesian Solution

The Bayesian estimate of the mean is calculated as follows. First, prior distributions on the data need to be specified. Vague priors will be used:

$$\Pr(\mu) \sim Uniform[0, 2000], \tag{4}$$

$$\Pr(\sigma) \sim Uniform[0, 1000]. \tag{5}$$

A t-distribution will be used for the likelihood below, and the degrees of freedom parameter (ν) of this distribution will, therefore, need to be specified. One could fix ν for the t-distribution at 12, since there are twelve data points and traditionally ν has been taken to signify this number. However, if outliers are present or if the data has more or less dispersion than the standard t-distribution with as many data points, this would not be realistic. It is therefore warranted to indicate the uncertainty in the data by specifying a prior distribution on ν also: a hyperprior. Kruschke [42] recommends an exponential distribution with the mean equal to the number of data points. This allows equal probability of ν being higher or lower than the default value:

$$\Pr(\nu) \sim Exponential[1/12]. \tag{6}$$

If the vector of the parameters is $\boldsymbol{\theta} = (\mu, \sigma, \nu)$, then the likelihood can be written as:

$$\Pr(\mathbf{D}|\boldsymbol{\theta}) \sim StudentT\left[\Pr(\mu),\ \Pr(\sigma),\ \Pr(\nu)\right]. \tag{7}$$

Note that the t-distribution is not used because of the t-test, but because its heavier tails are more accommodating of outliers. Any distribution could have been specified if there was good reason to do so. The posterior on μ is plotted in Figure 2. It was simulated in PyMC3 using the Automatic Differentiation Variational Inference (ADVI) algorithm with 100,000 draws, which is stable and converges on the posterior distribution in 10.76 s on a middle-range laptop computer. Although the mathematical notation may seem intimidating to practitioners who are not used to it, writing this in the probabilistic Python programming package PyMC3 [43] demonstrates the intuitive nature of such a model:

```
import pymc3 as pm
with pm.Model() as bayesian_sampling_model:
   # Hyperpriors and priors:
   mean = pm.Uniform('mean', 0, 2000)
   std = pm.Uniform('std', 0, 1000)
   nu = pm.Exponential('nu', 1/len(data))
   #  Likelihood
   likelihood = pm.StudentT('likelihood', mu=mean, sd=std, nu=nu, observed=data)
   # ADVI calculation
   trace = pm.variational.sample_vp(vparams=pm.variational.advi(n=100000))
```

It is important to note that no probability statements about the values inside the frequentist interval can be made, nor can one fit a distribution to the interval. The distribution indicated is strictly a Bayesian one. The Bayesian (highest density) interval is slightly wider than the frequentist confidence interval, at a precision of 8.5%. If ν were fixed at 12 (indicating that we are certain that the data does indeed reflect a t-distribution with 12 degrees of freedom exactly), Bayesian and frequentist intervals correspond exactly. However, the Bayesian alternative allows for a more realistic value.

With comparisons between two groups (two-sample t-tests), the effect of uncertainty in the priors becomes even more pronounced [42].

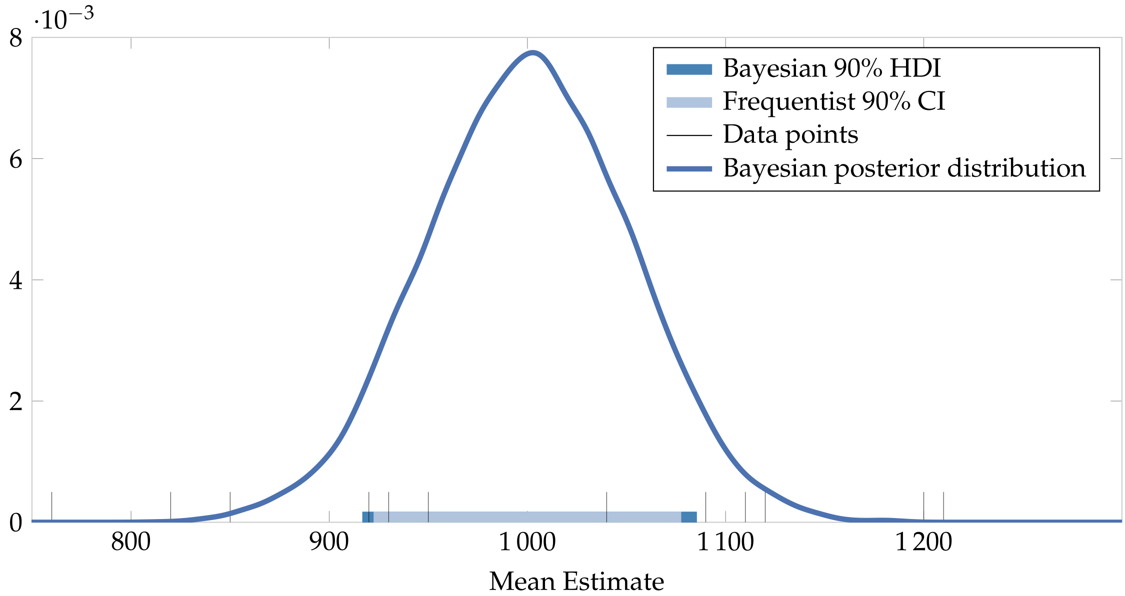


Figure 2. Illustration of Bayesian posterior density $\Pr(\mu|\mathbf{D})$, 90% Highest Density Interval (HDI), and frequentist 90% Confidence Interval (CI).

The posterior distribution can now be used to answer many interesting questions. For instance, what is the probability, given the data at hand, that the true mean is below 900? Or, is it safe to assume that the standard value of 950 is reflected by this sample, or should the null hypothesis be rejected? (If previous data to this effect is available, it could be included in the prior, maybe using the equivalent prior sample size method [44]). The frequentist may say that there is not enough evidence to reject the null, but cannot accept it either. In the Bayesian paradigm, 950 falls comfortably within the 90% confidence range, and can therefore be accepted at that level. As a further question, if this is an energy performance contracting project, and we assume that the data points are different facilities rather than different months, would it be worthwhile taking a larger sample to increase profits, if we believe that the true mean is 1100 (on which see Lindley [45], Bernardo [46] and Goldberg [47]).

It is therefore evident that the Bayesian result yields richer and more useful information using intuitive mathematics.

3.2. Regression

In M&V, one often uses the baseline data ($\mathbf{D}_b$) to infer the baseline (pre-retrofit) model parameters $\boldsymbol{\theta}$ through an inverse method:

$$\boldsymbol{\theta} = f^{-1}(\mathbf{D}_b, \tau), \tag{8}$$

where $f(\cdot)$ is a function relating the independent variables (energy governing factors) to the energy use of the facility, and τ is time. The model parameters describe the sensitivity of the energy model to the independent variables such as occupancy, outside air temperature, or production volume.

As an aside, this section will discuss an elementary parametric energy model using Bayesian regression, similar to standard linear regression. In practice, a two-parameter linear regression model seldom captures the different states of a facility's energy use, for example, heating at low temperatures, a comfortable range, and cooling at high temperatures. Piecewise linear regression techniques are often used [48–52], and they tend to work reasonably well if their assumptions are satisfied, but they are not stable in all cases, are approximate, and the assumptions are often restrictive. Shonder and Im [17] provide a Bayesian alternative. A non-parametric model using a Gaussian Process could also be used, and since one does not need to specify a parametric model, it allows very flexible models to be fit while still quantifying uncertainty. This is especially useful for models where energy use is a nonlinear function of the energy governing factors. However, to keep the example simple and focussed, only a simple parametric model will be considered below.

3.2.1. Example

Suppose one has a simple regression model where the energy use of a building **E** is correlated with the outside air temperature through the number of Cooling Degree Days (CDD). One cooling degree day is defined as an instance where the average daily temperature is one degree above the thermostat set point for one day, and the building therefore requires one degree of cooling (Technical note: a more accurate description would be the "building balance point", where the building's mass and insulation balance external forcings [53]). Let the intercept coefficient be θ_0, the slope coefficient θ_1, and the Gaussian error term $\boldsymbol{\epsilon}$. One could then write:

$$\mathbf{E} = \theta_0 + \theta_1 \mathbf{CDD} + \boldsymbol{\epsilon}. \tag{9}$$

In standard linear regression, one would write $\hat{\boldsymbol{\theta}}$ as the vector of two coefficients and do some linear algebra to obtain their estimates. There would be a standard error on each, which would indicate their uncertainties, and if the assumptions of linear regression, such as normality of residuals, independence of data, homoscedasticity, etc. hold, then it would be accurate. In Bayesian regression, one would describe the distributions on the parameters:

$$\Pr(\boldsymbol{\theta}|\mathbf{D}) \propto \Pr(\mathbf{D}|\boldsymbol{\theta}) \Pr(\boldsymbol{\theta}) \sim N[\hat{\boldsymbol{\theta}}, \boldsymbol{\sigma}], \tag{10}$$

where $\boldsymbol{\sigma}$ is the vector of the standard deviations on the estimates. Generating random pairs of values from the posterior, at a given value of CDD, according to the appropriate distributions, will yield the posterior predictive distribution. This is the distribution of energy use at a given temperature, or over the range of temperatures. Overlaying such realisations onto the actual data is called the posterior predictive check (PPC).

Now, consider a concrete example. The IPMVP 2012 [5] (Appendix B-6) contains a simple regression example of creating a baseline of a building's cooling load. The twelve data points themselves were not given, but a very similar data set yielding almost identical regression characteristics has been engineered and is shown in Table 1.

Table 1. Cooling Degree Day (CDD) Data for International Performance Measurement and Verification Protocol (IPMVP) Example B-6. Note that these data were reverse-engineered to yield the same regression results as reported in the IPMVP. The original data were not reported in the IPMVP.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
CDD	312	292	222	112	92	22	12	32	157	207	182	302
Energy Use	7823	7585	7486	6646	6185	5933	5381	5917	7158	7064	7231	8250

A linear regression model was fit to the data, and yielded the result shown in Table 2.

Table 2. Linear regression fit characteristics for data in Table 1. The coefficient of determination is $R^2 = 0.93$, which is identical to the IPMVP case. These results may be compared to Bayesian summary statistics in Table 3.

Parameter	Value	Standard Error	95% Interval
Slope coefficient	7.75	0.67	[6.26, 9.23]
Intercept coefficient	5634	129	[5347, 5921]

3.2.2. IPMVP Solution

The IPMVP then proceeds to calculate the uncertainty in the annual energy figure by multiplying the standard error on the estimate (the average standard error) by $t_{95\%}$ and the average consumption in the average month, and assumes that this value is constant for all months. As discussed in this study, this approach is problematic, and can at best be seen as approximate. Since it is treated in some detail in the IPMVP, the analysis will not be repeated here.

3.2.3. Bayesian Solution

The key to the Bayesian method is to approach the problem probabilistically, and therefore view all parameters in Equation (9) as probability distributions, and specify them as such. In this regression model, there are three parameters of interest: the intercept (θ_0), slope (θ_1), and the response (**E**). This response is the likelihood function, familiar to most readers as the frequentist approach. These distributions need to be specified in the Bayesian model. First, consider the priors on the slope and intercept. These can be vague. Technical note: the uniform prior on θ_0 in Equation (11) is actually technically incorrect: it may seem uniform in terms of gradient but is not uniform when the angle of the slope is considered. It is therefore not "rotationally invariant" and biases the estimate towards higher angles [54]. The correct prior is $Pr(\theta|I) \sim (1+\theta^2)^{-\frac{3}{2}}$; this is uniform on the slope angle. The reason that Equation (11) works in this case is that the exponential weight of the likelihood masks the effect. However, this is not always the case, and analysts should be careful of such priors in regression analysis:

$$\Pr(\theta_0) \sim Uniform[0, 10000], \tag{11}$$

and

$$\Pr(\theta_1) \sim Uniform[0, 20]. \tag{12}$$

Now, consider the likelihood. In frequentist statistics, one needs to assume that **E** in Equation (9) is normally distributed. In the Bayesian paradigm, one may do so, but it is not necessary. A Student's t-distribution is often used instead of a Normal distribution in other statistical calculations (e.g., t-tests) due to its additional ("degrees of freedom") parameter, which accommodates the variance arising from small sample sizes more successfully. As in Section 3.1.2, an exponential distribution on the degrees of freedom (ν_p) is specified. It has also been found that specifying a Half-Cauchy distribution on the standard deviation (σ_p) works well [55]. Therefore, the hyperpriors are specified as:

$$\Pr(\nu_p) \sim Exponential[12^{-1}] \tag{13}$$

and:

$$\Pr(\sigma_p) \sim HalfCauchy[1]. \tag{14}$$

The mean of the likelihood is the final hyperparameter that needs to be specified. This is simply Equation (9), written with the priors:

$$\mu_p = \Pr(\theta_0) + \Pr(\theta_1)\mathbf{CDD}. \tag{15}$$

The full likelihood can thus be written as:

$$\Pr(\mathbf{CDD}|\mathbf{E}) \sim StudentT\left(\mu = \mu_p, \nu = \Pr(\nu_p), \sigma = \Pr(\sigma_p)\right). \tag{16}$$

The PyMC3 code is shown below:

```
import pymc3 as pm
with pm.Model() as bayesian_regression_model:
    #  Hyperpriors and priors:
    nu = pm.Exponential('nu', lam=1/len(CDD))
    sigma = pm.HalfCauchy('sigma', beta=1)
    slope = pm.Uniform('slope', lower=0, upper=20)
    intercept = pm.Uniform('intercept', lower=0, upper=10000)
    #  Energy model:
    regression_eq = intercept + slope*CDD
    #  Likelihood:
    y = pm.StudentT('y', mu=regression_eq, nu=nu, sd=sigma, observed=E)
    #  MCMC calculation:
    trace = pm.sample(draws=10000, step=pm.NUTS(), njobs=4)
```

The last line of the code above invokes the MCMC sampler algorithm to solve the model. In this case, the No U-Turn Sampler (NUTS) [56] was used, running four traces of 10,000 samples each, simultaneously on a four-core laptop computer, in 3.5 min fewer samples, could also have been used.

A discussion of the inner workings and tests for adequate convergence of the MCMC is beyond the scope of the study and has been done in detail elsewhere in literature [4]. The key idea for M&V practitioners is that the MCMC, like MC simulation, must converge, and must have done enough iterations after convergence to approximate the posterior distribution numerically. For most simple models such as the ones used in most M&V applications, a few thousand iterations are usually adequate for inference. Two popular checks for posterior validity are the Gelman–Rubin statistic $\hat{R}$ [57,58] and the effective sample size (ESS). The Gelman–Rubin statistic compares the four chains specified in the program above, started at random places, to see if they all converged on the same posterior values. If they did, their ratios should be close to unity. This is easily done in PyMC3 with the `pm.gelman_rubin(trace)` command, which indicates $\hat{R}$ equal to one to beyond the third decimal place. However, even if the MCMC has converged, it does not mean that the chain is long enough to approximate the posterior distribution adequately because the MCMC mechanism produces a serially correlated (autocorrelated) chain. It is therefore necessary to calculate the *effective* sample size: the sample size taking autocorrelation into account. In PyMC3, one can invoke the `pm.effective_n(trace)` command, which shows that the ESSs for the parameters of interest are well over 1000 each for the current case study. As a first-order approximation, we can therefore be satisfied that the MCMC has yielded satisfactory estimates.

The MCMC results can be inspected in various ways. The posteriors on the parameters of interest are shown in Figure 3. If a normal distribution is specified on the likelihood in Equation (16) rather than the Student's t, the posterior means are identical to the linear regression point estimates—an expected result, since OLS regression is a special case of the more general Bayesian approach. Using a t-distributed likelihood yields slightly different, but practically equivalent, results. The mean or mode of a given posterior is not of as much interest as the full distribution, since this full distribution will be used for any subsequent calculation. However, the mean of the posterior distribution(s) is given in Table 3 for the curious reader.

Two brief notes on Bayesian intervals are necessary. As discussed in Section 1.1, the frequentist 'confidence' interval is a misnomer. To distinguish Bayesian from frequentist intervals, Bayesian intervals are often called 'credible' intervals, although they are much closer to what most people think of when referring to a frequentist confidence interval. The second note is that Bayesians often use HDIs, also known as highest posterior density intervals. These are related to the *area* under the probability density curve, rather than points on the x-axis. In frequentist statistics, we are used to equal-tailed confidence intervals since we compute them by taking the mean, and then adding or subtracting a fixed number—the standard error multiplied by the t-value, for example. This works well for symmetrical distributions such as the Normal, as is assumed in many frequentist methods.

However, real data distributions are often asymmetrical, and forcing an equal-tailed confidence interval onto an asymmetric distribution leads to including an unlikely range of values on the one side, while excluding more likely values on the other. An HDI solves this problem. It does not have equal tails but has equally-likely upper and lower bounds.

Table 3. Summary statistics for Bayesian posterior distributions shown in Figure 3 when a Student's *t*-distribution is used on the likelihood. Compare to linear regression results in Table 2. HDI: Highest Density Intervals.

Parameter	**Value**	**95% HDI**
Slope coefficient	7.69	[6.21, 9.24]
Intercept coefficient	5634	[5351, 5937]

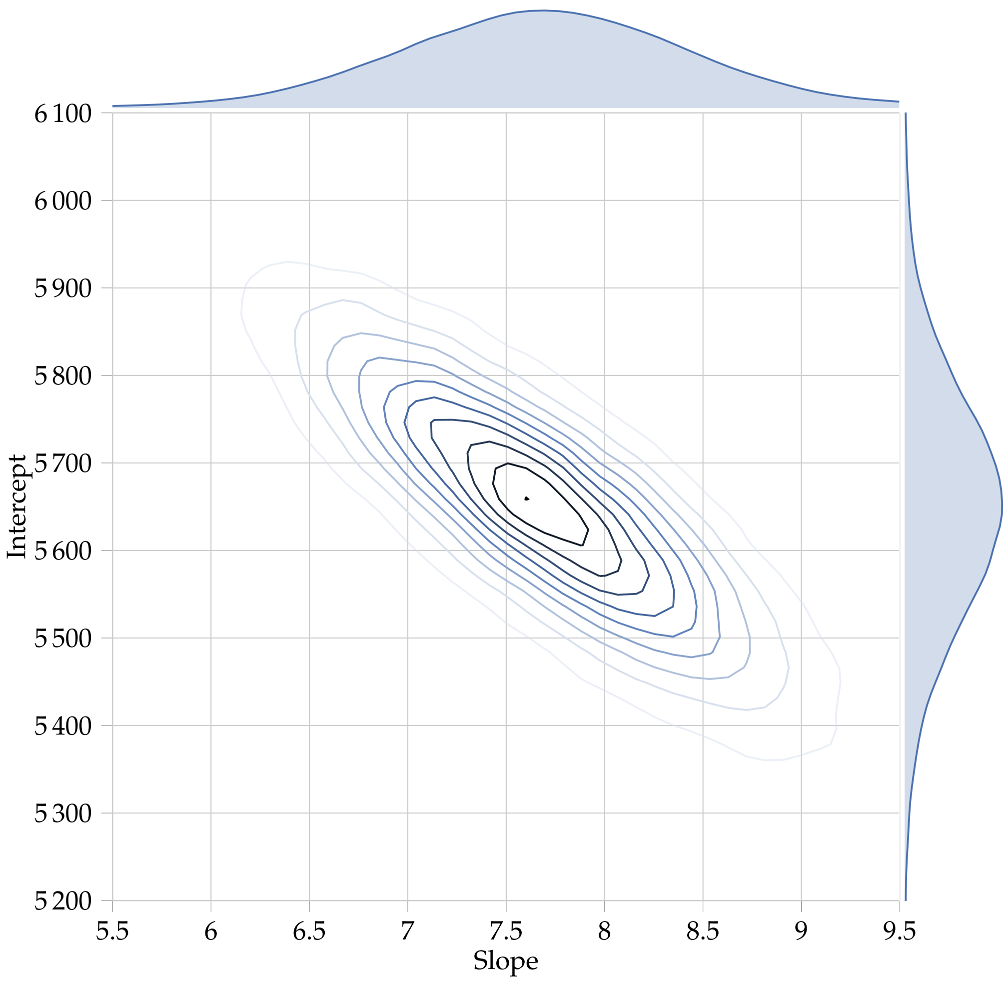


Figure 3. Joint plot of posterior distributions on the parameters of interest. The summary statistics are given in Table 3. Notice how the slope and intercept estimates are correlated: as the slope increases, the intercept decreases. The Markov Chain Monte Carlo (MCMC) algorithm explores this space, resulting in the real joint two-dimensional posterior distribution on the slope and intercept.

The posterior distributions shown in Figure 3 are seldom of use in themselves and are more interesting when combined in a calculation to determine the uncertainties in the baseline as shown in Figure 4, also known as the adjusted baseline. To do so, the posterior predictive distribution needs to be calculated using the `pm.sample_ppc()` command, which resamples from the posterior distributions, much like the MC simulation forward-step of Figure 1.

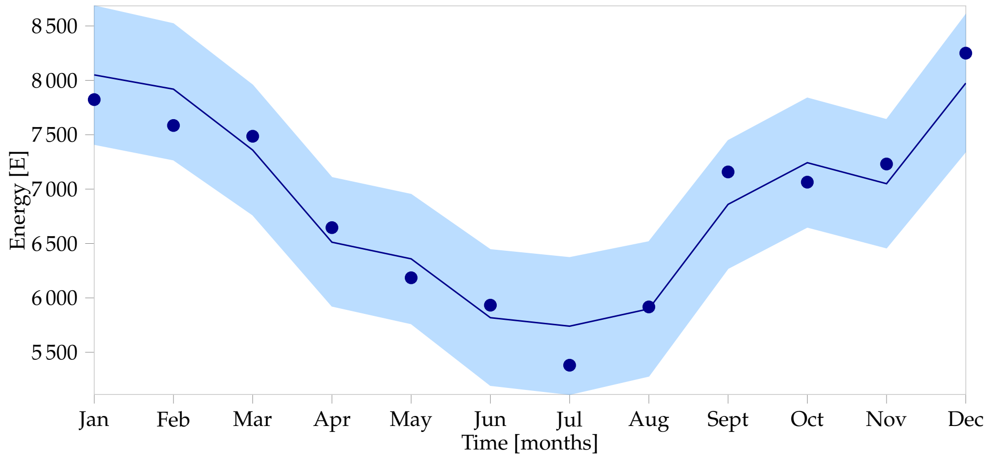


Figure 4. Measured data with overlaid Bayesian baseline model and its 95% HDI.

The Bayesian model can also be used to calculate the *adjusted* baseline, or what the post-implementation period energy use would have been, had no intervention been made. The difference between this value and the actual energy use during the reporting period is the energy saved. For the example under consideration, the IPMVP assumes that an average month in the post-implementation period: one with 162 CDDs. It also assumes that the actual reporting period energy use is 4300 kWh, measured with negligible metering error.

To calculate the savings distribution using the Bayesian method, one would do an MC simulation of:

$$Savings \sim \boldsymbol{\theta}_0 + 162\boldsymbol{\theta}_1 - 4300, \tag{17}$$

where $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ are the distributions shown in Figure 3. Note that they are correlated and so using the PPC method described above would be the correct approach. Running this simulation with 10,000 samples yields the distribution shown in Figure 5. The 95% HDI is [2229, 2959], while the frequentist interval is [1810, 3430] for the same data—a much wider interval. Furthermore, the IPMVP then assumes averages and multiplies these figures to get annual savings and uncertainties. In the Bayesian paradigm, the HDIs can be different for every month (or time step) as shown in Figure 4, yielding more accurate overall savings uncertainty values.

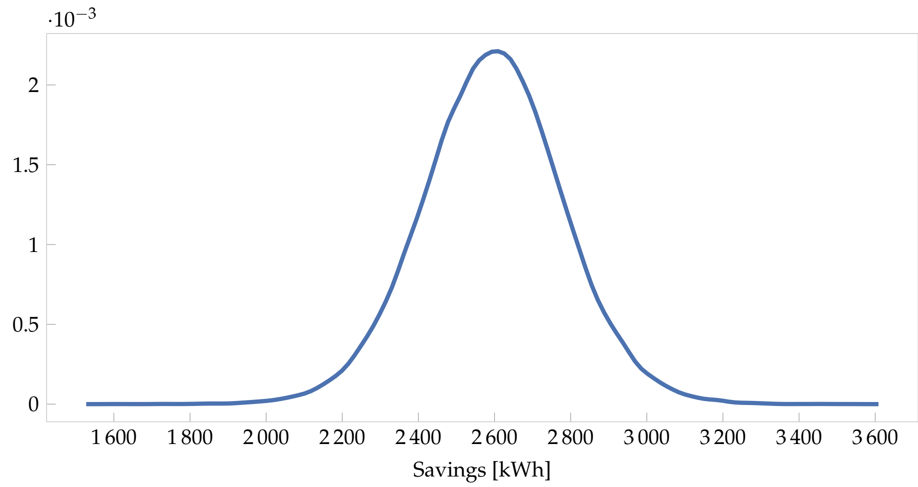


Figure 5. Distribution on the savings for a month with 162 Cooling Degree Days (CDDs).

3.2.4. Robustness to Outliers

As alluded to above, using the Student's *t*-distribution rather than the normal distribution allows for Bayesian regression to be robust to outliers [59]. The heavier tails more easily accommodate an outlying data point by automatically altering the degrees-of-freedom hyperparameter to adapt to the non-normally distributed data. Uncertainty in the estimates is increased, but this reflects the true state of knowledge about the system more realistically than alternative assumptions of light tails, and is therefore warranted. The robustness of such regression does not give the M&V practitioner carte blanche to ignore outliers. One should always seek to understand the reason for an outlier; if the operating conditions of the facility were significantly different, the analyst should consider neglecting (or 'condoning') the data point. However, it is not always possible to trace the reasons for all outliers, and inherently robust models are useful (Technical note: the treatment here is very basic, and for illustration. More advanced Bayesian approaches are also available. For example, if there are only a few outliers, a mixture model may be used [60]. If there is a systematic problem such an unknown error variable, one can "marginalise" the offending variable out. The right-hand and top distributions of Figure 3 are marginal distributions: e.g., the distribution on the slope, with the intercept marginalised out, and vice versa. For an M&V example of marginalisation where an unknown measurement error is marginalised out, see Carstens [61] (Section 3.5.3). von der Linden et al. provides a thorough treatment of all the options for dealing with outliers [3] (Ch. 22)).

To demonstrate the robustness of such a Bayesian model, consider the regression case above. Suppose that for some reason the December cooling load was 3250 kWh and not 8250 kWh, indicated by the red point in the lower right-hand corner of Figure 6. If OLS regression were used, and this point is not removed, it would skew the whole model. However, the *t*-distributed likelihood in the Bayesian model is robust to the outlier. The effect is demonstrated in Figure 6. Four lines are plotted: the solid lines are for the data set without the outlier. The dashed lines are for the data set with the outlier. In the Bayesian model, the two regression lines are almost identical and close to the OLS regression line for the standard set. However, the OLS regression on the outlier set is dramatically biased and would underestimate the energy use for hot months due to the outlier.

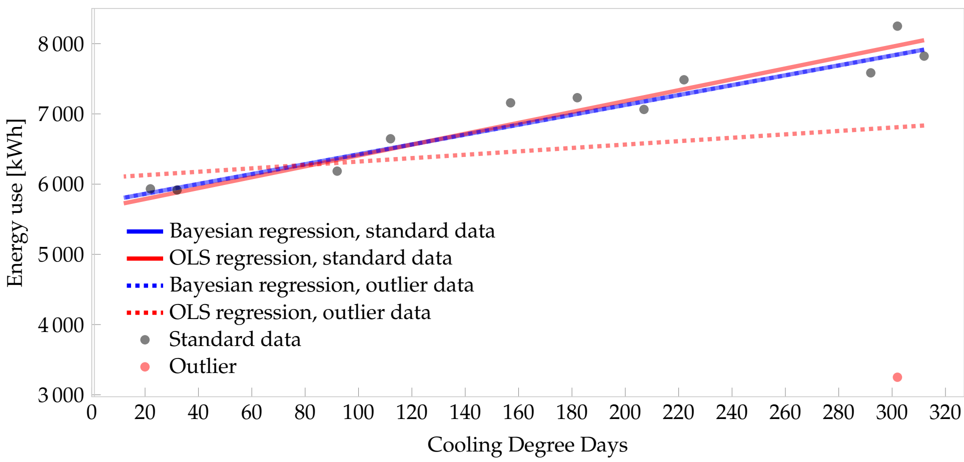


Figure 6. Demonstration of robustness of *t*-distributed Bayesian regression. Note that the two Bayesian regression lines (solid and dashed) coincide almost perfectly.

3.2.5. Hierarchical Models

A further advantage in the Bayesian paradigm is the use of hierarchical, or multilevel models. This is a feature of the model structure rather than the Bayesian calculation itself (it also works for MLE) [2], but it is nevertheless useful in M&V. Suppose that multiple measures are installed at multiple sites so that the IPMVP Option C: Whole Building Retrofit is used for M&V. The UMP Chapter 8 [62] reports that there are two ways to analyse such data. The two-stage approach involves first analysing each facility separately and then using these results for the overall analysis in stage two. The fixed effects approach analyses all buildings simultaneously but assumes that the effect sizes are constant across facilities, using an average effect for all buildings. Hierarchical modelling considers both the individual facility's energy saving and the overall effect simultaneously. It does this by assuming that the group effects are different realisations of an overarching distribution with a mean and variance, which is used as a prior. This can lead to 'shrinkage' in the parameter uncertainty estimates because the group effects are mutually informative. For groups with little data, the overarching effect distribution plays a larger role, and for groups with more data, a smaller role. In addition, the overall variance is reduced because the sources of inter-facility variance are isolated from that of inter-measure variance. The result for a hierarchical model is that the effect estimation for an individual facility is influenced by the overall estimate of the measured effect, as well as by the data for the facility. As another example, consider a program that retrofits air conditioning units in different provinces in South Africa. One could fix the savings effect across all facilities, but this will underestimate some and overestimate others. Otherwise, one could analyse by facility, then by province, and then overall. The hierarchical model provides a better alternative in these cases, and comprises the bulk of many Bayesian data analysis texts [2,4]. Booth, Choudhary, and Spiegelhalter have provided an excellent example of using hierarchical Bayesian models in energy M&V [63].

4. Bayesian Alternatives for Standard M&V Analyses

At this point, an M&V analyst may want to try the Bayesian method for an M&V problem, but where to start? In Table 4, some Bayesian alternatives to standard M&V analyses are given. The references cited are mostly from M&V studies, although some general statistical sources are also listed where applicable.

Table 4. Common M&V (Measurement and Verification) cases and their Bayesian alternatives.

Problem Type	Variant	Bayesian Alternative	Example Reference
Sampling	Single Sample		Section 3.1, [2]
	Randomised Control Trial	Bayesian Estimation	[42]
	ANOVA	Hierarchical modelling	[64]
Regression	Standard	Bayesian regression	Section 3.2, [19]
	With change points	Bayesian regression	[17]
	Pooled fixed effects	Hierarchical modelling	[63]
	Non-parametric	Gaussian Process	[39,65,66]
Longitudinal	Persistence	Dynamic Generalised	[20]
Meter calibration		Linear Model Simulation Extrapolation with Bayesian refinement	[67]

5. Conclusions

The Bayesian paradigm provides a coherent and intuitive approach to energy measurement and verification. It does so by defining the basic M&V question—the savings inference given measurements—using conditional probabilities. It also provides a simpler and more intuitive understanding of probability and uncertainty because it allows the analyst to answer real questions in a straightforward manner, unlike traditional statistics. Due to recent technological and mathematical advances being incorporated into software, analysts need not be expert statisticians to harness the power and flexibility of this method.

The probabilistic nature of Bayesian analysis allows for automatic and accurate uncertainty quantification in savings models. The richer nature of the Bayesian result is shown in a sampling and a regression problem, where it is found that the Bayesian method allows for more realistic modelling and a greater variety of questions that can be answered. Its flexibility is also demonstrated by constructing a robust regression model, which is much less sensitive to outliers that standard ordinary least squares regression traditionally used in M&V.

Acknowledgments: The authors would like to thank Mark Rawlins, Pieter de Villiers, and Franco Barnard for their inputs to the draft of this study, as well as the anonymous reviewers whose detailed feedback improved the paper substantially.

Author Contributions: Herman Carstens conceived of and wrote this paper. Xiaohua Xia and Sarma Yadavalli led the study and his doctoral work in which some of these ideas were developed.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ADVI	Automatic Differentiation Variational Inference
ANN	Artificial Neural Network
ASHRAE	American Society of Heating, Refrigeration, and Air Conditioning Engineers
CDD	Cooling Degree Days
CI	Confidence Interval
ESS	Effective Sample Size
GP	Gaussian Process
HDI	Highest Density Interval
IPMVP	International Performance Measurement and Verification Protocol
ISO	International Standards Organization
MC	Monte Carlo
MCMC	Markov Chain Monte Carlo
M&V	Measurement and Verification
PPC	Posterior Predictive Check
OLS	Ordinary Least Squares
UMP	Uniform Methods Project

References

1. Sivia, D.; Skilling, J. *Data Analysis: A Bayesian Tutorial*; Oxford University Press: Oxford, UK, 2006.
2. Kruschke, J. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*, 2th ed.; Academic Press: Cambridge, MA, USA, 2015.
3. Von der Linden, W.; Dose, V.; Von Toussaint, U. *Bayesian Probability Theory: Applications in the Physical Sciences*; Cambridge University Press: Cambridge, UK, 2014.
4. Gelman, A.; Carlin, J.B.; Stern, H.S.; Rubin, D.B. *Bayesian Data Analysis*; Taylor & Francis: Abingdon-on-Thames, UK, 2014; Volume 2.
5. Efficiency Valuation Organization. *International Performance Measurement and Verification Protocol*; Efficiency Valuation Organization: Toronto, ON, Canada, 2012; Volume 1.
6. International Organization for Standardization (ISO). *Guide 98–3 (2008) Uncertainty of Measurement Part 3: Guide to the Expression of Uncertainty in Measurement (GUM: 1995)*; ISO: Geneva, Switzerland, 2008.
7. American Society of Heating, Refrigeration and Air-Conditioning Engineers, Inc. Guideline 14-2014, Measurement of Energy, Demand, and Water Savings. 2014. Available online: https://ihsmarkit.com/products/ashrae-standards.html (accessed on 1 December 2017).
8. National Renewable Energy Laboratory. Uniform Methods Project. Available online: https://energy.gov/eere/about-us/ump-home (accessed on 1 December 2017).
9. Robert, C. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2007.
10. Neyman, J. Outline of a theory of statistical estimation based on the classical theory of probability. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Sci.* **1937**, *236*, 333–380.

11. Montgomery, D.C.; Runger, G.C. *Applied Statistics and Probability for Engineers*; John Wiley & Sons: Hoboken, NJ, USA, 2010.
12. Wasserstein, R.L.; Lazar, N.A. The ASA's statement on *p*-values: Context, process, and purpose. *Am. Stat.* **2016**, *70*, 129–133.
13. Joint Committee for Guides in Metrology. *OIML G1-101 Evaluation of Measurement Data—Supplement 1 to the "Guide to the Expression of Uncertainty in Measurement"—Propagation of Distributions Using a Monte Carlo Method*; International Organization of Legal Metrology: Paris, France, 2008.
14. Savage, S.L.; Markowitz, H.M. *The flaw of Averages: Why We Underestimate Risk in the Face of Uncertainty*; John Wiley & Sons: Hoboken, NJ, USA, 2009.
15. Kuang, Y.C.; Rajan, A.; Ooi, M.P.L.; Ong, T.C. Standard uncertainty evaluation of multivariate polynomial. *Measurement* **2014**, *58*, 483–494.
16. Rajan, A.; Ooi, M.P.L.; Kuang, Y.C.; Demidenko, S.N. Analytical Standard Uncertainty Evaluation Using Mellin Transform. *IEEE Access* **2015**, *3*, 209–222.
17. Shonder, J.A.; Im, P. Bayesian Analysis of Savings from Retrofit Projects. *ASHRAE Trans.* **2012**, *118*, 367.
18. Lira, I. The GUM revision: The Bayesian view toward the expression of measurement uncertainty. *Eur. J. Phys.* **2016**, *37*, 025803.
19. Tehrani, N.H.; Khan, U.T.; Crawford, C. Baseline load forecasting using a Bayesian approach. In Proceedings of the 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), Vancouver, BC, Canada, 15–18 May 2016; pp. 1–4.
20. Carstens, H.; Xia, X.; Yadavalli, S. Efficient Longitudinal Population Survival Survey Sampling for the Measurement and Verification of Building Retrofit Projects. *Energy Build.* **2017**, *150*, 163–176.
21. Carstens, H.; Xia, X.; Yadavalli, S. Efficient metering and surveying sampling designs in longitudinal Measurement and Verification for lighting retrofit. *Energy Build.* **2017**, *154*, 430–447.
22. Heo, Y. Bayesian Calibration of Building Energy Models for Energy Retrofit Decision-Making under Uncertainty. Ph.D. Thesis, Georgia Institute of Technology, Atlanta, GA, USA, 2011.
23. Lee, B.D.; Sun, Y.; Augenbroe, G.; Paredis, C.J. Toward Better Prediction of Building Performance: A Workbench to Analyze Uncertainty in Building Simulation. In Proceedings of the 13th Conference of the International Building Performance Simulation Association (BS2013), Le Bourget Du Lac, France, 25–30 August 2013.
24. Booth, A.; Choudhary, R. Decision making under uncertainty in the retrofit analysis of the UK housing stock: Implications for the Green Deal. *Energy Build.* **2013**, *64*, 292–308.
25. Heo, Y.; Graziano, D.J.; Guzowski, L.; Muehleisen, R.T. Evaluation of calibration efficacy under different levels of uncertainty. *J. Build. Perform. Simul.* **2015**, *8*, 135–144.
26. McShane, B.B.; Gal, D.; Gelman, A.; Robert, C.; Tackett, J.L. Abandon statistical significance. *arXiv* **2017**, arXiv:1709.07588.
27. Leek, J.; Colquhoun, D.; McShane, B.B.; Gelman, A.; Nuijten, M.B.; Goodman, S.N. Comment: Five Ways to Fix Statistics. *Nature* **2017**, *551*, 557–559.
28. Ioannidis, J.P. Why most published research findings are false. *PLoS Med.* **2005**, *2*, e124.
29. Button, K.S.; Ioannidis, J.P.; Mokrysz, C.; Nosek, B.A.; Flint, J.; Robinson, E.S.; Munafò, M.R. Power failure: Why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* **2013**, *14*, 365–376.
30. Pavlak, G.S.; Florita, A.R.; Henze, G.P.; Rajagopalan, B. Comparison of Traditional and Bayesian Calibration Techniques for Gray-Box Modeling. *J. Archit. Eng.* **2013**, *20*, 04013011.
31. Yildiz, B.; Bilbao, J.; Dore, J.; Sproul, A. Recent advances in the analysis of residential electricity consumption and applications of smart meter data. *Appl. Energy* **2017**, *208*, 402–427.
32. Gallagher, C.V.; Bruton, K.; Leahy, K.; O'Sullivan, D.T. The Suitability of Machine Learning to Minimise Uncertainty in the Measurement and Verification of Energy Savings. *Energy Build.* **2018**, *158*, 647–655.
33. Mathieu, J.L.; Price, P.N.; Sila, K.; Piette, M.A. Quantifying changes in building electricity use, with application to demand response. *IEEE Trans. Smart Grid* **2009**, *41*, 374–381.
34. Granderson, J.; Touzani, S.; Custodio, C.; Sohn, M.D.; Jump, D.; Fernandes, S. Accuracy of automated measurement and verification (M&V) techniques for energy savings in commercial buildings. *Appl. Energy* **2016**, *173*, 296–308.
35. Chipman, H.A.; George, E.I.; McCulloch, R.E. BART: Bayesian additive regression trees. *Ann. Appl. Stat.* **2010**, *4*, 266–298.

36. Tran, D.; Kucukelbir, A.; Dieng, A.B.; Rudolph, M.; Liang, D.; Blei, D.M. Edward: A library for probabilistic modeling, inference, and criticism. *arXiv* **2016**, arXiv:1610.09787.
37. Neal, R.M. *Bayesian Learning for Neural Networks*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 118.
38. Wolpert, D.H. The Lack of A Priori Distinctions Between Learning Algorithms. *Neural Comput.* **1996**, *8*, 1341–1390, doi:10.1162/neco.1996.8.7.1341.
39. Heo, Y.; Zavala, V.M. Gaussian process modeling for measurement and verification of building energy savings. *Energy Build.* **2012**, *53*, 7–18.
40. Zhang, Y.; O'Neill, Z.; Dong, B.; Augenbroe, G. Comparisons of inverse modeling approaches for predicting building energy performance. *Build. Environ.* **2015**, *86*, 177–190.
41. Rasmussen, C.E.; Williams, C.K. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006; Volume 1.
42. Kruschke, J.K. Bayesian Estimation Supersedes the *t*-Test. *J. Exp. Psychol. Gen.* **2013**, *142*, 573–603.
43. Salvatier, J.; Wiecki, T.V.; Fonnesbeck, C. Probabilistic programming in Python using PyMC3. *PeerJ Comput. Sci.* **2016**, *2*, e55.
44. Winkler, R.L. The assessment of prior distributions in Bayesian analysis. *J. Am. Stat. Assoc.* **1967**, *62*, 776–800.
45. Lindley, D.V. The choice of sample size. *J. R. Stat. Soc. Ser. D (Stat.)* **1997**, *46*, 129–138.
46. Bernardo, J. Statistical inference as a decision problem: The choice of sample size. *J. R. Stat. Soc. Ser. D (Stat.)* **1997**, *46*, 151–153.
47. Goldberg, M.L. Reasonable Doubts: Monitoring and Verification for Performance Contracting. In Proceedings of the 1996 ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA, USA, 25–31 August 1996; Volume 4, pp. 133–143.
48. Ruch, D.; Kissock, J.; Reddy, T. Prediction uncertainty of linear building energy use models with autocorrelated residuals. *J. Sol. Energy Eng.* **1999**, *121*, 63–68.
49. Kissock, J.K.; Haberl, J.S.; Claridge, D.E. Inverse modeling toolkit: Numerical algorithms (RP-1050). *Trans.-Am. Soc. Heat. Refrig. Air Cond. Eng.* **2003**, *109*, 425–434.
50. Reddy, T.; Claridge, D. Uncertainty of "measured" energy savings from statistical baseline models. *HVAC&R Res.* **2000**, *6*, 3–20.
51. Walter, T.; Price, P.N.; Sohn, M.D. Uncertainty estimation improves energy measurement and verification procedures. *Appl. Energy* **2014**, *130*, 230–236.
52. Mathieu, J.L.; Callaway, D.S.; Kiliccote, S. Variability in automated responses of commercial buildings and industrial facilities to dynamic electricity prices. *Energy Build.* **2011**, *43*, 3322–3330.
53. Berk, H.; Ascazubi, M.; Bobker, M. *Inverse Modeling of Portfolio Energy Data for Effective Use with Energy Managers*; Building Simulation: San Francisco, AR, USA, 2017.
54. VanderPlas, J. Frequentism and Bayesianism: a Python-driven primer. *arXiv* **2014**, arXiv:1411.5018.
55. Gelman, A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* **2006**, *1*, 515–534.
56. Hoffman, M.D.; Gelman, A. The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **2014**, *15*, 1593–1623.
57. Gelman, A.; Rubin, D.B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* **1992**, 457–472.
58. Brooks, S.P.; Gelman, A. General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* **1998**, *7*, 434–455.
59. Lange, K.L.; Little, R.J.; Taylor, J.M. Robust statistical modeling using the t distribution. *J. Am. Stat. Assoc.* **1989**, *84*, 881–896.
60. Press, W.H. Understanding data better with Bayesian and global statistical methods. *arXiv* **1996**, arXiv:astro-ph/9604126.
61. Carstens, H. A Bayesian Approach to Energy Monitoring Optimization. Ph.D. Thesis, University of Pretoria, Pretoria, South Africa, 2017.
62. Agnew, K.; Goldberg, M. *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*; Whole-Building Retrofit with Consumption Data Analysis Evaluation Protocol; National Renewable Energy Laboratory: Golden, CO, USA, 2013; Chapter 8.

63. Booth, A.; Choudhary, R.; Spiegelhalter, D. A hierarchical Bayesian framework for calibrating micro-level models with macro-level data. *J. Build. Perform. Simul.* **2013**, *6*, 293–318.
64. Gelman, A. Analysis of variance—Why it is more important than ever. *Ann. Stat.* **2005**, *33*, 1–53.
65. Burkhart, M.C.; Heo, Y.; Zavala, V.M. Measurement and verification of building systems under uncertain data: A Gaussian process modeling approach. *Energy Build.* **2014**, *75*, 189–198.
66. Carstens, H.; Rawlins, M.; Xia, X. A user's guide to the SANAS STC WG guideline for reporting uncertainty in measurement and verification. In Proceedings of the South African Energy Efficiency Confederation Conference, Kempton Park, South Africa, 14–15 November 2017.
67. Carstens, H.; Xia, X.; Yadavalli, S. Low-cost energy meter calibration method for measurement and verification. *Appl. Energy* **2017**, *188*, 563–575.

Applied Energy 218 (2018) 66–77

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Estimate and characterize PV power at demand-side hybrid system

Qian Li[a,b], Zhou Wu[a,b,*], Xiaohua Xia[c]

[a] Key Laboratory of Dependable Service Computing in Cyber Physical Society (Chongqing University) of Ministry of Education, China
[b] College of Automation, Chongqing University, Chongqing 400044, China
[c] Department of Electrical Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa

HIGHLIGHTS

- A new way is developed to directly perform the forecast of PV power at demand side.
- Effects of temperature, humidity, historical value on PV power forecast are explored.
- Estimation results are qualitatively investigated via data mining approaches.
- Experimental studies show that the new method could achieve more accurate prediction.

ARTICLE INFO

Keywords:
Renewable energy
Distributed generation
Solar irradiation
Echo state network
Data mining

ABSTRACT

Power forecasting, in a hybrid photovoltaic (PV) system, is an important issue regarding to the control and optimization of energy systems. In this work, multi-clustered echo state network (MCESN) models are proposed to directly perform the forecast of PV power generation. Furthermore, data characteristics of measured and estimated PV power are qualitatively investigated via data mining approaches. These characteristics include seasonality, stationarity (or non-stationarity) and complexity analysis. Simulation results indicate that the proposed MCESN model is able to precisely forecast PV power one-hour-ahead. The performance on the 24-h-ahead forecast is competitive with the correlation coefficient 99% for sunny days, and 91–98% for cloudy days. Results of data analysis unveil that critical characteristics between the measured and estimated PV power data are analogous. Comparison studies also show that MCESN could achieve more accurate prediction, compared with auto-regressive moving average (ARMA), back propagation (BP) neural networks.

1. Introduction

In recent years, due to globally increasing energy demand, renewable energy sources(e.g., wind and solar energy) have gained great attention, as they are freely available, omnipresent, and environmental friendly. Thanks to easy accessibility, government's support, and technical development, large-scale photovoltaic (PV) systems have been installed around the world. However, the power generation of PV system is a nonlinear and complex process, depending on time-varying factors, such as, temperature, humidity, wind speed and direction, and historical data of PV system. In order to ensure reliable and efficient operation of PV energy systems, it is essential and urgent to forecast PV power precisely [1,2].

There have been a large number of studies on PV power prediction, in which high accuracy and low computational complexity are two main concerns.

A common approach is to transform PV power prediction into solar irradiance prediction, which consists of two steps. The first step is to forecast solar irradiance, and the second step is to calculate the PV power according to solar irradiation and system parameters. Different models of prediction have been developed by traditional techniques and linear methods, e.g., various clear-day models [3], auto-regressive moving average (ARMA) [4] and other econometric technologies. However, as many statistical assumptions and empirical parameters are involved in these models, it is rather difficult to precisely forecast the dynamic behavior of solar irradiance. Some improved models have been proposed based on advanced technologies in [5–7].

Artificial intelligence (AI) and neural network (NN) provide powerful tools of approximating nonlinear systems. Various AI and NN models have been successfully applied to forecasting solar irradiance in literature. A wavelet-coupled support vector machine (W-SVM) model was adopted to forecast global incident solar radiation [8]. A NN model is proposed to achieve a 24-h-ahead solar irradiance prediction for a PV system [9]. Based on recurrent neural networks (RNNs) and wavelet

* Corresponding author at: Key Laboratory of Dependable Service Computing in Cyber Physical Society (Chongqing University) of Ministry of Education, China.
E-mail address: wuzhsky@gmail.com (Z. Wu).

https://doi.org/10.1016/j.apenergy.2018.02.160
Received 4 December 2017; Received in revised form 23 February 2018; Accepted 24 February 2018

neural networks (WNNs), a new diagonal recurrent wavelet neural network (DRWNN) was established to perform the forecast of hourly and daily global solar irradiance [10]. Advanced approximation techniques based on wavelet analysis [11,12], fuzzy technique [13], and empirical analysis [14] can also be employed to enhance NN models. In addition, some other forecasting approaches have also been proposed, such as, peer-to-peer (P2P) solar forecasting [15], machine learning [16,17], probabilistic approach, and so forth. The predicted values of solar irradiance are used to obtain PV power output. On the one hand, canonical PV formula could be utilized to compute the power output of PV system. On the other hand, some commercial PV simulation softwares, such as HOMER and PVFORM, could be used to forecast PV power based on the forecasted solar irradiance and system parameters.

Echo state networks (ESNs) is an improved and simplified form of RNNS [18]. Unlike classical RNNs, ESNs adopt non-trainable sparse connections in the hidden layer (called dynamic reservoir), and only connections in the output layer need to be trained through linear regression. As a result, the high computational complexity is conquered, and ESNs is much faster than traditional RNNs. ESNs also show obvious advantages in dealing with nonlinear time series and dynamic prediction system due to its high prediction accuracy and efficiency. ESNs have been widely applied to various practical fields, including dynamic pattern classification and recognition [19,20], image processing [21], optimal energy management [22], and especially nonlinear time series prediction [23,24]. To our best knowledge, there exist few results in ESN-based prediction of solar irradiance and PV power.

For a PV hybrid system, one practical issue is the uncertainty of PV power. While considering the external environment and different demand-side features, the PV power cannot be directly calculated from a linear form of solar irradiance. Therefore, recent studies have focused on the direct prediction of PV power [25–28]. In this paper, the uncertain PV power at the demand side will be specifically modeled in a direct approach. In the application of PV hybrid system, few results are reported to evaluate inner rules and hidden patterns of the demand-side PV power. Influenced by many factors, such as seasons, geographic locations, weather and surroundings, the PV power profile presents its own data characteristics, which are closely related to the power generation process [29]. In order to unveil the inner dynamics, data features of measured and estimated PV power are quantitatively analyzed. In this paper, some main data characteristics between measured and forecasted PV power will be studied to check statistical similarity.

The contributions are in three folds. First, the ESN models are established to directly perform the one-hour-ahead and 24-h-ahead forecast in the PV hybrid system. The direct effects of measured temperature, humidity, historical 24-h-lag information are also explored in detail. Comparison between ARMA model, BP neural networks and MCESN have been conducted. Secondly, the estimation performance is evaluated with comprehensive criteria, such as normalized root mean square error (NRMSE), mean absolute error (MAE), root mean square error (RMSE), and correlation coefficient (r). Thirdly, the data characteristics are investigated with respect to descriptive statistics, seasonality, non-stationarity and complexity.

The rest of this paper is organized as follows. In Section 2, background is introduced. Section 3 describes the basic theory of ESN in terms of network structure, mathematical model, and training methods. The experimental design and numerical results are shown in Section 4. The data characteristics of measured and estimated PV power are qualitatively analyzed in Section 5. Finally, the conclusion is presented in Section 6.

2. Uncertainty in the PV hybrid system

The electricity consumption have been increasing in past decades, which could result in over exploration of traditional fossil fuel resources. Therefore, the exploration of renewable energy (RE) resources is necessary to control fossil fuel consumption and pollutant emission.

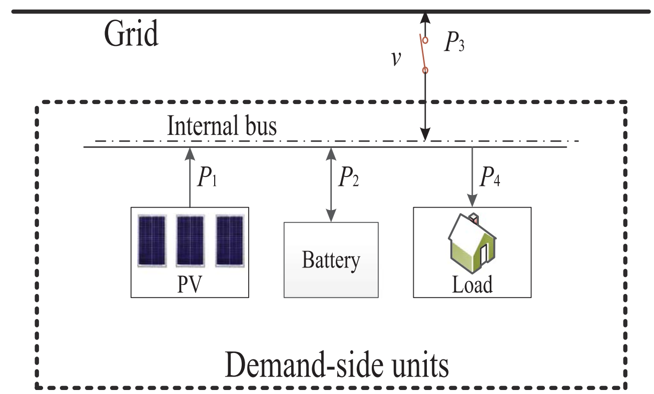


Fig. 1. Schematic of PV hybrid system.

Due to large potential and free availability, wind and solar energy are the popular choices among available RE resources. However, the storage components are required for renewable energy hybrid system due to the intermittent nature. A renewable energy hybrid system is composed of multiple power resources and storage components for stable power supply.

Hybrid renewable energy system (HRES), commonly used for remote power supply, is playing an important role in demand side management with the grid connection, such as, green building and smart community. The PV hybrid system is the most popular application due to easy accessibility, low cost, and high safety. The PV hybrid system consists of PV panel and battery bank that are both connected to the grid, as shown in Fig. 1. As the first priority, the PV power is used to feed the load demand. If the demand is less than the PV power, the surplus PV power will be charged into the battery. If the demand is larger than the PV power, the deficient amount will be then covered by the battery. For saving electricity cost, the battery can be charged by the grid when the electricity has a low price, and be discharged when the electricity has a high price. The grid takes part into the power supply when the load demand cannot be satisfied by the PV and the battery. Note that the PV hybrid system could work in the stand-alone mode and the grid-connected mode, depending on the on/off status of switch v, as shown in Fig. 1.

In the PV hybrid system, a critical problem arisen is the power flow control, which refers to scheduling the power flow between each component for satisfying requirements of cost saving and safety. Let P_1 denote the PV power generation, and P_2 denote the charging/discharging power of battery. Let P_3 denote the grid power flow, and P_4 denote the load demand. With respect to cost, the electricity cost can be expressed as

$$J = \int_{t=0}^{T} \rho(t) P_3(t), \tag{1}$$

where $\rho(t)$ is the real time price of electricity, and J is the electricity cost. With respect to safety, the power balance should be first satisfied as

$$P_1 + P_2 + P_3 = P_4, \tag{2}$$

Power flow control methods have significant effects on electricity cost and operational safety at demand side. In literature, rule-based and optimization-based methods are proposed to reduce the cost and enhance the safety. However, the uncertainty of PV power has presented several challenges on the power flow control. First, the uncertainty could violate the condition of power balance and risk the security of grid and demand-side units. Secondly, the uncertainty could influence actual energy consumption, so that the electricity cost might deviate from the reference one.

In this paper, the prediction of uncertain PV power is specifically studied at the demand side, as solar irradiation at a certain location (a weather station or solar farm) cannot be directly used in the PV power for other distant customers. A PV panel usually consists of several PV cells to convert solar irradiation into direct current power. With a number of PV panels, the hourly PV power output can be simply formulated as:

$$P_{pv}(t) = \eta_{pv}(t) I_{pv}(t) A_c, \tag{3}$$

where $P_{pv}(t)$ is the hourly power output from the PV panels; $\eta_{pv}(t)$ is the efficiency of solar generation; $I_{pv}(t)$ is the hourly solar irradiation incident on the PV panels (kW h/m^2); A_c is the total size of PV panels.

Many researchers have studied the prediction of solar irradiation, and several kinds of methods have been proposed. The PV power can be linearly derived from the solar irradiation, if customers have the same characteristics. Considering different demand-side characteristics, such as, location, weather, external environment, the efficiency $\eta_{pv}(t)$ is time-varying. For example, when partial shading occurs due to cloud and other objects, the efficiency will decrease. Therefore, the uncertain PV has to be modeled specifically at the demand side, while the solar irradiation can only be regarded as a reference. In this study, distributed generation at a university of South Africa is investigated, and the uncertain PV power is directly modeled with an approach of echo state networks. Note that the proposed approach can also be extended to the prediction of solar irradiation.

3. Echo state neural network

As a kind of neural networks, the ESN has a typical architecture that is composed of an input layer, a hidden layer (referred to as a dynamical reservoir), and an output layer, as shown in Fig. 2(a). In the ESN, the input signal, the output signal, and reservoir states are denoted as $\mathbf{u}(t)$, $\mathbf{y}(t)$, $\mathbf{x}(t)$, respectively. For the task of PV power prediction, the input signal could be current and historical values of PV power, temperature, humidity, and other meteorological indicators. The output signal is the future PV power that needs to be predicted, and the reservoir states are states of neurons in the dynamic reservoir, i.e., the hidden layer. First, the ESN has adopted a dynamic reservoir to transfer the input signal into a high-dimensional state vector, which is expected to include all characteristics. Then, an optimal combination of states is chosen for representing output dynamics that is task-related. In other words, the output signal, extracted from the reservoir, is expected to match the desired target signal. In the rest of the paper, vectors are denoted by boldface lowercase letters, e.g., $\mathbf{x}$, while matrices are denoted by boldface uppercase letters, e.g., $\mathbf{X}$.

In the reservoir, there are a large number of neurons with sparse connections. The weight of each connection is randomly initialized, and remains unchanged in the process of training and testing. Inspired by the nature of biological neural system, such as small-world and modular characteristics, a multi-clustered structure of reservoir was designed in the authors' recent study [30]. Compared with the traditional ESN with a random structure, the multi-clustered ESN (MCESN) achieved more accurate prediction. As illustrated in Fig. 2(b), the MCESN has a similar architecture with the traditional ESN, and their difference is the structure of reservoir. In this paper, the MCESN is adopted for the prediction of PV power. The multi-clustered structure is generated according to Kaisers clustering algorithm [31]. All neurons in the reservoir are divided into two different kinds of neurons, i.e., pioneer neurons and normal neurons. The pioneer neurons, with mutual connections, are the critical neurons that determine the number of clusters. The normal neurons have connections within a cluster according to spatial distance between neurons and associated time windows probability model. Note that the spatial distance is defined as the Euclidean distance in the graphic space, and the time window size determines the value of the probability function and affects the connection probability between neurons. The procedure for reservoir generation is given as the following steps [30]:

Step 1: The reservoir is initialized by a small number (denoted as n) of pioneer neurons, which are bi-directionally connected to each other.

Step 2: A random neuron is added and categorized into the nearest cluster, which is determined by the evaluation of the nearest pioneer neuron. This neuron has a probability to connect each node belonging to the same cluster. The probability is calculated based on the spatial distance and the time window size. Any new neuron that fails to establish a connection will be given up. Step (2) is repeated until the number of existing nodes reaches the defined reservoir size (denoted as N).

Step 3: Each node is connected with itself with a self-connecting probability.

Step 4: The reservoir connection matrix $\mathbf{W}^{res}$ is calculated as follows:

$$\mathbf{W}^{res} = \begin{pmatrix} \mathbf{W}_{1,1} & \cdots & \mathbf{W}_{1,i} \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{i,1} & \cdots & \mathbf{W}_{i,i} \end{pmatrix} \tag{4}$$

where $\mathbf{W}_{i,i}$ is the weight matrix of the ith cluster ($i = 1,...,n$); $\mathbf{W}_{i,j}$ are the weight matrix between the ith and jth cluster. Fig. 3 shows the topology of 200 nodes in the two-dimensional graphic plane [0,1]. The clustered phenomenon is obvious, and it is also clear that the intra-cluster connections are more intensive than the inter-cluster connections.

Assume that the MCESN has K, N, and L neurons in the input, hidden, and output layer, respectively. There exist connection weights from the input units to reservoir (denoted as $\mathbf{W}^{in}$,$\mathbf{W}^{in} \in \mathbb{R}_{N\times K}$), reservoir connection weights collected in an $N \times N$ weight matrix $\mathbf{W}^{res} \in \mathbb{R}_{N\times N}$, and connection weights from the reservoir to the readout neurons given in a $L \times N$ output weight matrix $\mathbf{W}^{out} \in \mathbb{R}_{L\times N}$. For $\mathbf{W}^{in}$ and $\mathbf{W}^{res}$, each component is a random number in the MCESN. Furthermore, the connection weights projected back from the readout neurons to the reservoir units are given in an $N \times L$ feedback weight matrix $\mathbf{W}^{back} \in \mathbb{R}_{N\times L}$. The update of the reservoir states is expressed as follows:

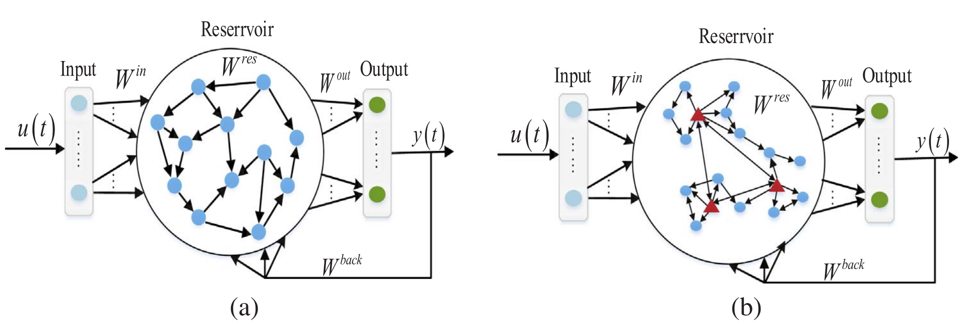


Fig. 2. Network architecture: (a) regular echo state network model with random reservoir structure; (b) multi-clustered echo state network, where the triangles denote the pioneer neurons.

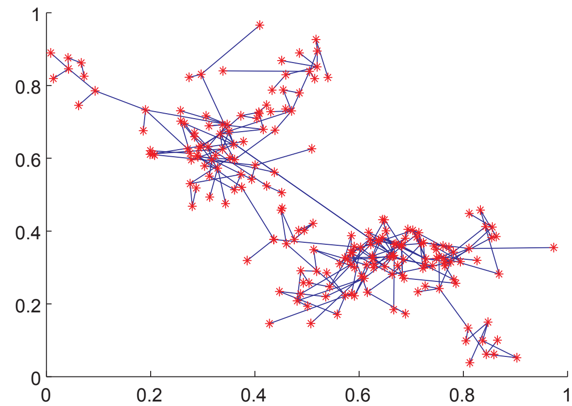


Fig. 3. Two-dimensional projection of multi-clustered network with cluster size $n = 2$.

$$\mathbf{x}(t+1) = f(\mathbf{W}^{in}\mathbf{u}(t+1) + \mathbf{W}^{res}\mathbf{x}(t) + \mathbf{W}^{back}\mathbf{y}(t) + \mathbf{v}(t)), \tag{5}$$

where f is the activation function of each reservoir neuron (usually defined as a sigmoid or Fermi function), and $\mathbf{v}(t)$ is noise signals. The Fermi function is adopted as the hidden neurons function in the paper. The network output is calculated as

$$\mathbf{y}(t+1) = f^{out}(\mathbf{W}^{out}\mathbf{x}(t+1)), \tag{6}$$

where f^{out} is the activation function of the output units. Note that the identity function is adopted in this paper. In the MCESN, the main task is to determine the output weight matrix $\mathbf{W}^{out}$ by training the networks.

At the training stage, the teaching signal, i.e., the future PV power, is given in prior, and the reservoir states can be updated according to Eq. (5). Regression methods could be employed to calculate the output weight matrix. Let l_{tr} represent the length of training datasets, and $\mathbf{X}$ represent the internal state matrix. The corresponding teacher signal vector matrix Λ is denoted as

$$\Lambda = \begin{bmatrix} d_1(1) & d_2(1) & \dots & d_L(1) \\ d_1(2) & d_2(2) & \dots & d_L(2) \\ \vdots & \vdots & \vdots & \vdots \\ d_1(l_{tr}) & d_2(l_{tr}) & \dots & d_L(l_{tr}) \end{bmatrix}_{l_{tr}\times L} \tag{7}$$

and the internal states matrix $\mathbf{X}$ is collected as

$$\mathbf{X} = \begin{bmatrix} x_1(1) & x_2(1) & \cdots & x_N(1) \\ x_1(2) & x_2(2) & \cdots & x_N(2) \\ \vdots & \vdots & \vdots & \\ x_1(l_{tr}) & x_2(l_{tr}) & \cdots & x_N(l_{tr}) \end{bmatrix}_{l_{tr}\times N} \tag{8}$$

where $d(t)$ is the teacher signal, i.e., the future PV power at the training stage.

According to the classical pseudo-inverse method, the output weight matrix $\mathbf{W}^{out}$ is computed as

$$(\mathbf{W}^{out})^T = \mathbf{X}^{+}\Lambda, \tag{9}$$

where $\mathbf{X}^{+}$ is defined as generalized inverse matrix of $\mathbf{X}$.

To overcome the over-fitting phenomenon, a ridge regression training method [32] is applied as

$$(\mathbf{W}^{out})^T = (\mathbf{X}^T\mathbf{X} + \rho\mathbf{I})^{-1}\mathbf{X}^T\Lambda, \tag{10}$$

where $\mathbf{I}$ denotes the identity matrix, ρ is the regularization parameter which should be determined through a large number of experiments for the specific learning tasks.

4. Experimental design and estimation results

In this study, a MCESN model is established to forecast the hourly PV power at the PV hybrid system, installed in University of Pretoria at South Africa. The PV system comprises a large number of equal PV modules with rated power 250 W, providing the cooling, heating, and electrical needs for the campus. The historical data, mainly including temperature, humidity, and PV power, is collected for the year 2014. The meteorological sensors are installed for measuring temperature and humidity, the Danfoss Comlynx Monitor logger [33] is used for recording these PV power, temperature, and humidity. As an example of recorded data, Fig. 4 shows the profiles of hourly PV power $P_{pv}(t)$, temperature (T), and humidity (H) from January 1st 2014 to December 31st 2014.

4.1. ESN setup

To generate the multi-clustered reservoir, the parameter settings are given in Table 1 based on [30]. Weight matrices $\mathbf{W}^{in}$ and $\mathbf{W}^{back}$ are sampled from a uniform distribution over $[-1,1]$, and the spectral radius of $\mathbf{W}^{res}$ is set as 0.8 [34]. The ridge regression training method is adopted to obtain the output weights in the current study. The prediction accuracy is indicated by the normalized root mean square error (NRMSE) [18], which can be expressed as

$$NRMSE = \sqrt{\sum_{t=1}^{l_t}(y(t)-d(t))^2/l_t\sigma^2}, \tag{11}$$

where $y(t)$ is the forecast of PV power; $d(t)$ is the actual PV power; l_t is the number of samples; and σ^2 is the variance of the actual PV power. In this application, 60% of data is used for training, and the remaining data is used for testing.

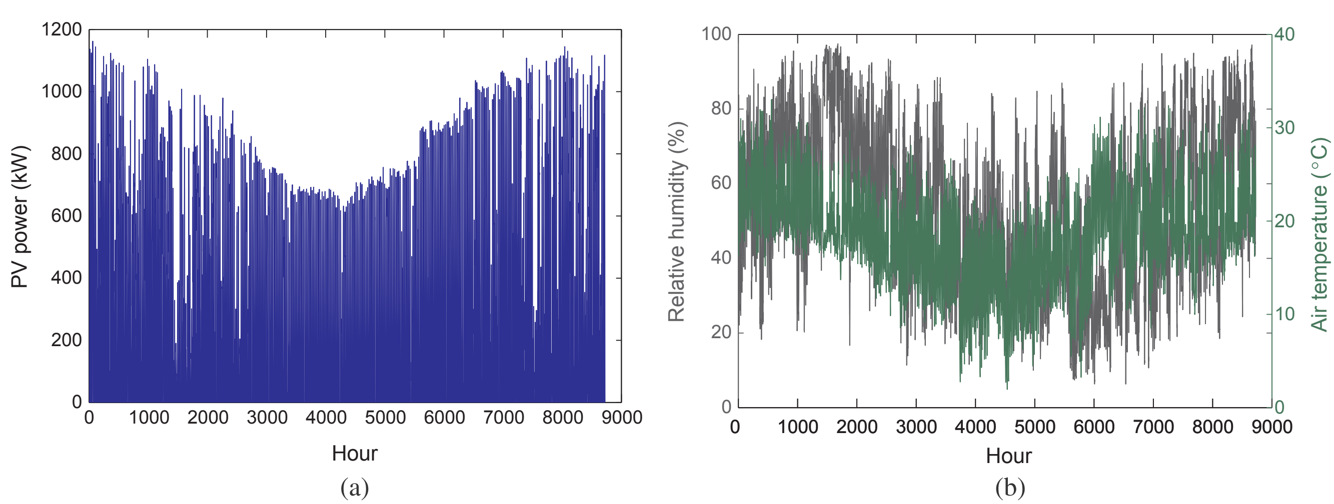


Fig. 4. Recorded data set used in this study: (a) hourly PV power data; (b) corresponding air temperature, humidity.

Table 1
Model parameters for multi-clustered network.

Parameter meaning	Values
Reservoir size	200
Cluster number	2
Time window size	0.3
Self-connecting probability	0.8
Connection probability coefficient1	6
Connection probability coefficient2	10

4.2. One-hour-ahead prediction

In this section, the feasibility and prediction performance of MCESN is evaluated in the one-hour-ahead prediction. As the PV power of each month shows different characteristics, the hourly PV power is modeled for each month in this paper. The PV power at a certain time, denoted as $P_{pv}(t-1)$, is regarded as the input signal, and the PV power at the subsequent hour is regarded as the teacher signal. Take sub-data in summer (January) and winter (July) as two examples, respectively. Results of MCESN are presented in terms of the actual and predicted values at the training and testing stages, as shown in Fig. 5. It can be observed that the prediction output could well match the actual output, and that large fluctuations could be feasibly discovered.

For each month, the prediction accuracy is evaluated in the terms of training and testing NRMSE, respectively. The average NRMSE over 20 independent runs is calculated and shown in Fig. 6. As a result, it can be seen that the prediction accuracy is the lowest in summer.

In addition, in order to directly analyze the factors that may affect PV power, a reasonable input layer of MCESN should be designed. In this study, measured temperature (T) and humidity (H) are used as examples to analyze the direct effect on PV power. The historical values of temperature and humidity are also regarded as the input signals. Besides the model previously derived, 3 other models are evaluated, as illustrated in Table 2. In the first model, the input signal includes the PV power. In the second model, the input signal includes the PV power and temperature. In the third model, the input signal includes the PV power and humidity. In the fourth model, the input signal includes the PV

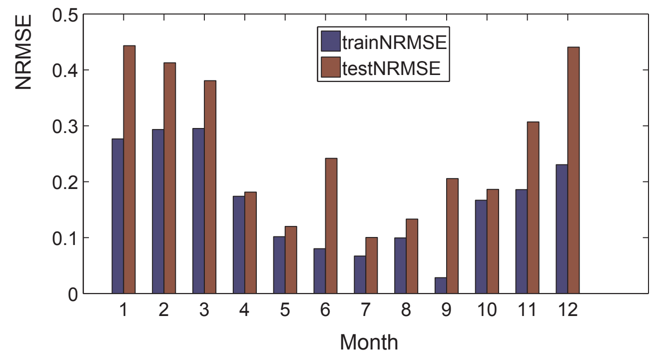


Fig. 6. NRMSEs comparison of training and testing set for each month.

Table 2
Different input and output for MCESN model considering temperature (T) and humidity (H).

Model	MCESN	MCESN + T	MCESN + H	MCESN + T+H
Input	$P_{pv}(t-1)$	$P_{pv}(t-1),T(t-1)$	$P_{pv}(t-1),H(t-1)$	$P_{pv}(t-1),T(t-1),H(t-1)$
Output	$P_{pv}(t)$	$P_{pv}(t)$	$P_{pv}(t)$	$P_{pv}(t)$

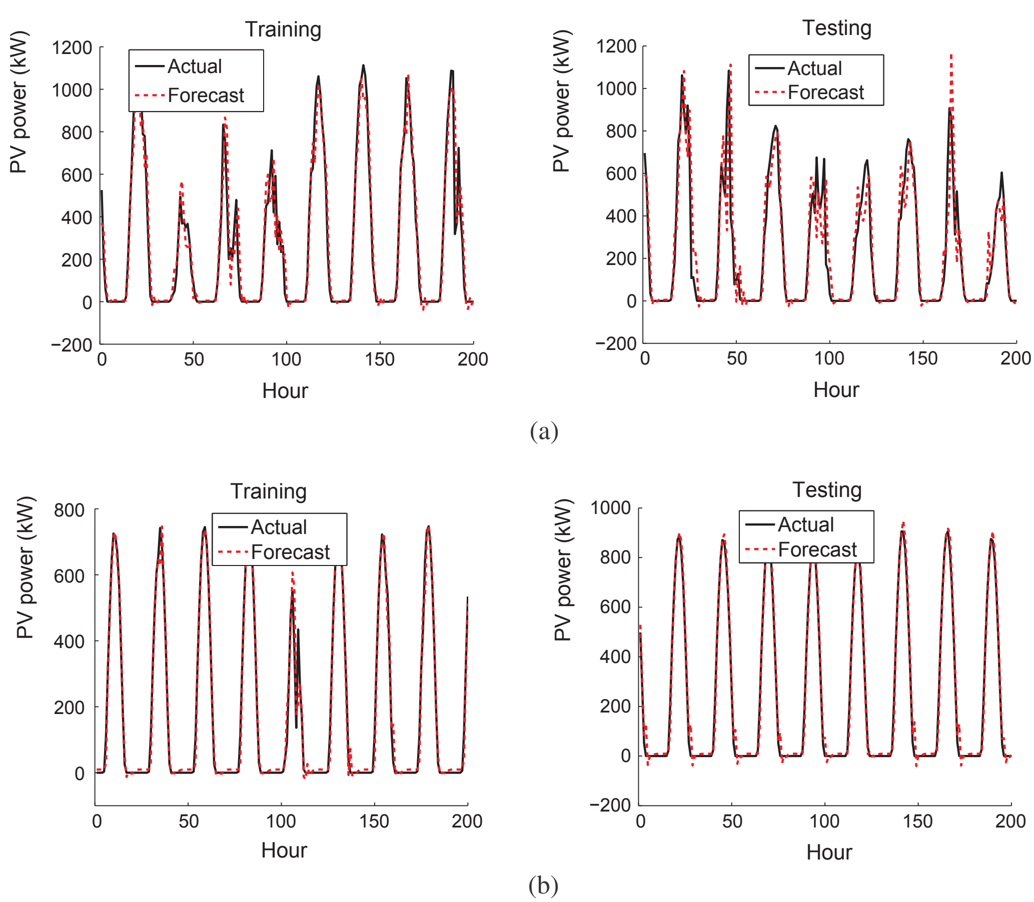


Fig. 5. One-hour-ahead prediction results by MCESN versus actual values for 200 training and testing points: (a) January; (b) July.

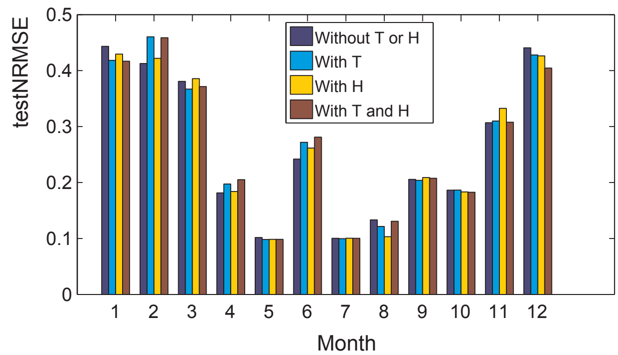


Fig. 7. Testing NRMSEs comparison under four diverse input-output models.

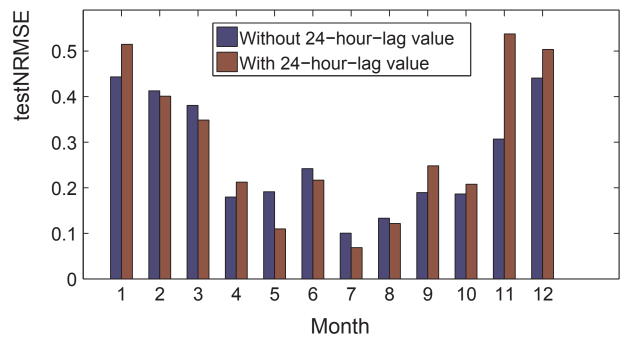


Fig. 9. Testing NRMSEs comparison without/with considering historical 24-h-lag information.

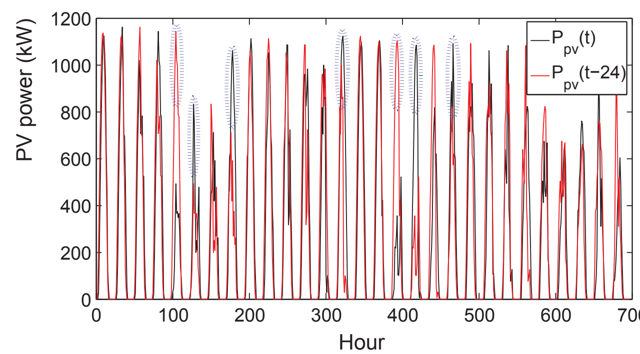


Fig. 10. Comparison of current signal and 24-h-lag information in January.

power, temperature, and humidity. In these different situations, the input and output signals are listed in Table 2. For each model, the prediction accuracy is reported with respect to the NRMSE, as shown in Fig. 7. It can be observed that the input effects of setting measured temperature and humidity are minor, as each model has similar accuracy. In the same way, the direct effect of other measured factors (cloud cover, geographic location) could also be analyzed.

Furthermore, in order to evaluate the periodic phenomenon, the hourly data of PV power is represented as a 24∗365 matrix, in which the component at the mth column and the sth row represents the PV power at the mth hour of the sth day (m = 1,...,24, and s = 1,...,365). The 2-D matrix is plotted as a surface mesh shown in Fig. 8. The daily profile of PV power has a periodic pattern, so the effects of 24-h-lag information on the prediction accuracy are further explored. The multiple inputs are selected as $P_{pv}(t-1)$ and $P_{pv}(t-24)$, and the single output is selected as $P_{pv}(t)$. The testing NRMSE without/with the lag information is presented in Fig. 9, where the blue bar represents the results without the lag information and the red bar represents the results with the lag information. It can be obtained that the 24-h-lag information has positive effects on the prediction accuracy in winter and negative effects in summer. The reason behind this phenomenon is that there exist intensive fluctuations that cause greater prediction error, as shown in Fig. 10.

4.3. 24-h-ahead prediction

The MCESN approaches are utilized to predict the hourly PV power with a good accuracy. However, the hourly PV power is insufficient for certain cases of daily schedule and optimization. Therefore, 24-h-ahead forecast of PV power is further investigated. According to [9], MCESN permits to estimate 24-h-ahead of PV power based on the actual mean value of daily current PV power, daily air temperature, and the day of each month.

Experimental results are presented in Fig. 11 to compare the forecasted profiles and the measured profiles for 4 sunny days (July 19th–22nd). As can be seen, the forecast profiles of PV power can approximate the measured profiles with well accuracy. The scatter plots of prediction results are given in Fig. 12. Most points are close to the diagonal line with the coefficient of determination $R^2 = 0.99$.

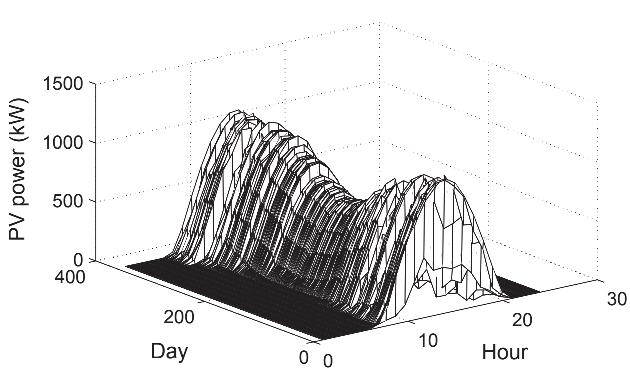


Fig. 8. 2-D surface plot of hourly measured PV power data.

To quantify the prediction performance, several different statistical criteria, i.e., root mean square error (RMSE), correlation coefficient r, and the mean absolute error (MAE), are calculated for different case studies. These statistical results are listed in Table 3. It can be observed that the RMSE for the sunny day is smaller than the cloudy day, and that the correlation coefficient for the sunny day is larger than the cloudy day. The results indicate that the MCESN model delivers less accuracy on the cloudy days. One possible reason is that the weather information, such as, rain and cloud, which is missing in this study, is required for the prediction task of cloudy days.

4.4. Comparisons of MCESN and other typical models

In order to validate the effectiveness of proposed method, two popular models, i.e., auto-regressive moving average (ARMA) and BP neural networks, are selected in the comparison study.

In Fig. 13 and Table 4, the MAE, RMSE, r values between measured and forecasted profiles are compared with respects of ARMA, BP and MCESN. Obviously, MCESN has the highest precision, while ARMA performs the worst. This demonstrates that MCESN has obviously better performance to deal with nonlinear PV power prediction task.

For the PV hybrid system, future PV power is essential information for most problems of design and operation, e.g., sizing and power flow dispatching. For power flow dispatching, day-ahead optimal control is usually applicable to minimize the electricity cost of customers, who already install the hybrid PV system at demand side. Under a certain pricing policy, the PV power prediction affects the optimal dispatching strategy and its associative cost. For example, the PV hybrid system with the proposed MCESN could be used in a time-of-use (TOU) program, which is a typical demand response program to alleviate peak burden. In TOU, the electricity prices are fixed in advance for the

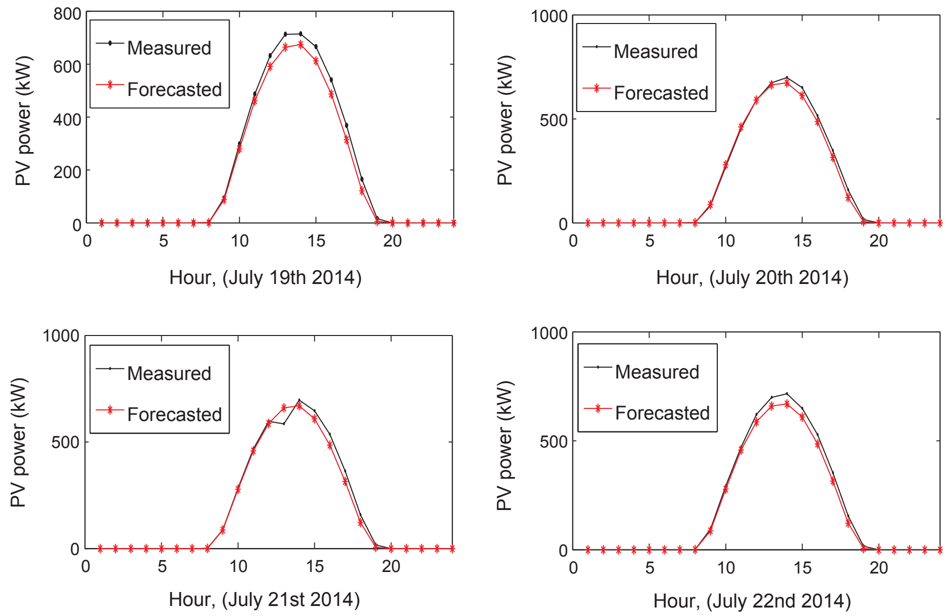


Fig. 11. Comparison between measured and forecasted PV power values 24-h-ahead during period July 19th–22nd (sunny days).

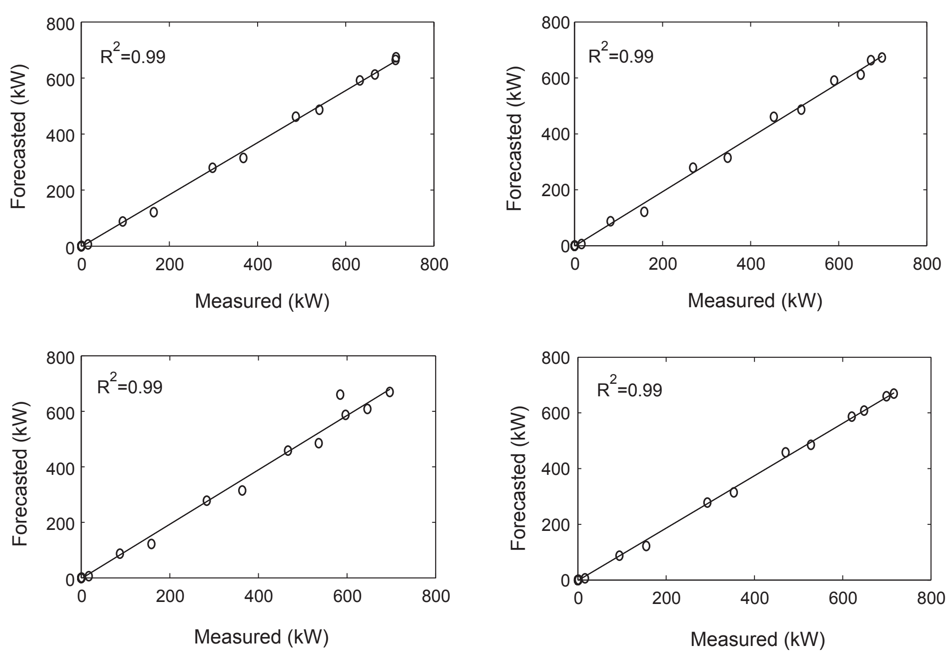


Fig. 12. Scatter plots comparison between measured and forecasted PV power values 24-h-ahead for 4 sunny days (July 19th–22nd).

Table 3
Statistical test between measured and forecasted PV power values 24-h-ahead for 4 sunny days: July 19th–22nd 2014 and 4 cloudy days: November 19th, December 23rd, January 20th, February 20th.

Seasons	Days	MAE (kW)	r	RMSE (kW)
Winter	July 19th	16.33	0.9993	26.72
	July 20th	8.61	0.9988	15.49
	July 21st	12.97	0.9957	24.54
	July 22nd	13.32	0.9996	21.83
Summer	November 19th	65.16	0.9567	113.61
	December 23rd	97.58	0.9225	158.19
	January 20th	74.04	0.9153	120.34
	February 20th	37.14	0.9864	60.12

customer reference. Note that future PV power and load demand could be forecasted using MCESN.

5. Analysis of data characteristics

The prediction performance has been evaluated by quantifying the difference between the predicted results and the measured results. However, internal dynamic of the measured results are not essentially the same with the predicted results. Therefore, some data characteristics, including descriptive statistics, seasonality, stationarity (or non-stationarity), and complexity, are qualitatively investigated in this section. The one-hour-ahead forecast is taken as an example to analyze these characteristics.

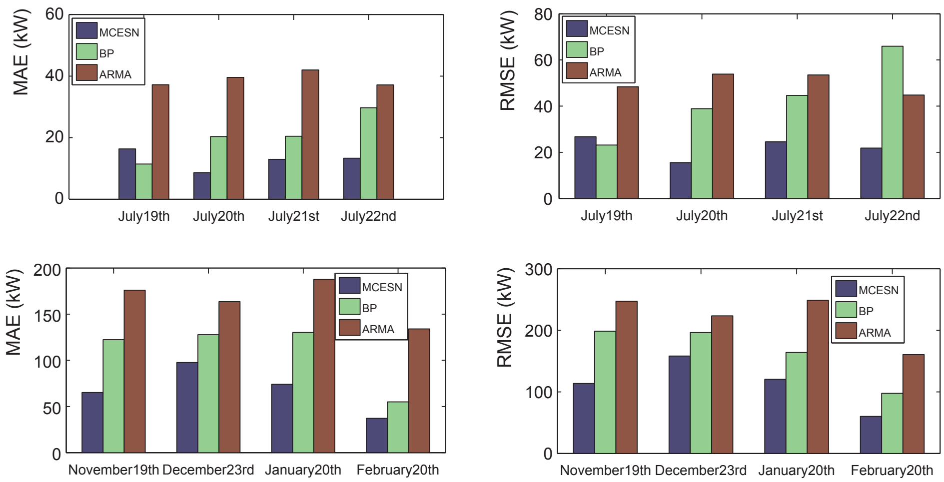


Fig. 13. MAE, RMSE comparison between measured and forecasted PV power values for 4 sunny days: July 19th–22nd 2014 and 4 cloudy days: November 19th, December 23rd, January 20th, February 20th.

Table 4
Correlation coefficient (r) comparison between measured and forecasted PV power values for 4 sunny days: July 19th–22nd 2014 and 4 cloudy days: November 19th, December 23rd, January 20th, February 20th.

		Correlation coefficient r		
Seasons	Days	ARMA	BP	MCESN
Winter	July 19th	0.9866	0.9957	*0.9993*
	July 20th	0.9836	0.9928	*0.9988*
	July 21st	0.9824	0.9835	*0.9957*
	July 22nd	0.9845	0.9690	0.9996
Summer	November 19th	0.8686	0.9227	*0.9567*
	December 23rd	0.8273	0.9143	*0.9225*
	January 20th	0.8213	0.9048	*0.9153*
	February 20th	0.9020	0.9440	*0.9864*

5.1. Descriptive statistics

The histogram between the measured and predictive PV power values are firstly studied. The histogram for January (in the summer) and July (in the winter) is given in Fig. 14(a) and (b), respectively. From Fig. 14, it can be seen that the PV power distribution of the forecasted results is similar with that of the measured results. The distribution of January is also different with that of July, which indicates there exist varying dynamics between seasons. The mean and standard deviation of monthly PV power are computed in Table 5. It can be concluded that the mean, standard deviation of the forecasted results are close to those metrics of the measured results. In addition, statistical test between measured and forecasted values is conducted, e.g., *F*-test and *T*-test, and the results are reported in Table 5. Note that 0 means two data sets are statistically similar, and 1 means they are significantly different. The *F*-test results indicate that there is no significant difference between measured values and forecasted values for most months except March. The *T*-test results also show that there is no significant difference between measured and forecasted PV power values for all months. Therefore, it can be concluded that the forecasted values is similar with the measured valued.

5.2. Periodicity and stationarity

In order to explore the periodic or seasonal characteristics, a surface mesh and a gray image are plotted in Fig. 15. When the region is brighter, the PV power is more intensive, and vice versa. The profiles of PV power show seasonally periodic, although some fluctuations occur in summer (January, November, and December). There is a wider white blob during the summer compared with the winter, as the period from dawn to dusk is longer. Meanwhile, autocorrelation coefficients of measured and forecasted data are plotted in Fig. 16, which can indicate the cyclical pattern has a period of 24 h and non-stationarity. Both the measured and forecasted autocorrelation coefficients values with lag of 24 h are far higher than those with other lags, further demonstrating the 24-h-lag information has strongly positive correlation. Note that non-stationarity means that the statistical properties of PV power dynamics remain diverse during the data generation process. It can be concluded that internal dynamic characteristics, with respect to periodicity and stationarity, keep similar between the measured and forecasted results.

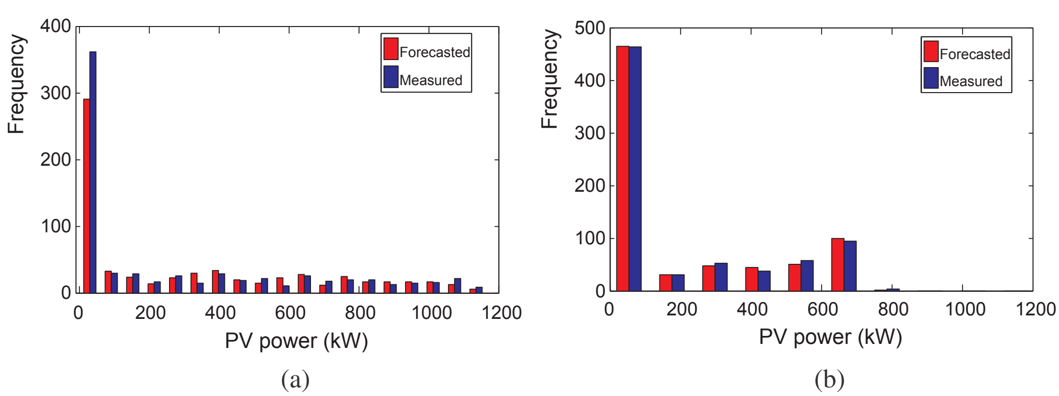


Fig. 14. Histogram comparison between forecasted and measured hourly PV power: (a) January; (b) July.

Table 5
Monthly mean and standard deviation comparison between the forecasted and actual PV power.

The measured values			The forecasted values			Statistical test		
Month	Mean (kW)	Std (kW)	Month	Mean (kW)	Std (kW)	Month	*F*-test	*T*-test
1	269.92	346.92	1	280.39	343.94	1	0	0
2	253.65	328.49	2	253.18	319.09	2	0	0
3	173.96	259.98	3	163.76	231.44	3	1	0
4	201.59	282.26	4	202.75	277.53	4	0	0
5	186.78	260.31	5	188.83	260.06	5	0	0
6	174.72	244.89	6	177.59	245.64	6	0	0
7	183.38	256.32	7	185.28	256.15	7	0	0
8	209.44	287.96	8	211.61	296.99	8	0	0
9	261.14	340.75	9	262.92	343.72	9	0	0
10	287.49	370.75	10	285.03	363.40	10	0	0
11	236.96	332.01	11	235.03	320.09	11	0	0
12	265.97	355.24	12	273.59	348.61	12	0	0

5.3. Complexity

The complexity characteristic could reflect the complex state between regular and irregular relationships. Different techniques have been applied to measure the data complexity, including the phase-space reconstruction method [35], the G-P algorithm [36], and so on. A simple and fast method, i.e., visibility graph method [37], is used to analyze the complexity of forecasted and measured results in this study. The basic idea of the algorithm is to map a time series signal into an associated graph, and graph theory can be employed to characterize the associated graph. The visibility graph method can reflect the structure of the mapped time series according to [37].

For the visible graph method, scatter diagrams and corresponding degree distributions are shown in Figs. 17 and 18. Besides forecasted

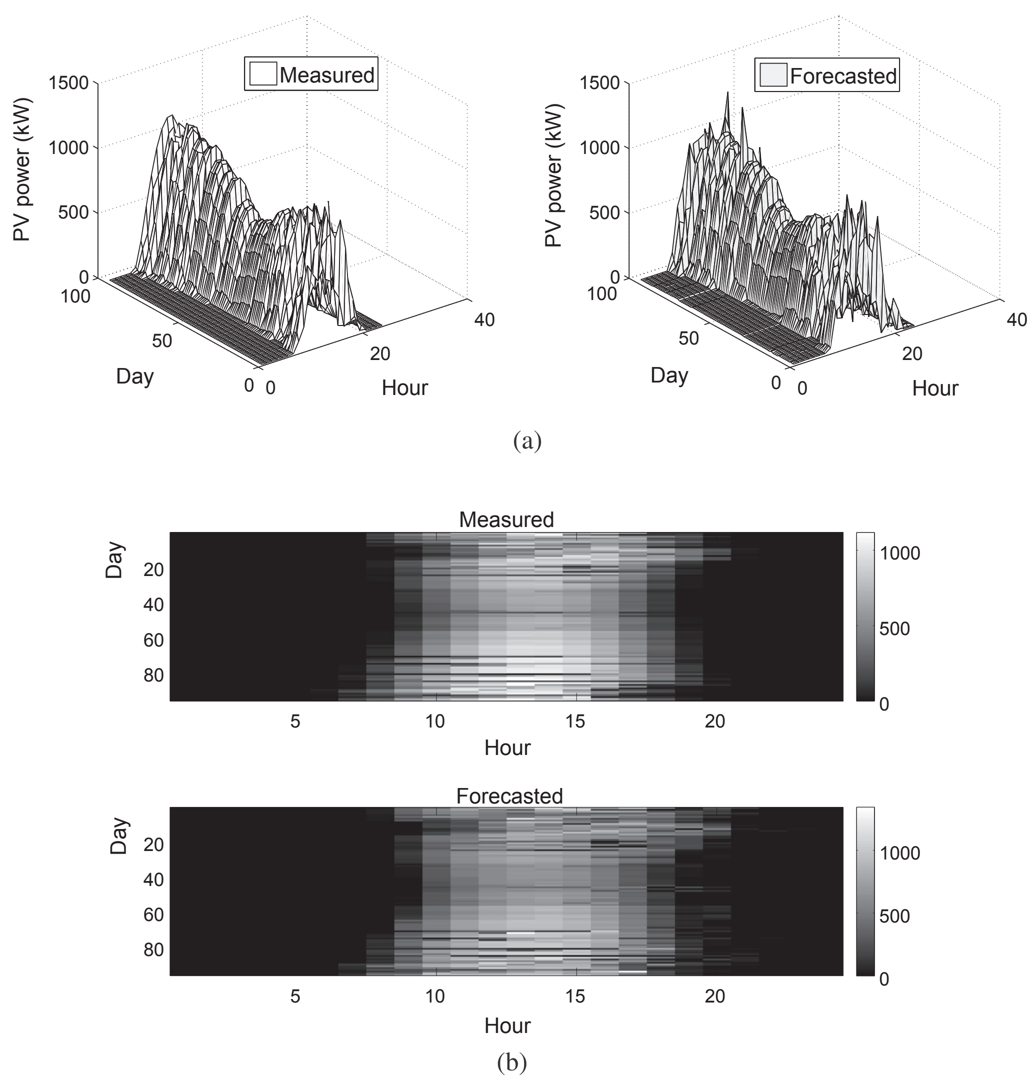


Fig. 15. (a) 2-D surface plot comparison between the estimated and measured PV power. (b) Image visualization comparison between the estimated and measured PV power.

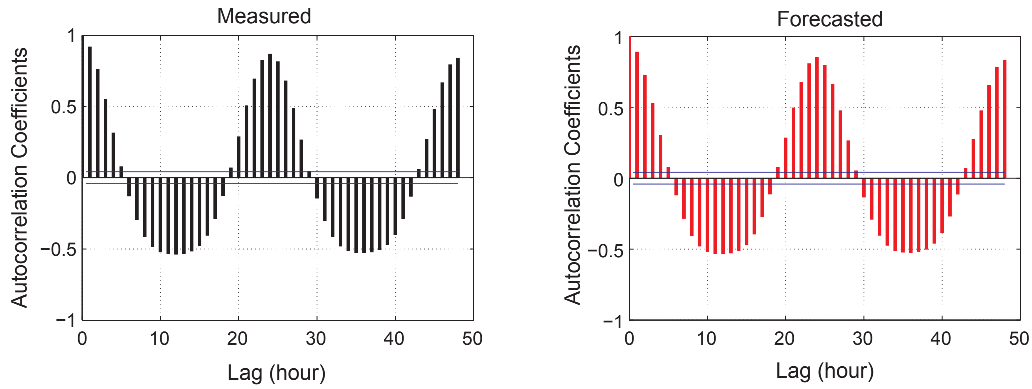


Fig. 16. Autocorrelation coefficients comparison between the measured and forecasted PV power.

and measured singles, several time series are modeled for comparison, such as, a random sequence uniformly distributed in [0,1], and a chaotic sequence generated from the Mackey-Glass system (MGS). From Fig. 17, we can conclude that both forecasted and measured signals present more intensive intra-cluster connections compared with random series. In Fig. 18, the degree distribution of random sequence fits an exponential distribution, while the degree distribution of forecasted and measured signals fits the Gauss-like distribution, which is similar to the distribution of MGS chaotic series. The Gauss-like distribution shows certain chaotic property of PV power.

Note that the scatter diagrams and degree distributions differ from each month. To further evaluate the seasonal complexity, the basic graph metrics, including average path length (AP), clustering coefficient (CC), and average degree (AD), are calculated for each month. The formulas of AP and CC are given in the following equations:

$$AP = \frac{1}{O(O-1)} \sum_{\varsigma \neq \zeta} \vartheta_{\varsigma,\zeta}, \quad (12)$$

$$CC = \frac{1}{O} \sum_{\varepsilon=1}^{O} \frac{2e_\varepsilon}{\xi_\varepsilon(\xi_\varepsilon-1)}, \quad (13)$$

where $\vartheta_{\varsigma,\zeta}$ denotes the shortest length between point ς and ζ of time series, O is the length of sequence; ξ_ε, e_ε represent the degree of point ε and the actual number of edges among the points connected to point ε.

In Table 6, AP, CC and AD are calculated for different time series. It can also be seen that the AP and AD of measured and forecasted signals are between those of MGS and random signals, which could indicate certain small-world properties between random and chaos. In Table 7, AP, CC and AD are calculated for each month. The measured results are comparable with respect to these three metrics. The small values of AP, CC, and AD in summer mean the high randomness, which can explain the poor performance in summer.

In this section, several main data features, including descriptive statistics, seasonality, non-stationarity and complexity of measured and forecasted results are qualitatively analyzed. Experimental results show that the measured and forecasted signals have similar dynamics and complexity. Some linear models, such as, moving average (MA), autoregressive (AR), and auto-regressive moving average (ARMA) may not be suitable for modeling the demand-side PV power precisely. Therefore, the MCESN model is proposed to predict the demand-side PV power due to its nonlinear mapping capacity.

6. Conclusions

For the PV power forecast in the demand-side hybrid system, this paper presents a direct approach for one-hour-ahead prediction and 24-h-ahead prediction based on multi-clustered echo state network

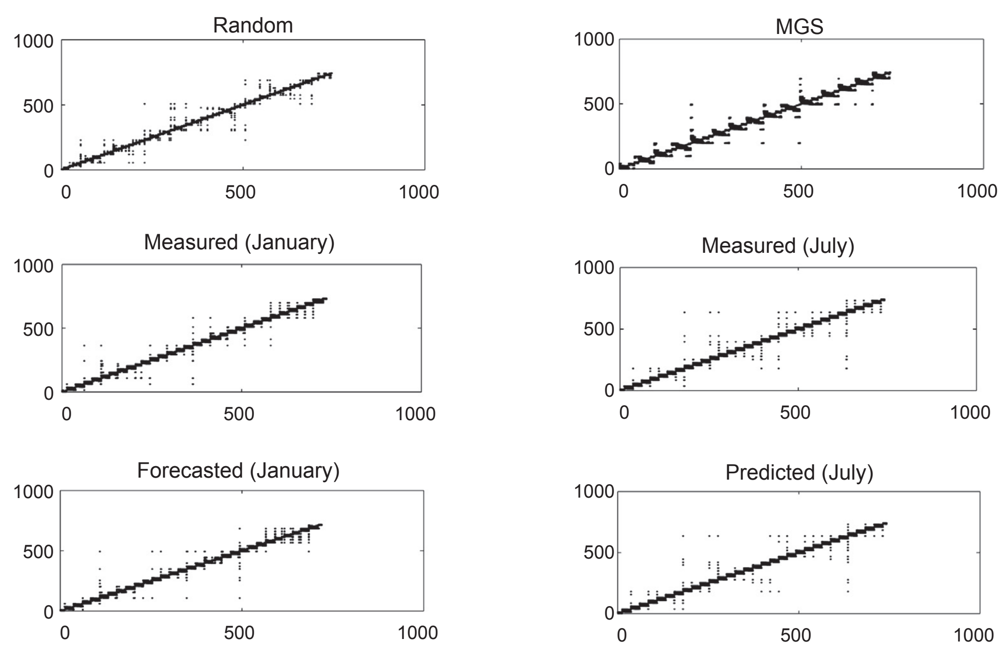


Fig. 17. Network topologies comparison from six time series according to visible graph method.

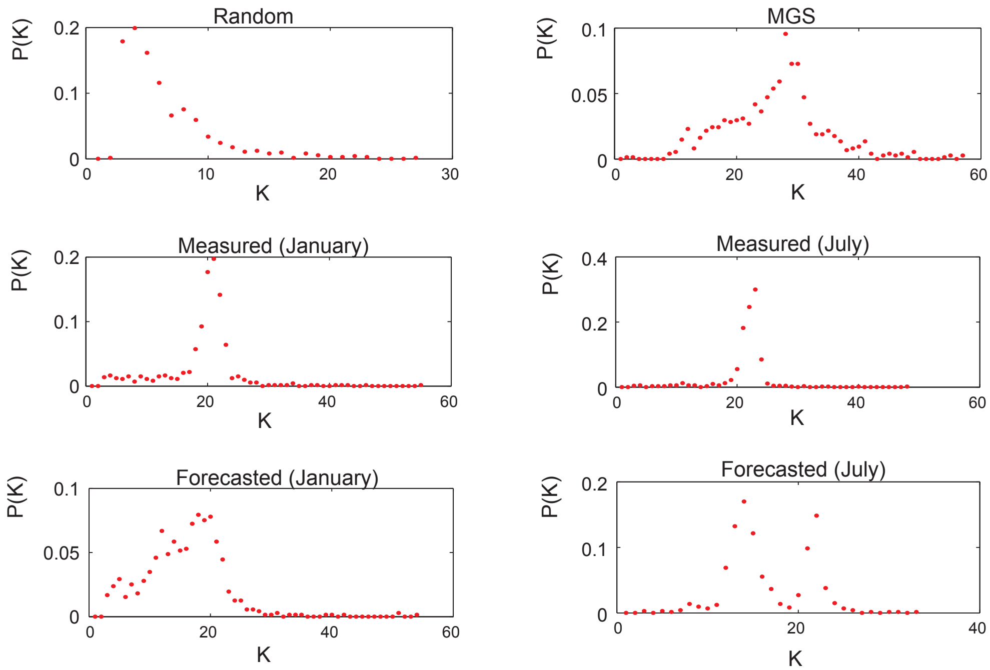


Fig. 18. Network degree distributions comparison from six time series according to visible graph method.

Table 6
Complexity characteristics comparison from six time series according to visible graph method.

Data sequences	AP	CC	AD
Random	5.4455	0.7443	5.4993
MGS	4.3041	0.6862	25.2128
Measured (January)	4.6102	0.8382	18.1641
Forecasted (January)	4.5138	0.7692	14.5063
Measured (July)	4.6141	0.8858	20.8652
Forecasted (July)	4.8181	0.7427	15.6137

Table 7
Complexity characteristics comparison from each month of PV power according to visible graph method.

The measured values				The forecasted values			
Month	AP	CC	AD	Month	AP	CC	AD
1	4.6102	0.8382	18.1641	1	4.5138	0.7692	14.5063
2	4.8815	0.8467	18.3612	2	4.3329	0.7696	14.3134
3	4.5481	0.8599	19.3315	3	4.2975	0.7865	14.6334
4	4.7880	0.8639	20.0139	4	4.4859	0.7748	15.6384
5	5.1523	0.8853	20.9757	5	4.1334	0.7722	16.5714
6	5.5073	0.8843	20.7549	6	4.6996	0.7676	15.8217
7	4.6141	0.8858	20.8652	7	4.8181	0.7427	15.6137
8	5.1591	0.8761	20.5175	8	4.3979	0.7534	15.5283
9	4.6141	0.8858	20.8652	9	4.0719	0.7386	16.0056
10	5.0368	0.8483	19.4536	10	4.5815	0.7421	15.3244
11	4.2022	0.8326	18.3477	11	4.5901	0.7680	13.9026
12	4.2323	0.8449	18.4339	12	4.2502	0.7797	14.2587

(MCESN). The proposed approach can achieve competitive performance of prediction. The effects of measured temperature, humidity, and 24-h-lag information are also studied in the MCESN model. The results show that consideration of temperature and humidity information has negligible effects on the prediction accuracy, and that the historical 24-h-lag information has positive effects on the prediction accuracy in winter and negative effects in summer. The simulation results also indicate that the proposed model could perform accurate 24-h-ahead prediction for sunny days with the correlation coefficient being 99%, and acceptable precision for cloudy days with the correlation coefficient being in the range 91–98%. MCESN could achieve more accurate prediction, compared with ARMA, BP neural networks. Finally, several data characteristics of measured and estimated PV power are qualitatively analyzed. Experimental results show that the seasonality, non-stationarity, complexity, and descriptive statistics characteristics are analogous between measured and estimated values.

There are some open issues for the PV hybrid system. One issue is big data analysis in the PV power forecast. Additional factors, such as cloud cover, sunshine duration, should be considered in the ESN model. Some advanced neural networks, such as convolutional neural network (CNN) and long short-term memory (LSTM) can also be studied for large and complicated applications. Another issue is the load modeling, which is closely related to customer behavior and demand response. The ESN model will be investigated for the load forecast. Furthermore, after day-ahead PV power output and load demand are forecasted, power flow dispatching in the PV hybrid system will be studied under different demand-side programs, e.g., the time-of-use program. Energy efficiency and economic performance must be considered in some rule-based or optimization-based strategies.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.apenergy.2018.02.160.

References

[1] Roldn MI, Monterreal R. Heat flux and temperature prediction on a volumetric receiver installed in a solar furnace. Appl Energy 2014;120:65–74.
[2] Gulin M, Vaak M, Peric N. Dynamical optimal positioning of a photovoltaic panel in all weather conditions. Appl Energy 2013;108:429–38.
[3] Younes S, Muneer T. Clear-sky classification procedures and models using a world-wide data-base. Appl Energy 2007;84(6):623–45.
[4] Mellit A, Benghanem M, Arab AH, Guessoum A. A simplified model for generating sequences of global solar radiation data for isolated sites: using artificial neural network and a library of markov transition matrices approach. Solar Energy 2005;79(5):469–82.
[5] Hassan MA, Khalil A, Kaseb S, Kassem MA. Exploring the potential of tree-based ensemble methods in solar radiation modeling. Appl Energy 2017;203:897–916.
[6] Hassan GE, Youssef ME, Mohamed ZE, Ali MA, Hanafy AA. New temperature-based models for predicting global solar radiation. Appl Energy 2016;179:437–50.

[7] Wong L, Chow W. Solar radiation model. Appl Energy 2001;69(3):191–224.
[8] Deo RC, Wen X, Qi F. A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. Appl Energy 2016;168:568–93.
[9] Mellit A, Pavan AM. A 24-h forecast of solar irradiance using artificial neural network: application for performance prediction of a grid-connected pv plant at trieste, Italy. Solar Energy 2010;84(5):807–21.
[10] Cao J, Lin X. Study of hourly and daily solar irradiation forecast using diagonal recurrent wavelet neural networks. Energy Convers Manage 2008;49(6):1396–406.
[11] Capizzi G, Napoli C, Bonanno F. Innovative second-generation wavelets construction with recurrent neural networks for solar radiation forecasting. IEEE Trans Neural Netw Learn Syst 2012;23(11):1805–15.
[12] Cao S, Weng W, Chen J, Liu W, Yu G, Cao J. Forecast of solar irradiance using chaos optimization neural networks. Power and energy engineering conference, Asia-Pacific. 2009. p. 1–4.
[13] Cao J, Lin X. Application of the diagonal recurrent wavelet neural network to solar irradiation forecast assisted with fuzzy technique. Eng Appl Artif Intell 2008;21(8):1255–63.
[14] Shah ASBM, Yokoyama H, Kakimoto N. High-precision forecasting model of solar irradiance based on grid point value data analysis for an efficient photovoltaic system. IEEE Trans Sust Energy 2015;6(2):474–81.
[15] Elsinga B, van Sark WG. Short-term peer-to-peer solar forecasting in a network of photovoltaic systems. Appl Energy 2017;206:1464–83.
[16] Lou S, Li DHW, Lam JC, Chan WWH. Prediction of diffuse solar irradiance using machine learning and multivariable regression. Appl Energy 2016;181:367–74.
[17] Salcedo-Sanz S, Deo RC, Cornejo-Bueno L, Camacho-Gmez C, Ghimire S. An efficient neuro-evolutionary hybrid modelling mechanism for the estimation of daily global solar radiation in the sunshine state of australia. Appl Energy 2018;209:79–94.
[18] Jaeger H, Haas H. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. Science 2004;304(5667):78–80.
[19] Ozturk MC, Principe JC. An associative memory readout for ESNS with applications to dynamical pattern recognition. Neural Netw 2007;20(3):377–90.
[20] Gallicchio C, Micheli A. Tree echo state networks. Neurocomputing 2013;101:319–37.
[21] Meftah B, Lzoray O, Benyettou A. Novel approach using echo state networks for microscopic cellular image segmentation. Cognit Comput 2016;8(2):237–45.
[22] Shi G, Liu D, Wei Q. Echo state network-based q-learning method for optimal battery control of offices combined with renewable energy. IET Control Theory Appl 2017;11(7):915–22.
[23] Zhong S, Xie X, Lin L, Wang F. Genetic algorithm optimized double-reservoir echo state network for multi-regime time series prediction. Neurocomputing 2017;238:191–204.
[24] Li D, Min H, Wang J. Chaotic time series prediction based on a novel robust echo state network. IEEE Trans Neural Netw Learn Syst 2012;23(5):787–99.
[25] Utpal KumarDas Mea, SoonTey Kok. Forecasting of photovoltaic power generation and model optimization: a review. Renew Sust Energy Rev 2018;81(1):912–28.
[26] Long H, Zhang Z, Su Y. Analysis of daily solar power prediction with data-driven approaches. Appl Energy 2014;126:29–37.
[27] Alessandrini S, Monache LD, Sperati S, Cervone G. An analog ensemble for short-term probabilistic solar power forecast. Appl Energy 2015;157:95–110.
[28] Ferlito S, Adinolfi G, Graditi G. Comparative analysis of data-driven methods online and offline trained to the forecasting of grid-connected photovoltaic plant production. Appl Energy 2017;205:116–29.
[29] Tang L, Yu L, He K. A novel data-characteristic-driven modeling methodology for nuclear energy consumption forecasting. Appl Energy 2014;128:1–14.
[30] Xue F, Li Q, Zhou H, Li X. Reservoir computing with both neuronal intrinsic plasticity and multi-clustered structure. Cogn Comput 2017;9(3):400–10.
[31] Nisbach F, Kaiser M. Developmental time windows for spatial growth generate multiple-cluster small-world networks. Euro Phys J B 2007;58(2):185–91.
[32] Wyffels F, Schrauwen B, Stroobandt D. Stable output feedback in reservoir computing using ridge regression. In: International conference on artificial neural networks; 2008. p. 808–17.
[33] Smith MR. Photovoltaic system performance monitoring guidelines for measurement, data exchange and analysis. Int Organ 1998;29(3):211–8.
[34] Venayagamoorthy GK, Shishir B. Effects of spectral radius and settling time in the performance of echo state networks. Neural Netw 2009;22(7):861–3.
[35] Packard NH, Crutchfield JP, Farmer JD, Shaw RS. Geometry from a time series. Phys Rev Lett 1980;45(9):712–6.
[36] Grassberger P, Procaccia I. Measuring the strangeness of strange attractors. Phys D Nonlinear Phenom 1983;9(1):189–208.
[37] Lacasa L, Luque B, Ballesteros F, Luque J, Nuo JC. From time series to complex networks: the visibility graph. Proc Natl Acad Sci USA 2008;105(13):4972–5.

Building and Environment 136 (2018) 312–321

Contents lists available at ScienceDirect

Building and Environment

journal homepage: www.elsevier.com/locate/buildenv

Energy-efficiency building retrofit planning for green building compliance

Yuling Fan*, Xiaohua Xia

Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria, 0002, South Africa

ARTICLE INFO

Keywords:
Building retrofit
Energy performance certificate
Green building compliance
Rooftop PV system
Economic analysis

ABSTRACT

To promote sustainable development and expedite the progress on moving to a green building sector, the government of South Africa has developed an energy performance certificate (EPC) standard for buildings. A building is required to obtain a certain rating from the EPC in order to comply with the country's green building policy. Therefore, finding optimal retrofit plans for existing buildings are essential given the high investments involved in the retrofit of buildings that do not currently comply with the policy. This paper presents an optimization model to help decision makers to identify the best combination of retrofit options for buildings to ensure policy compliance in the most cost-effective way. The model determines optimal retrofit plans for a whole building in a systematic manner, taking into account both the envelope components and the indoor facilities. A roof top PV system is utilized to reduce the usage of electricity produced from fossil fuels. The model breaks down the long-term investment into yearly short-term investments that are more attractive to investors. Tax incentive program available in the country is taken into account to offset the long payback period of the investment. Economic analysis is also built into the model to help decision makers to make informed decisions. The retrofit of an existing office building is taken as a case study. The results show that 761.6 MWh energy savings and an A rating from the EPC can be obtained with a payback period of 70 months, which demonstrates the effectiveness of the model developed.

1. Introduction

The building sector is responsible for about 30%–40% of energy consumption throughout the world, being one of the largest energy consuming sectors [1,2]. It was also concluded that existing buildings are the main cause of the high energy consumption in the sector given that the replacement rate of existing buildings with new buildings is about 1%–3% per year [3]. In view of this, improving the energy efficiency of existing buildings is a priority task to mitigate environmental impacts of the building sectors [4]. Aligning to this purpose, many countries, such as the US, Australia, China, etc., are developing green building policies to promote the transition to a green building sector. For example, the Leadership in Energy and Environmental Design (LEED) certification program developed by the US Green Building Council, the Green Star rating system developed in Australia, and the evaluation standard for green buildings developed in China all aim to bring down the energy demand of the building sector. For the same purpose, the government of South Africa has recently developed a similar rating system called energy performance certificate (EPC) of buildings [5]. Unlike green building rating systems developed by other countries, the EPC program only focuses on the energy intensity, without considering other indicators such as water usage and indoor air quality, thus, enforcing the building sector to use energy more efficiently.

The EPC system classifies the energy intensity of a building into seven grades ranging from grade A to grade G. Grade A is for the most energy efficient buildings and grade G is for buildings that are the most inefficient. These grades are rated according to the energy intensity of a building in comparison with a reference consumption level determined for buildings of different occupancy classes specified in Ref. [6]. The national green building policy requires that all buildings which are owned, operated and leased by the South African Department of Public Works must reach at least a D rating from the EPC. Enforcement of this green building policy will be extended to all buildings in the country shortly.

While a Grade D rating is mandatory, the government is also promoting higher ratings for the targeted buildings. The existing buildings targeted are usually quite old and are inefficient. Achieving a desired rating for these buildings requires considerable investments. In light of the financial uncertainties and long payback periods of some existing building retrofit projects, a decision support tool is essential to help decision makers to come up with the optimal retrofit plan. This paper aims to fill in this gap by developing an optimization model to identify the optimal retrofit plan aiming at achieving the desired grade with the

* Corresponding author.
E-mail address: ylfan.up@gmail.com (Y. Fan).

https://doi.org/10.1016/j.buildenv.2018.03.044
Received 19 October 2017; Received in revised form 12 March 2018; Accepted 26 March 2018
Available online 29 March 2018
0360-1323/

Nomenclature

α_1	power load densities of people (W/m^2)
α_2	power load densities of lightings (W/m^2)
α_3	power load densities of appliances (W/m^2)
$\overline{C_f}(M)$	the absolute value of the cumulative cash flow at the end of the M-th month ($)
β_t	budget allocated in year t for retrofit ($)
$\delta(t)$	coefficient taking the values from Table 1
$\Delta W(t)$	difference of humidity ratio between the inside air and outdoor air in year t (kg/kg)
η_p	efficiency of the p-th solar panel alternative
η_s	average solar energy to electrical power conversion efficiency
λ_j	thermal conductivity of the j-th alternative of the external wall insulation materials (W/m°C)
λ_k	thermal conductivity of the k-th alternative of the roof insulation materials (W/m°C)
ζ_a	allowance rate set by the government
ζ_t	tax rate for general businesses in South Africa
A_l^{pv}	area of the l-th solar panel alternative (m^2)
A_p^{pv}	area of one solar panel of the p-th alternative (m^2)
A_e	available roof area for the PV power supply system installation (m^2)
A_{flr}	areas of the floor of the building (m^2)
A_g	gross area of the building (m^2)
A_{rof}	areas of the roof of the building (m^2)
A_{wal}	areas of the walls of the building (m^2)
A_{win}	areas of the windows of the building (m^2)
C	number of chiller alternative
$C(t)$	retrofit cost in year t ($)
C_c^{chi}	cost of the c-th chiller alternative ($)
$C_f(M+1)$	the discounted cash flow in the ($M+1$)-th month ($)
C_h^{pum}	cost of the h-th heat pump alternative ($)
C_i^{win}	cost of the i-th window alternative ($/m^2)
C_j^{wal}	cost of the j-th wall insulation material alternative ($/m^2)
C_k^{rof}	cost of the k-th roof insulation material alternative ($/m^2)
C_l	latent heat factor of air (W/(L/s))
C_l^{pv}	unit cost of the l-th solar panel alternative ($)
C_s	sensible heat factor of air (W/(°C L/s))
$C_{dd}(t)$	cooling degree days in year t (°Ch)
$C_{l_m}^{lig_m}$	unit cost of the l_m-th alternative of the lightings used to retrofit the m-th type of existing lighting technologies ($)
d	discount rate
d_j	thickness of the j-th alternative of the external wall insulation materials (m)
d_k	thickness of the k-th alternative of the roof insulation materials (m)
$E_p(t)$	net energy consumed by the building in year t (kWh/m^2)
E_r	reference of net annual energy consumption (kWh/m^2)
$E_{cool}(t)$	energy consumed by the chillers in year t
$E_d(t)$	energy usage of lighting systems and appliances in year t
$E_{heat}(t)$	electrical energy used for the heating purpose in year t
$E_i(t)$	internal heat gain in year t
$E_{lc}(t)$	latent heat gain in year t
$E_{lh}(t)$	latent heat gain in year t
E_{pre}	baseline energy consumption of the building before retrofit
$E_{pv}(t)$	energy produced by the PV system in year t
$E_{sc}(t)$	sensitive heat gain in year t
$E_{sh}(t)$	sensitive heat loss in year t
$E_{sl}(t)$	solar heat gain of the cooling load in year t
$E_{tc}(t)$	transmission heat gain of the cooling load in a general building in year t
$E_{th}(t)$	transmission heat loss through the envelope in year t
$E_{tot}(t)$	total energy consumption of the building during year t
$ES(t)$	resultant energy savings in year t
H	number of heat pump alternative
$H_{dd}(t)$	heating degree days in year t (°Ch)
$HSPF(t)$	heating seasonal performance factor in year t (Btu/Wh)
$HSPF_h$	performance coefficient of the h-th heat pump alternative
I	number of window alternative
$I_{pv}(t)$	solar irradiation on the PV power supply system during year t (Wh/m^2)
$I_{win}(t)$	solar irradiance on windows in year t (W/m^2)
J	number of wall insulation material alternative
K	number of roof insulation material alternative
L_m	number of lighting alternative for retrofitting the m-th type of existing lights
M	the month after the investment at which the last negative cumulative discounted cash flow occurs
m	number of existing lightings' type
N_{l_m}	maximum quantity of the m-th type of existing lamps available for retrofit
$N_{lig_m}(t)$	number of the m-th type of existing lighting technology retrofitted in year t
$N_{pv}(t)$	number of the selected solar panel to be installed in year t
P	number of solar panel alternative
$p(t)$	electricity price in year t ($/kWh)
P_a	total power of the appliances in the building in year t (W)
$P_l(t)$	total power of the lights in the building in year t (W)
Q_s	air flow rate (L/s)
$R(t)$	the actual monetary incentive in year t
$SEER(t)$	seasonal energy efficiency ratio (Btu/Wh)
$SEER_c$	performance coefficient of the c-th chiller alternative
$SHGC(t)$	solar heat gain coefficient as a function of incident angle in year t
T	project period
$T_c(t)$	cooling time in year t (h)
$T_h(t)$	heating time in year t (h)
T_p	payback period measured (months)
$T_s(t)$	solar radiation time in year t
$T_d(t)$	occupancy time of the lightings and appliances in year t (h)
$T_{oc}(t)$	occupancy time during the cooling season in year t (h)
U_i	thermal transmittance of the i-th window alternative (W/m^2°C)
U_r	thermal transmittance of the roof before retrofit (W/m^2°C)
U_w	thermal transmittance of the wall before retrofit (W/m^2°C)
$U_{flr}(t)$	thermal transmittances of the floor in year t (W/m^2°C)
$U_{rof}(t)$	thermal transmittances of the roof in year t (W/m^2°C)
$U_{wal}(t)$	thermal transmittances of the walls in year t (W/m^2°C)
$U_{win}(t)$	thermal transmittances of the windows in year t (W/m^2°C)
w_1	positive weight
w_2	positive weight
$x_c^{chi}(t)$	retrofit state of the c-th chiller alternative in year t, similar to $x_i^{win}(t)$
$x_h^{pum}(t)$	retrofit state of the h-th heat pump alternative in year t, similar to $x_i^{win}(t)$
$x_i^{win}(t)$	retrofit state of the i-th alternative of the windows
$x_j^{wal}(t)$	retrofit state of the j-th alternative of the insulation materials for the external walls in year t, similar to $x_i^{win}(t)$
$x_k^{rof}(t)$	retrofit state of the k-th alternative of the insulation materials for the roof in year t, similar to $x_i^{win}(t)$
$x_p^{pv}(t)$	retrofit state of the p-th solar panel alternative in year t, similar to $x_i^{win}(t)$
$x_{l_m}^{lig_m}(t)$	retrofit state of the l_m-th alternative of the lightings for the m-th type of existing lightings in year t, similar to $x_i^{win}(t)$

maximum possible financial benefits.

To reduce energy usage of buildings, it is noted that energy consumption in buildings are attributed to two main subsystems. That is, the energy dissipated by the envelope/enclosure that separates the interior and exterior environments and the energy consumed by the facilities and appliances inside the building.

In the literature, the energy efficiency of buildings was classified into performance efficiency, operation efficiency, equipment efficiency, and technology efficiency, definitions of which are given in published works such as [7–9]. Efforts on improving the energy efficiency of the existing buildings, regarding the two subsystems mentioned earlier, were mainly focused on the technology and equipment efficiency levels from both power supply and demand sides. At technology efficiency level, efforts have been made to introduce renewable power generating technologies to buildings, including solar systems [10,11], wind systems [12], etc. from the energy supply side. At the same time, many energy efficiency technologies have been developed and reported from the demand side. These include development of insulation materials [13–15], energy-efficient appliances [16], etc. At the equipment level, the maintenance of envelope system, ventilation and air conditioning (HVAC) systems and lighting systems was also studied [17–20]. At the operation level, studies have been focused on the optimal scheduling, coordination, and control of the indoor appliances/facilities including heating, HVAC systems, lighting systems, smart appliances [21,22], etc. to reduce both energy consumption and cost for individual or a group of buildings. Performance level studies were mainly focusing on the impacts of existing building on the environment and the electricity grid. Such as the ones reported in Refs. [23–25].

The retrofit planning, a technology level problem, has not been well studied in the literature. Majority of the reported studies in this area focuses on developing guidelines to facilitate the retrofit process or on the cost-benefit analysis of retrofits. This means that the reported studies mostly focusing on policy or management level, that deals with long term impacts of the retrofit or the procedures of a building retrofit at a high level. For example [26], concluded that the key enabling factors for the implementation of green building retrofits include introduction of a project facilitation team, performance contracting, etc. [27] presented a state-of-the-art review of all building energy retrofit activities and developed a conceptual method for determining the most cost-effective retrofit measures for a particular project [28]. emphasized the importance of the selection of optimization objectives in the decision making process for building retrofits and developed a decision matrix to guide the objective selection process [29]. looked into developing a building retrofit index to guide the selection of building with the best retrofit potential at regional and national scales to support green building policies making use of a clustering method. Cost-benefit analysis of building retrofits was reported in Ref. [30] aiming at offer policy makers and managers to develop incentive mechanisms and management interventions to promote the implementation of building retrofit programs [31]. presented a life cycle analysis approach for building retrofits with similar objective of helping identify retrofit options in early planning stage. As pointed out by Webb in a review paper [32], echoed by another review paper [27], that although the development of building retrofit criteria, performance simulation and analysis tools, and consistent guidelines certainly aids the building retrofit process, methods to identify the most cost-effective retrofit measures for particular projects is still a major technical challenge.

In this regard, several studies presented detailed mathematical models to determine the optimal retrofit options in a building from several aspects. In particular, a mixed integer model was developed for the indoor appliances retrofit to reduce energy consumption in Ref. [33] from a control system point of view. In the building retrofit for green building certification context, a particular study [34] reported an optimization model to reduce energy and water consumptions of an existing building aiming at LEED certification. In Ref. [34], the optimal retrofit planning problem was formulated as a mixed integer programming problem, which only considered indoor appliances such as light bulbs and washing machines. Because of the envelope structure's significant contribution to a building's energy consumption, the retrofit planning for the envelope components of buildings was also studied recently in Refs. [35–37].

However, no study on the systematic retrofit planning for the whole building including the envelope and the indoor systems has been reported so far. Only indoor facilities were considered in Refs. [19,33,34]. The thermal dynamics of the building envelope, which contributes up to 40% of energy consumption of buildings, was ignored in those studies. Previous studies on the building envelope energy consumption reduction, however, didn't consider the energy usage inside the building [35–38]. Consequently, no study was done on retrofit planning considering the interactions between the indoor and envelope systems of the building in terms of energy consumption. This is because of the technical difficulties associated with the building retrofit problem considering both the envelope and indoor systems. When only indoor appliances are considered, the problem can be formulated as a linear mixed integer problem. However, the problem becomes highly nonlinear and of high-dimensional when both the indoor and envelope systems are involved. In addition, no study on the optimal building retrofit plan considering the EPC rating system, which looks at the energy intensity of a whole building and calls for a systematic whole-building retrofit approach, has been conducted.

Therefore, the purposes of this study are to

- develop a mathematical model that can determine an optimal retrofit plan for the whole building aiming at maximizing the energy savings, minimizing the payback period of the project, and achieving a desired energy rating from the EPC systematically;
- help decision makers to directly obtain the best retrofit solution to a specific building without the need of complex human decision making process;
- provide a detailed analysis of the retrofit plan given by the model developed in terms of its financial implications such as payback period, NPV, etc.

Although operational level optimization is also an important aspect to improve energy efficiency of existing buildings by optimal sizing, matching and timing control of facilities in the building. This is however out of the scope of this study and not considered.

The main contributions of this study are stated in the following. Firstly, a systematic approach to determine the optimal retrofit plan for existing buildings considering both the envelope systems and the indoor systems and their interactions to reduce the energy consumption and to ensure compliance with a green building policy with reference to the EPC rating system is presented. The optimal retrofit plan obtained can help a building to achieve a desired energy rating from the EPC rating system in a cost-effective manner. Secondly, factors including energy savings and economic benefits, which are important to decision makers, are built into the proposed optimization model to make sure that the economic benefits of an investment project are maximized and the desired energy savings is achieved. Thirdly, the proposed model treats the retrofit plan as a multi-year project with improving efficiency targets in the consecutive years. That is to say, the model breaks down the one-time long-term project into smaller projects over multiple financial years with shorter payback periods. This is of great help to mitigate the concerns of the investors. In view that obtaining the best rating (grade A) usually requires a significant amount of investment with a long payback period and the high economic uncertainties, breaking the investment down in short-term ones helps to attract investments for similar building retrofit projects. The proposed approach in this study will make sure that at least the so-called 'low-hanging fruits' projects, which generate noticeable savings with a relatively small investment, for energy efficiency improvement will be implemented in the starting years of the retrofit project. Lastly, the government of South Africa, struggling from sever energy supply constraints, has implemented a series of initiatives to promote efficient utilization of the country's limited power generating capacity in recent years. The tax incentive program introduced under the section 12L of the income tax act is one of these initiatives. It allows business owners to claim a deduction of their taxable income according to their energy savings over a year comparing to their baseline consumption in the previous year. The 12L tax incentive program helps to bring in an additional cash flow by means of reduced tax paid by the building owner, which can be used to fund the new retrofit projects in the coming years and can further shorten the payback period of the retrofit project. This tax incentive program is also considered in the optimal retrofit planning method.

Relevance of this research to the building retrofit field can be stated from two aspects. From the application point of view, this study develops a powerful decision support tool for the whole building energy efficiency retrofit aiming at a green building rating taking into account all possible retrofit activities, interactions between the indoor and envelope systems of a building, and financial incentives over several years. From the academic perspective, the presented optimization model adds value to the literature on the green building retrofit by introducing a systematic model capable of optimizing the retrofit actions of both envelope and indoor facilities of a building simultaneously. This systematic approach essentially develops a retrofit planning tool for buildings involving multi-technologies, which was found to be difficult [39]. It also features a multi-year planning architecture that helps to ease the mind of investors and helps to evaluate the financial and energy savings benefits of the retrofit over a realistic multi-year scale [39]. Moreover, the formulated optimal retrofit planning problem is a nonlinear mixed-integer programming (NMIP) problem that cannot be solved by conventional optimization techniques and consequently, a genetic algorithm (GA) is developed in this study to solve this NMIP problem. It should be noted that the focus of this study is developing the optimization model to “identify the most cost-effective retrofit measures for particular projects” and not the optimization algorithm to solve this problem. Although a GA based algorithm is adopted, it should be noted that this problem can be solved by other algorithms as well. Investigation of the most efficient algorithm to solve the formulated problem will be reported in our future works.

The remainder of this paper includes five parts. Modeling of the building energy consumption is presented in Section 2 followed by the optimal retrofit problem formulation in Section 3. After that, a case study covering all aspects of the whole building retrofit problem is given in Section 4 and conclusions are drawn in Section 5.

2. Energy modeling of the building

The energy consumption of the various components of a building must be mathematically modeled before the retrofit problem can be formulated. This section presents the equations that govern the energy usage of a building. Specifically, the heating and cooling energy usages are modeled considering the heat flows through the envelope materials and the characteristics of the heating and cooling facilities. The energy consumption of lighting system and appliances inside the building is then modeled. Lastly, a photovoltaic (PV) system is included in the model to produce electricity for the building in order to help it to reach the desired grade. The motivation of such a PV system is because that South Africa is one of the countries in the world that has the best solar resource, and that other energy saving technologies such as district heating infrastructures are not available in the country. It is however noted that if other energy saving systems are available, they can be modeled and then incorporated in the optimal retrofit plan model developed in this study, which sets a general framework for the optimal retrofit plan with reference to the EPC rating system.

The impacts of the envelope components on the energy consumption of the building are modeled first followed by the energy consumption model of the lighting and appliances. Modeling of the rooftop PV power supply system comes at the end of this section.

In the following subsection, equations for the cooling and heating loads calculation are derived from Refs. [40,41] if not specifically stated otherwise.

2.1. Cooling energy consumption

In a general building, the energy consumption for the cooling load includes the following parts: transmission heat gain, infiltration and ventilation heat gain, solar heat gain, and internal heat gain.

2.1.1. Transmission heat gain

The transmission heat gain of the cooling load in a general building in year t can be determined by

$$E_{tc}(t) = C_{dd}(t)(A_{win}U_{win}(t) + A_{wal}U_{wal}(t) + A_{rof}U_{rof}(t) + A_{flr}U_{flr}(t)) \tag{1}$$

In this study, the floor of the building is not considered to be retrofitted. Hence, the thermal transmittance of the floor $U_{flr}(t)$ keeps unchanged. The thermal transmittances of the other envelope components of the building after the retrofit are calculated by

$$U_{win}(t) = \sum_{i=1}^{I} x_i^{win}(t)U_i, \tag{2}$$

$$U_{wal}(t) = \sum_{j=1}^{J} x_j^{wal}(t)\frac{U_w\lambda_j}{U_w d_j + \lambda_j}, \tag{3}$$

$$U_{rof}(t) = \sum_{k=1}^{K} x_k^{rof}(t)\frac{U_r\lambda_k}{U_r d_k + \lambda_k}, \tag{4}$$

in which $x_i^{win}(t)$ denotes the state of the i-th alternative of the windows, i.e., when $x_i^{win}(t) = 1$, it is chosen to retrofit the existing window in year t, while if $x_i^{win}(t) = 0$, it is not chosen.

2.1.2. Infiltration and ventilation heat gain

The infiltration and ventilation heat gains of the cooling load in a general building consist of sensible and latent components. The sensitive heat gain in year t can be calculated by

$$E_{sc}(t) = C_sQ_sC_{dd}(t). \tag{5}$$

The latent heat gain in year t can be calculated by

$$E_{lc}(t) = C_lQ_s\Delta W(t)T_c(t). \tag{6}$$

2.1.3. Solar heat gain

The solar heat gain of the cooling load in a general building in year t can be calculated by

$$E_{sl}(t) = A_{win}I_{win}(t)SHGC(t)T_s(t). \tag{7}$$

In the calculation of $SHGC(t)$, the shading factor is not considered in this study.

2.1.4. Internal heat gain

The internal heat gain of the cooling load in a general building mainly results from people, lightings and appliances. It can be calculated by

$$E_i(t) = (\alpha_1 + \alpha_2 + \alpha_3)A_gT_{oc}(t). \tag{8}$$

2.1.5. Energy consumption of the cooling load

The cooling loads detailed in Sections from 2.1.1 to 2.1.4 are supplied by chillers installed in the building. The following equation is used to determine the energy consumed by the chillers to supply these cooling loads [42].

$$E_{cool}(t) = \frac{E_{tc}(t) + E_{sc}(t) + E_{lc}(t) + E_{sl}(t) + E_i(t)}{SEER(t)}. \tag{9}$$

SEER is a ratio of the cooling output in BTU over the cooling season to the used watt-hours electricity input during the same period measured in Btu/Wh. When the exiting chiller is retrofitted by a new one, the resulting SEER is determined by

$$SEER(t) = \sum_{c=1}^{C} x_c^{chi}(t) SEER_c. \tag{10}$$

2.2. Heating energy consumption

The heating load for a building includes two parts, namely, transmission heat loss and infiltration and ventilation heat loss.

2.2.1. Transmission heat loss

The transmission heat loss through the envelope in year t is calculated by

$$E_{th}(t) = H_{dd}(t)(A_{win} U_{win}(t) + A_{wal} U_{wal}(t) + A_{rof} U_{rof}(t) + A_{flr} U_{flr}(t)) \tag{11}$$

2.2.2. Infiltration and ventilation heat loss

The infiltration and ventilation heat loss consists of sensitive and latent heat losses. The sensitive heat loss in year t can be calculated by

$$E_{sh}(t) = C_s Q_s H_{dd}(t), \tag{12}$$

and the latent heat gain in year t can be calculated by

$$E_{lh}(t) = C_l Q_s \Delta W(t) T_h(t). \tag{13}$$

2.2.3. Energy consumption of the heating load

The heat loads determined in Sections 2.2.1 and 2.2.2 are supplied by heat pumps in the HVAC system. Accounting for the efficiency of the heat pump, the electrical energy used for the heating purpose can be determined by Ref. [42].

$$E_{heat}(t) = \frac{E_{th}(t) + E_{sh}(t) + E_{lh}(t)}{HSPF(t)}. \tag{14}$$

HSPF is defined as the heating output in BTU during the heating season divided by the total electricity energy input in watt-hours during the same period measured in Btu/Wh. When the heat pump is retrofitted, the resulting HSPF can be calculated by

$$HSPF(t) = \sum_{h=1}^{H} x_h^{pum}(t) HSPF_h. \tag{15}$$

2.3. Lighting and appliance energy consumption

In addition to heating and cooling energy consumption, lighting systems and appliances in the building also consume energy. This part of energy usage in year t is calculated by

$$E_d(t) = (P_l(t) + P_a) T_d(t). \tag{16}$$

2.4. PV system energy production

The energy produced by the PV system in year t depends on the local solar radiation and is calculated by Refs. [43,44]:

$$E_{pv}(t) = I_{pv}(t)\eta_s \sum_{p=1}^{P} x_p^{pv}(t)\eta_p \sum_{p=1}^{P} x_p^{pv}(t) A_p^{pv} \sum_{t=1}^{t} N_{pv}(t). \tag{17}$$

Table 1
Energy performance scale.

Grade	Requirement
A	Energy intensity < $0.3E_r$
B	$0.3E_r \le$ Energy intensity < $0.6E_r$
C	$0.6E_r \le$ Energy intensity < $0.9E_r$
D	$0.9E_r \le$ Energy intensity < $1.1E_r$
E	$1.1E_r \le$ Energy intensity < $1.4E_r$
F	$1.4E_r \le$ Energy intensity < $1.7E_r$
G	Energy intensity $\ge 1.7E_r$

2.5. Total energy consumption of a building

Summing up all the energy consumption and generation in the building from Section 2.1 to Section 2.4, the total energy consumption of the building during year t can be calculated by

$$E_{tot}(t) = E_{cool}(t) + E_{heat}(t) + E_d(t) - E_{pv}(t). \tag{18}$$

3. The energy-efficiency retrofit problem

The objective of the retrofit is to obtain a desired EPC rating in order to comply with the green building policy. Therefore, the details on the EPC rating system are briefly discussed first.

3.1. EPC for buildings

The EPC rating system assigns a grade from A (most efficient) to G (most inefficient) to a building by comparing its actual net annual energy usage in kilowatt hours per square meter to a reference value set by the national standard SANS10400-XA [6]. To be exact, the requirements to reach different energy performance grades are detailed in Table 1, in which E_r is the reference net annual energy consumption in kilowatt hours per square meter. The value of E_r for a target building is determined by the occupancy type and climate zone of the building which can be found in Ref. [6]. For instance, the value of E_r is set to 190 kWh/m^2 for an office building operating in climate zone 2 while it is set to 630 kWh/m^2 for a hotel operating in climate zone 6.

The minimum requirement for target buildings is to obtain a D rating from the EPC at least. Therefore, the main aim of the presented optimization model in this study is to design an optimal energy efficiency retrofit plan for existing buildings that will ensure compliance with the green building policy and maximize the economic benefits of the retrofit.

As mentioned in Section 1, the one-time long-term investment project is breakdown into yearly investments with shorter payback periods. Keep in mind that tighter regulation may come into effect in the coming years, the targeted rating for each consecutive year can be different. Therefore, the retrofit plan problem can be put in the following optimization problem format.

$$\begin{aligned} &\textbf{max} \quad \text{energy savings} \\ &\textbf{min} \quad \text{payback period} \\ &\textbf{s. t.} \quad \text{desired EPC rating, and} \\ &\qquad\quad \text{budget available} \end{aligned} \tag{19}$$

In this study, the retrofit actions focus on the retrofit of envelope components, including windows, walls, and roof; the replacement of the chiller, heat pump in the HAVC system and the lighting fixtures in the building by more efficient models; and installation of a rooftop PV power supply system to produce electricity for the building. Details of this optimization problem are formulated in the following subsections with the following assumptions:

1) The occupancy type of the building over the planning period remains unchanged, i.e., an office building will continue to serve as an

office building and will not be used for other purposes.
2) Proper maintenance of the retrofitted items is practiced such that the resulting energy savings is persistent.
3) Any existing item will only be retrofitted once during the project period. For instance if the heat pump is retrofitted by a certain alternative in year one, no further retrofit of this alternative will happen during the project period.

3.2. Decision variables

Assume that there are I alternatives of windows, J alternatives of wall insulation materials, K alternatives of roof insulation materials, C alternatives of chillers, H alternatives of heat pumps, and P alternatives of solar panels available for the retrofit. And that, there are m types of existing lightings to be retrofitted and L_m alternatives for retrofitting the m-th type. Let

$$
\begin{aligned}
X_{win} &= [x_1^{win}(1),...,x_I^{win}(1),...,x_1^{win}(T),...,x_I^{win}(T)]\\
X_{wal} &= [x_1^{wal}(1),...,x_J^{wal}(1),...,x_1^{wal}(T),...,x_J^{wal}(T)]\\
X_{rof} &= [x_1^{rof}(1),...,x_K^{rof}(1),...,x_1^{rof}(T),...,x_K^{rof}(T)]\\
X_{chi} &= [x_1^{chi}(1),...,x_C^{chi}(1),...,x_1^{chi}(T),...,x_C^{chi}(T)]\\
X_{pum} &= [x_1^{pum}(1),...,x_H^{pum}(1),...,x_1^{pum}(T),...,x_H^{pum}(T)]\\
X_{pv} &= [x_1^{pv}(1),...,x_P^{pv}(1),...,x_1^{pv}(T),...,x_P^{pv}(T)]\\
X_{lig_m} &= [x_1^{lig_m}(1),...,x_{Lm}^{lig_m}(1),...,x_1^{lig_m}(T),...,x_{Lm}^{lig_m}(T)]\\
N &= [N_{pv}(1),...,N_{pv}(T), N_{lig_1}(1),...,N_{lig_1}(T),...,\\
&\quad N_{lig_m}(1),...,N_{lig_m}(T)]
\end{aligned}
$$

The decision variable of the optimization problem is then given by:

$$
X = [X_{win}, X_{wal}, X_{rof}, X_{chi}, X_{pum}, X_{pv}, X_{lig_1},...,\\
X_{lig_m}, N].
$$

3.3. The objective function

As seen in (19), the objectives of the retrofit problem will maximize energy savings and minimize the payback period of the retrofit.

Energy savings resulted from the retrofit is calculated by

$$ES(t) = E_{pre} - E_{tot}(t). \tag{20}$$

Taking into annual discounts of the cash flow, the following formula is used to determine the discounted payback period of the retrofit project [45].

$$T_p = M + \frac{|\overline{C_f}(M)|}{C_f(M+1)}. \tag{21}$$

In the calculation of cash flows of the investment, the tax incentive program is taken into account. The incentive program promotes green development by reducing the amount of total taxable incomes of the owner of the buildings according to the energy savings achieved annually. Therefore, the actual monetary incentive for the building owner is calculated by multiplying the offset amount by the tax rate of the individual/business. It can be obtained by

$$R(t) = (E_{tot}(t-1) - E_{tot}(t))\zeta_a\zeta_t. \tag{22}$$

Combining Eqs. (20) and (22), the discounted cash flows of the retrofit problem can be obtained by

$$C_f(t) = \frac{-C(t) + p(t)ES(t) + R(t)}{(1+d)^t}. \tag{23}$$

The retrofit cost in year t is calculated by

$$
\begin{aligned}
C(t) &= A_{win}\sum_{i=1}^{I} x_i^{win}(t)C_i^{win} + A_{wal}\sum_{j=1}^{J} x_j^{wal}(t)C_j^{wal}\\
&\quad + A_{rof}\sum_{k=1}^{K} x_k^{rof}(t)C_k^{rof} + \sum_{c=1}^{C} x_c^{chi}(t)C_c^{chi}\\
&\quad + \sum_{h=1}^{H} x_h^{pum}(t)C_h^{pum} + N_{pv}(t)\sum_{p=1}^{P} x_p^{pv}(t)C_p^{pv}\\
&\quad + \sum_{m=1}^{m}\sum_{l_m=1}^{L_m} x_{l_m}^{lig_m}(t)C_{l_m}^{lig_m}N_{lig_m}(t).
\end{aligned} \tag{24}
$$

Eventually, the multiple objective optimization problem that maximizes the energy savings and minimizes the payback period is converted to a single objective optimization problem making use of the weighted sum method [46–48] with the following combined cost function

$$J = -w_1\sum_{t=1}^{T} ES(t) + w_2T_p. \tag{25}$$

During the optimization process, the values of the two objectives are normalized with respect to their base case for the convenience of tuning the weighting factors in the optimization process.

3.4. The constraints

The constraints of the optimization problem consist of three parts. The first constraint is the limit on the available budget, which is described as

$$C(t) \le \beta_t. \tag{26}$$

The second one is to ensure target buildings to obtain desired EPC ratings. It is described as

$$E_p(t) < \delta(t)E_r, \tag{27}$$

where $\delta(t)$ takes the values from Table 1. For example, $\delta(t) = 1.1$ ensures that the energy performance of the building must reach grade D at least in year t. The energy performance of the building in year t can be described by Ref. [5]:

$$E_p(t) = \frac{E_{tot}(t)}{A_g}. \tag{28}$$

The third kinds of constraints are some physical limits of the retrofit, including the limit on usable area of the roof for PV system installation

$$\sum_{t=1}^{T}\sum_{p=1}^{P} x_p^{pv}(t)A_p^{pv}N_{pv}(t) \le A_e, \tag{29}$$

and boundary limits on the decision variables

$$
\begin{cases}
\sum_{i=1}^{I} x_i^{win}(t) \in \{0,1\},\ i = 1,2,...,I\\
\sum_{j=1}^{J} x_j^{wal}(t) \in \{0,1\},\ j = 1,2,...,J\\
\sum_{k=1}^{K} x_k^{rof}(t) \in \{0,1\},\ k = 1,2,...,K\\
\sum_{c=1}^{C} x_c^{chi}(t) \in \{0,1\},\ c = 1,2,...,C\\
\sum_{h=1}^{H} x_h^{pum}(t) \in \{0,1\},\ h = 1,2,...,H\\
\sum_{p=1}^{P} x_p^{pv}(t) \in \{0,1\},\ p = 1,2,...,P\\
\sum_{l_m=1}^{L_m} x_{l_m}^{lig_m}(t) \in \{0,1\},\ l_m = 1,2,...,L_m.
\end{cases} \tag{30}
$$

The each of the $x_i^{win}(t)$, $x_j^{wal}(t)$, $x_k^{rof}(t)$, $x_c^{chi}(t)$, $x_h^{pum}(t)$, $x_p^{pv}(t)$, $x_p^{pv}(t)$ and $x_{l_m}^{lig_m}(t)$ takes the value of either zero or one.

4. Case study

To analyze the effectiveness and feasibility of the optimization model, an existing office building situated in Pretoria, South Africa, the Koppen-Geiger climate of which is Cwa, is used as a case study in this section. The building has a gross area of 568 m^2 and consists of two floors with the same structure, which is shown in Fig. 1.

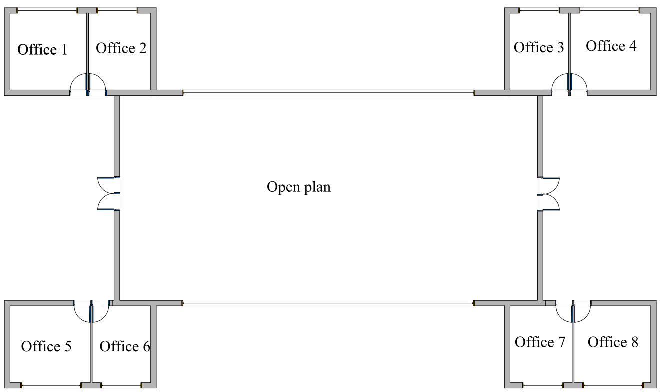


Fig. 1. Floor plan of the office building under study.

Table 2
Alternatives of windows.

i	Alternatives	U_i (W/m°C)	C_i^{win} (\$/m^2)
1	Single glazing, aluminum frame	1.25	21.00
2	Double glazing, uncoated air-filled metallic frame	0.82	38.00
3	Double glazing, tinted uncoated air–filled metallic frame	0.49	50.00
4	Double glazing, tinted coated air–filled metallic frame	0.38	80.00
5	Double glazing, low-e window, air-filled metallic frame	0.32	97.00

Table 3
Alternatives of external wall insulation materials.

j	Alternatives	d_j(m)	λ_j(W/m°C)	C_j^{wal}(\$/m^2)
1	Stone wool	0.03	0.034	14.49
2	Glass wool	0.05	0.038	16.32
3	EPS	0.03	0.036	9.84
4	EPS	0.07	0.036	13.45
5	EPS	0.08	0.036	14.37
6	EPS	0.08	0.033	21.10
7	EPS	0.04	0.036	10.44
8	EPS	0.06	0.036	12.32
9	SPF	0.02	0.042	8.23
10	Cork	0.01	0.040	3.93
11	Cork	0.10	0.040	23.13
12	Cork	0.15	0.040	34.70
13	Cork	0.30	0.040	69.38

Table 4
Parameters of roof insulation materials.

k	Alternatives	d_k(m)	λ_k(W/m°C)	C_k^{rof}(\$/m^2)
1	SPF	0.020	0.042	8.23
2	EPS	0.030	0.033	5.57
3	EPS	0.040	0.033	7.22
4	EPS	0.050	0.033	8.85
5	EPS	0.060	0.033	10.49
6	EPS	0.070	0.033	12.15
7	EPS	0.080	0.033	13.79
8	EPS	0.040	0.034	15.00
9	Stone wool	0.065	0.037	31.78
10	Stone wool	0.105	0.037	44.84

Table 5
Parameters of chiller alternatives.

c	Alternatives	SEER	C_h^{pum}(\$)
1	Trane chiller type 1	17.0	8580
2	Trane chiller type 2	15.0	7590
3	Trane chiller type 3	12.0	6435

Table 6
Parameters of heat pump alternatives.

h	Alternatives	HSPF	C_c^{chi}(\$)
1	Trane heat pump type 1	9.5	7920
2	Trane heat pump type 2	8.6	7425
3	Trane heat pump type 3	7.9	5775

Table 7
Parameters of solar panels.

l	Alternatives	C_l^{pv}(\$)	η_l	A_l^{pv}(m^2)
1	STP255-20/WD	900.78	15.7%	1.627
2	YL190P-23B	592.62	14.7%	1.297
3	YL265C-30B	942.30	16.3%	1.624
4	CS6X-300P	870.33	15.6%	1.919
5	HSL60P6-PB-1-240B	704.82	14.8%	1.616
6	Sharp ND 245 Poly	1023.12	14.9%	1.642
7	SW 275 MONO	1042.50	16.4%	1.593

The existing windows of the building are single glazing and the existing roof, walls and floor have no thermal insulation. The retrofit plan for this building includes a set of actions. For the envelope, retrofit of the windows using better alternatives is considered and insulation materials are considered to be installed to the walls and roof. The existing lighting system is to be upgraded by more energy efficient models and the chiller and heat pump in HVAC system are to be retrofitted with their more efficient counterparts. Installation of a PV power supply system is also part of the retrofit options. The detailed information of the systems/components used for the retrofit, including windows, wall and roof insulation materials, chiller, heat pump, and PV panels, is given in Tables 2–8. In Table 8, three alternative lighting technologies are listed to retrofit the corresponding existing technologies. The baseline energy consumption of the building before the retrofit is 120.6 MWh per year. The discount rate involved in the optimization process is set at 6% according to South Africa statistics.[1] The rate of increase in the electricity price in South Africa is determined as 12.69% according to the average increase rate of electricity published by Eskom, which is the largest utility in South Africa.[2] The values of other parameters involved in the optimization model are taken from the national code on the energy efficiency in buildings [49].

For this particular building studied, EPC rating system gives it a E rating before the retrofit. Therefore, to improve the energy efficiency in order to reach D rating for policy compliance and subsequently C, B and A rating in the following years, the retrofit plan considers an implementation period of the retrofit of four years. In particular, the retrofit plan will improve the EPC rating of this building to D in year one and to grade C in year two, and eventually to grade A in year four to first ensure policy compliance and then pursuit better energy efficiency.

In this study, a genetic algorithm (GA) is employed to solve the multi-objective optimization problem [50,51]. With the genetic algorithm, the optimization problem is solved with the weighting factors set to $w1 = 0.7$ and $w2 = 0.3$. The budgets allocated to each year for the

[1] http://www.statssa.gov.za/.

[2] Eskom. Historical average price increase. http://www.eskom.co.za/CustomerCare/TariffsAndCharges/Pages/Tariff_History.aspx. Accessed 7th Dec. 2016.

Table 8
Parameters of lighting technologies.

l_m	Existing lighting	N_{lm}	Alternatives	C_{lm}^{ligm}($)
l_1	2-lamp 4′ T8 fixture 70 W	80		
			2-lamp 4′ T5 14 W	19.0
			2-lamp 4′ T5 18 W	20.5
			2-lamp 4′ T5 36 W	10.0
l_2	PAR 38–65 W	48		
			CFL lamp 7 W	35.4
			CFL lamp 14 W	37.1
			CFL lamp 20 W	27.6
l_3	Halogen 50 W − 12 V	56		
			LED flood 7 W	8.5
			LED flood 10 W	12.2
			LED flood 14 W	17.7
l_4	Incandescent 100 W	32		
			LED bulb 12 W	79.5
			LED bulb 17 W	53.0
			LED bulb 20 W	42.4
l_5	Incandescent 60 W	68		
			LED bulb 12 W	79.5
			LED bulb 17 W	53.0
			LED bulb 20 W	42.4

Table 9
The optimal solution.

	Year 1	Year 2	Year 3	Year 4
$\beta(t)$ ($)	2000	7000	30000	70000
$C(t)$ ($)	1425	6991	18959	69742
Window	0	0	0	0
Wall	0	10	0	0
Roof	0	0	0	2
Chiller	0	0	1	0
Heat pump	0	0	1	0
PV	0	0	0	5
N_{pv}	0	0	0	97
L_1	1	3	0	0
N_{lig_1}	75	5	0	0
L_2	0	2	0	0
N_{lig_2}	0	48	0	0
L_3	0	3	0	0
N_{lig_3}	0	56	0	0
L_4	0	2	0	0
N_{lig_4}	0	32	0	0
L_5	0	2	3	0
N_{lig_5}	0	10	58	0
$t_p(t)$ (month)	8	20	44	90
$ES(t)$ (kWh)	12096	34433	58111	93852
$ES_p(t)$	10%	29%	48%	78%
$E_p(t)$	1.01	0.80	0.58	0.25

Table 10
RSD of investment's indicators.

	Payback period	Energy saving	NPV	Energy intensity
RSD	1.72%	1.19%	1.88%	5.12%

building retrofit are $2000, $7000, $30000 and $70000, respectively. The results obtained by the optimization procedure are given in Table 9. In Table 9, the numbers shown in the last four columns from the fourth row onward indicate the retrofit decision on the corresponding items listed in the first column. L_1, L_2, L_3, L_4 and L_5 represent the five existing lighting technologies. N_{pv}, N_{lig_1}, N_{lig_2}, N_{lig_3}, N_{lig_4} and N_{lig_5}, represent the numbers of installed solar panels and the numbers of lamps replaced. For example, the number '10' in the fifth row of the third column means that the tenth alternative of the wall insulation materials will be applied to the walls of the office building under study in the second year. The '1'

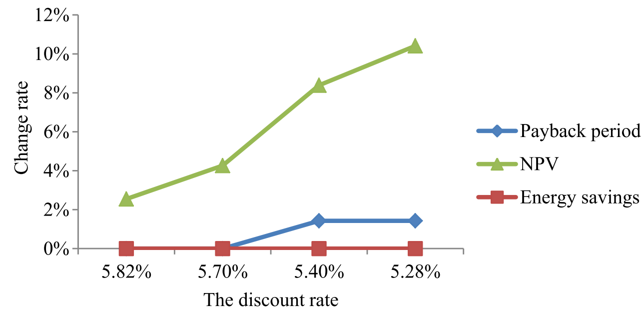


Fig. 2. Sensitivity analysis of the discount rate.

for L_1 and '75' for N_{lig_1} in the year one means that 75 of the first type of the existing lighting technologies will be replaced by its first alternative shown in Table 8. A '0' in the table indicates that the corresponding item will not be retrofitted in that year. In addition, Table 9 also shows the payback periods of the individual investments made at each year ($t_p(t)$). For instance, in year one, t_p is eight months, which corresponds to the payback period of the $1452 investment. The resulting absolute and percentage energy savings, $ES(t)$ and $ES_p(t)$, together with the energy intensity, $E_p(t)$, are also listed in the table.

The results obtained indicate that the lighting retrofit is the most cost-effective option followed by retrofit of HVAC facilities. Installation of PV system and retrofitting the envelope of the building require a long payback period. However, it can be concluded from Table 9 that the PV system can generate remarkable energy savings by comparing the values of ES_p in years 3 and 4, which positively contributes to the sustainability and environmental friendliness of the building.

Therefore, the optimization chooses the best combination of retrofit actions for the optimal plan. Table 9 shows that only the first lighting technology is retrofitted in the first year to achieve the desired EPC rating 'D'. Most of the lightings are replaced and the insulation is installed for the walls in the second year. It is noticed that not all of the last lighting technologies are retrofitted in the second year because of target grade 'C' requiring more energy savings, which is satisfied by the wall insulation. To reach grade 'B' rating in year three, the remaining quantities of the fifth lighting technology is retrofitted and the HVAC facilities are upgraded. The roof and solar panels are lastly considered in the forth year.

Intuitively, the payback period of the lighting system is the shortest while that of the envelope is the longest. Without help of the proposed optimization model, the decision maker is limited to this intuition and can only make retrofit plans accordingly, which, as demonstrated by the optimization result, will result in non-optimal retrofit activities. This demonstrates the effectiveness as well as the importance of the proposed optimal retrofit plan model.

The cumulative energy savings and net present value over ten-year period, and payback period of the total investment are given in Fig. 3. It is shown that the optimal retrofit plan results in 761.6 MWh energy savings, a net present value of $81003 with a payback period of 70 months.

Since GA is adopted to solve the optimization problem formulated, a statistical analysis of the results obtained is done through 20 run of the simulations. The relative standard deviations of the cumulative energy savings, net present value and the payback period of the building retrofit project and the energy intensity of the building are presented in Table 10, which are 1.19%, 1.88%, 1.72% and 5.12%, respectively. The results verify the effectiveness and convergence of the solution obtained by the GA algorithm.

As the parameters considered in the optimization process influence the optimal results, this study analyzes the effects of the discount rate, weighting factors and tax incentive on the proposed model.

Firstly, the discount rates with values of 5.82%, 5.70%, 5.40% and 5.28% are introduced. The resulting changes in the investment

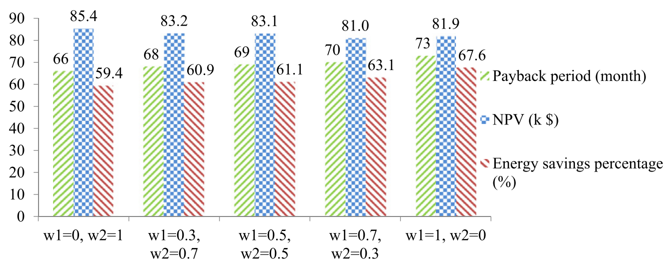


Fig. 3. Optimal retrofit plans with different weighting factors.

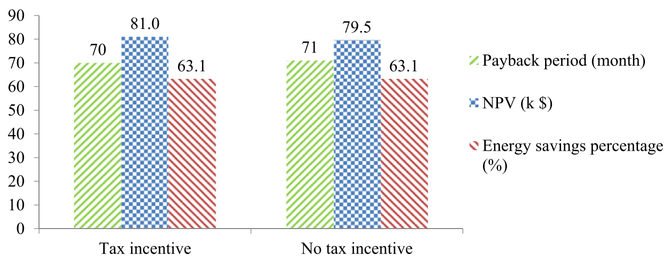


Fig. 4. Comparison between retrofit plans with and without tax allowance, $w_1 = 0.7.w_2 = 0.3$

indicators of applying the new discount rates are detailed in Fig. 2. To be specific, the optimal solution to the whole-building retrofit problem remained the same, thus leading to no change in the energy savings obtained. However, the payback period and NPV of the project change when the discount rate varies. It can be concluded from Fig. 2 that the energy savings are robust against the uncertainty on the discount rate while the economic factors are sensitive to its change. For instance, the NPV grows by 10.41% and the payback period decreases by 1.43% when the discount rate decreases to 5.28%."

Four more sets of results with the weighting factors in the objective function (25) set to different values are presented in Fig. 2. It can be seen that a shorter payback period and more energy savings can be achieved when the values of their corresponding weighting factors grow. For instance, the payback period of the project increases by 2.9% (from 68 to 70 months) and the percentage of energy savings increases by 3.6% (from 60.9% to 63.1%) when the values of the weighting factor change from $w1 = 0.3$ and $w2 = 0.7$ to $w1 = 0.7$ and $w2 = 0.3$. Comparing the five sets of results with different weighting factors, one can find that the shortest payback period of the retrofit can be obtained when the decision makers emphasizes minimization of the payback period with $w1 = 0$ and $w2 = 1$ and the most energy savings can be achieved when emphasis is put on the energy savings with $w1 = 1$ and $w2 = 0$.

Lastly, the optimization problem with $w_1 = 0.7$ and $w_2 = 0.3$ is solved again without taking into account the tax incentive program in view that some of the government owned buildings do not qualify for tax allowance. The ten-year energy savings and economic indicators obtained are shown in Fig. 4. It can be seen that the payback period is longer and the net present value is less than the case when the tax incentive is considered (see Fig. 3). However, it is seen that the tax incentive program has very limited impact on the building energy efficiency retrofit. To be exact, the payback period increased marginally by one month (1.4%) and the net present value decreased slightly by 1.5 thousand dollars (1.9%).

5. Conclusion

The focus of this paper is to develop a systematic optimization method for whole-building retrofit planning, aiming at reducing the energy consumption of existing buildings for green building policy compliance in a cost-effective manner. The main conclusions are given as follows:

- The model developed is able to identify the best retrofit plans for whole-building retrofit projects, taking into account both the envelope components and the indoor appliances. In this study, the retrofit actions considered include upgrade of lighting systems, HVAC facilities, installation of insulation materials to the walls and roof of the building, replacement of windows by more energy-efficient alternatives and installation of a roof top solar power system.
- The optimal retrofit plans obtained by the model can help target buildings to achieve a desired energy rating from the energy performance certificate (EPC) standard set by the South Africa government in the most profitable way.
- The proposed model is capable of breaking down the long-term building retrofit project requiring substantial investment into smaller projects over multiple financial years. This helps decision makers to select the best retrofit activities on a yearly basis to ensure that the energy performance of the building is improved and complies with the green building policy. In such a way, the most energy savings is obtained with a reasonable payback period of the investment.
- The tax incentive program available in South Africa is taken into account by the retrofit planning model to further shorten the payback period of the investment. It is however found that the tax incentive program has little impact on the building energy efficiency retrofit project.

The results of a case study show that 761.6 MWh energy savings and $81003 cost savings can be achieved in 70 months after applying the optimal retrofit plan, which validate the effectiveness and the importance of the model for decision makers because intuitive plans will lead to non-optimal retrofit actions.

References

[1] J. Woo, C. Menassa, Virtual retrofit model for aging commercial buildings in a smart grid environment, Energy Build. 80 (2014) 424–435.
[2] A. Jafari, V. Valentin, An optimization framework for building energy retrofits decision-making, Build. Environ. 115 (2017) 118–129.
[3] P. Cooper, D. Daly, Z. Ma, L. Ledo, Existing building retrofits: methodology and state-of-the-art, Energy Build. 55 (2012) 889–902.
[4] M. Almeida, M. Ferreira, Cost effective energy and carbon emissions optimization in building renovation (annex 56), Energy Build. 152 (2017) 718–738.
[5] Energy Performance Certificates for Buildings, 1 ed., South African Bureau of Standards, 2014 SANS 1544.
[6] The Application of the National Building Regulations Part XA: Energy Usage in Buildings, 1 ed., South African Bureau of Standards, 2011 SANS 10400-XA.
[7] X. Xia, J. Zhang, Energy efficiency and control systems–from a poet perspective, Conference on Control Methodologies and Technology for Energy Efficiency vol. 43, (2010), pp. 255–260 Vilamoura, Portugal.
[8] X. Xia, L. Zhang, Industrial energy systems in view of energy efficiency and operation control, Annu. Rev. Contr. 42 (2016) 299–308.
[9] X. Xia, Control problems in building energy retrofit and maintenance planning, Annu. Rev. Contr. 44 (2017) 78–88.
[10] H. Tazvinga, B. Zhu, X. Xia, Energy dispatch strategy for a photo-voltaicwinddieselbattery hybrid power system, Sol. Energy 108 (2014) 412–420.
[11] Z. Wu, H. Tazvinga, X. Xia, Demand side management of photovoltaic-battery hybrid system, Appl. Energy 148 (2015) 294–304.
[12] S. Sichilalu, H. Tazvinga, X. Xia, Optimal control of a fuel cell/wind/pv/grid hybrid system with thermal heat pump load, Sol. Energy 135 (2016) 59–69.
[13] A. Papadopoulos, State of the art in thermal insulation materials and aims for future developments, Energy Build. 37 (1) (2005) 77–86.
[14] F. Kuznik, D. David, K. Johannes, J.J. Roux, A review on phase change materials integrated in building walls, Renew. Sustain. Energy Rev. 15 (1) (2011) 379–391.
[15] P.L. Hurtado, A. Rouilly, V. Vandenbossche, C. Raynaud, A review on the properties of cellulose fibre insulation, Build. Environ. 96 (2016) 170–177.
[16] K. Chua, S. Chou, W. Yang, J. Yan, Achieving better energy-efficient air conditioning ? a review of technologies and strategies, Appl. Energy 104 (2013) 87–104.
[17] Y. Fan, X. Xia, An optimal maintenance plan for building envelope insulation materials after retrofitting, 2015 Chinese Automation Congress. Wuhan, China, 2015.
[18] B. Wang, X. Xia, A multistate-based control system approach toward optimal maintenance planning, IEEE Trans. Contr. Syst. Technol. 25 (1) (2017) 374–381.
[19] B. Wang, X. Xia, Optimal maintenance planning for building energy efficiency retrofitting from optimization and control system perspectives, Energy Build. 96 (2015) 299–308.
[20] X. Ye, X. Xia, L. Zhang, B. Zhu, Optimal maintenance planning for sustainable

energy efficiency lighting retrofit projects by a control system approach, Contr. Eng. Pract. 37 (0) (2015) 1–10.
[21] D. Setlhaolo, X. Xia, Combined residential demand side management strategies with coordination and economic analysis, Int. J. Electr. Power Energy Syst. 79 (2016) 150–160.
[22] D. Setlhaolo, X. Xia, Optimal scheduling of household appliances with a battery storage system and coordination, Energy Build. 94 (2015) 61–70.
[23] C. Thormark, A low energy building in a life cycleits embodied energy, energy need for operation and recycling potential, Build. Environ. 37 (4) (2002) 429–435.
[24] I.Z. Bribin, A.V. Capilla, A.A. Usn, Life cycle assessment of building materials: comparative analysis of energy and environmental impacts and evaluation of the eco-efficiency improvement potential, Build. Environ. 46 (5) (2011) 1133–1140.
[25] D. Chwieduk, Towards sustainable-energy buildings, Appl. Energy 76 (1) (2003) 211–217 Energex 2002-Energy Policies and Economics and Rational Use of Energy of Energy Topics VI and VII.
[26] E. Bertone, R.A. Stewart, O. Sahin, M. Alam, P.X. Zou, C. Buntine, et al., Guidelines, barriers and strategies for energy and water retrofits of public buildings, J. Clean. Prod. 174 (2018) 1064–1078.
[27] Z. Ma, P. Cooper, D. Daly, L. Ledo, Existing building retrofits: methodology and state-of-the-art, Energy Build. 55 (2012) 889–902.
[28] A. Jafari, V. Valentin, Selection of optimization objectives for decision-making in building energy retrofits, Build. Environ. 130 (2018) 94–103.
[29] F. Khayatian, L. Sarto, G. Dall'O', Building energy retrofit index for policy making and decision support at regional and national scales, Appl. Energy 206 (2017) 1062–1075.
[30] Y. Liu, T. Liu, S. Ye, Y. Liu, Cost-benefit analysis for energy efficiency retrofit of existing buildings: a case study in China, J. Clean. Prod. 177 (2018) 493–506.
[31] C. Rodrigues, R. Kirchain, F. Freire, J. Gregory, Streamlined environmental and cost life-cycle approach for building thermal retrofits: a case of residential buildings in south european climates, J. Clean. Prod. 172 (2018) 2625–2635.
[32] A.L. Webb, Energy retrofits in historic and traditional buildings: a review of problems and methods, Renew. Sustain. Energy Rev. 77 (2017) 748–759.
[33] B. Wang, Z. Wu, B. Zhu, X. Xia, Optimal control of maintenance instants and intensities in building energy efficiency retrofitting project, The 54th IEEE Conference on Decision and Control (CDC 2015). Osaka, Japan, 2015.
[34] M. Michael, L. Zhang, X. Xia, An optimal model for a building retrofit with LEED standard as reference protocol, Energy Build. 139 (2017) 22–30.
[35] E. Asadi, M.G. da Silva, C.H. Antunes, L. Dias, Multi-objective optimization for building retrofit strategies: a model and an application, Energy Build. 44 (2012) 81–87.
[36] Y. Fan, X. Xia, A multi-objective optimization model for building envelope retrofit planning, Energy Procedia 75 (2015) 1299–1304.
[37] Y. Fan, X. Xia, A multi-objective optimization model for energy-efficiency building envelope retrofitting plan with rooftop PV system installation and maintenance, Appl. Energy 189 (2017) 327–335.
[38] C. Rodrigues, F. Freire, Integrated life-cycle assessment and thermal dynamic simulation of alternative scenarios for the roof retrofit of a house, Build. Environ. 81 (2014) 204–215.
[39] Z. Wu, B. Wang, X. Xia, Large-scale building energy efficiency retrofit: concept, model and control, Energy 109 (2016) 456–465.
[40] ASHRAE handbook-fundamentals, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc, 2009.
[41] N. Wang, J. Zhang, X. Xia, Energy consumption of air conditioners at different temperature set points, Energy Build. 65 (2013) 412–418.
[42] Performance Rating of Unitary Air-conditioning & Air-source Heat Pump Equipment, Air Conditioning, Heating and Refrigeration Institute, 2008 ARI Standard 210/240.
[43] H. Tazvinga, B. Zhu, X. Xia, Optimal power flow management for distributed energy resources with batteries, Energy Convers. Manag. 102 (2015) 104–110.
[44] J. Hofierka, J. Kaňuk, Assessment of photovoltaic potential in urban areas using open-source solar radiation tools, Renew. Energy 34 (10) (2009) 2206–2214.
[45] M. Christersson, J. Vimpari, S. Junnila, Assessment of financial potential of real estate energy efficiency investments:a discounted cash flow approach, Sustain. Cities Soc. 18 (2015) 66–73.
[46] I. Kim, O. de Weck, Adaptive weighted sum method for multiobjective optimization: a new method for pareto front generation, Struct. Multidiscip. Optim. 31 (2) (2006) 105–116.
[47] R.T. Marler, J.S. Arora, The weighted sum method for multi-objective optimization: new insights, Struct. Multidiscip. Optim. 41 (6) (2010) 853–862.
[48] B. Surekha, L.K. Kaushik, A.K. Panduy, P.R. Vundavilli, M.B. Parappagoudar, Multi-objective optimization of green sand mould system using evolutionary algorithms, Int. J. Adv. Manuf. Technol. 58 (1) (2012) 9–17.
[49] Energy Efficiency in Buildings, 1 ed., South African Bureau of Standards, 2011 SANS 204.
[50] D. Whitley, A genetic algorithm tutorial, Stat. Comput. 4 (2) (1994) 65–85.
[51] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Trans. Evol. Comput. 6 (2) (2002) 182–197.

Applied Energy 221 (2018) 450–463

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

An autonomous hierarchical control for improving indoor comfort and energy efficiency of a direct expansion air conditioning system

Jun Mei[a,*], Xiaohua Xia[a], Mengjie Song[b]

[a] Centre of New Energy Systems, Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0028, South Africa
[b] Energy Research Institute at NTU (ERI@N), Research Techno Plaza, Nanyang Technological University, 50 Nanyang Drive, Singapore 637553, Singapore

HIGHLIGHTS

- We propose a novelty control strategy to save more energy consumed and cost.
- The results validate the proposed method for improving comfort levels.
- The proposed hierarchical control method is easy to implement in practice.
- Performance of designed control strategy is better than the previous strategies.
- The designed control method is not very sensitive to the system parameters.

ARTICLE INFO

Keywords:
Autonomous hierarchical control
PMV index
Model predictive control
Energy saving
time-of-use

ABSTRACT

This paper presents an autonomous hierarchical control method for a direct expansion air conditioning system. The control objective is to maintain both thermal comfort and indoor air quality at required levels while reducing energy consumption and cost. This control method consists of two layers. The upper layer is an open loop controller that allows obtaining tradeoff steady states by optimizing the energy cost of the direct expansion air conditioning system and the value of predicted mean vote under the time-of-use price structure of electricity. On the other hand, the lower layer designs a model predictive controller, which is in charge of tracking the tradeoff steady states calculated by the upper layer. Control performance of the proposed control method is compared to a conventional control strategy. The results show that the proposed control strategy reduces the energy consumption and energy cost of the direct expansion air conditioning system by 31.38% and 33.85%, respectively, while maintaining both the thermal comfort and indoor air quality within acceptable ranges, which validate the proposed methodology in terms of both comfort and energy efficiency.

1. Introduction

It is well known that the building sector is responsible for almost 40% of the global total energy consumption, costing $350 billion per year. Since energy management of building air conditioning (A/C) systems is a key factor in improving the energy efficiency and reducing the energy cost of buildings, optimal control of the A/C systems has increasingly attracted research attention. Energy efficiency improvement of buildings can also be performed at different levels of time scale and building subsystems such as ambient intelligence [1–3], energy balance [4–8], building portfolio management and planning [9–14] and energy-water nexus [15,16].

Since people spend much time indoors, thermal comfort and indoor air quality (IAQ) are important issues in A/C control. Thermal comfort has been accomplished by regulating temperature and relative humidity of indoor air. In view of air quality, CO_2 concentration is used as an indicator because carbon dioxide is the main fluid waste from occupants in a building. The indoor air temperature, humidity and CO_2 concentration are affected by A/C systems, lighting, the number of occupants and natural ventilation. They are also affected by outdoor environment, including the outside temperature, humidity, CO_2 concentration and solar irradiation. The A/C system needs to provide a comfortable environment for occupants with the minimum energy consumption and cost. There are strong interactions of energy cost and energy consumption with thermal comfort and IAQ. This crucial fact has been recognised by industrial and academic researchers.

Researchers proposed various control strategies to improve energy efficiency and comfort temperature [17–20]. In [21], the authors

* Corresponding author.
E-mail address: junmei027@gmail.com (J. Mei).

https://doi.org/10.1016/j.apenergy.2018.03.162
Received 29 January 2018; Received in revised form 15 March 2018; Accepted 30 March 2018

Nomenclature

A_1	heat transfer area of the DX evaporator in the dry-cooling region, m^2
A_2	heat transfer area of the DX evaporator in the wet-cooling region, m^2
A_0	total heat transfer area of the DX evaporator, m^2
A_{win}	total window area, m^2
C_c	CO_2 concentration of conditioning space, ppm
C_s	CO_2 concentration of supply air, ppm
C_z	specific heat of air, kJ kg^{-1} °C^{-1}
G	amount of CO_2 emission rate of people, m^3/s
h_{fg}	latent heat of vaporization of water, kJ/kg
h_{r1}	enthalpy of refrigerant at evaporator inlet, kJ/kg
h_{r2}	enthalpy of refrigerant at evaporator outlet, kJ/kg
k_{spl}	coefficient of supply fan heat gain, kJ/m^3
k_P,k_I	proportional and integral gains of PI controller
m_r	mass flow rate of refrigerant, kg/s
M_{load}	moisture load of conditioned space, kg/s
$Occp$	number of occupants
Q_{load}	sensible heat load of conditioned space, kW
Q_{rad}	solar radiative heat flux density, W/m^2
Q_{spl}	heat gain of supply fan, kW
T_d	air temperature leaving the dry-cooling region on air side, °C
T_s	temperature of supply air from the DX evaporator, °C
T_w	temperature of the DX evaporator wall, °C
T_z	air temperature of conditioned space, °C
T_0	temperature of outside, °C
v_a	air face velocity for DX cooling coil, m/s
v_f	air volumetric flow rate, m^3/s
V	volume of conditioned space, m^3
V_{h1}	air side volume of the DX evaporator in the dry-cooling region on air side, m^3
V_{h2}	air side volume of the DX evaporator in the wet-cooling region on air side, m^3
W_s	moisture content of supply air from the DX evaporator, kg/kg dry air
W_z	air moisture content of conditioned space, kg/kg dry air
W_0	air moisture content of outside, kg/kg dry air
α_1	heat transfer coefficient between air and the DX evaporator wall in the dry-cooling region, kW m^{-2} °C^{-1}
α_2	heat transfer coefficient between air and the DX evaporator wall in the wet-cooling region, kW m^{-2} °C^{-1}
ε_{win}	transmissivity of glass of window
ρ	density of moist air, kg/m^3

proposed an optimization method on room air temperature to improve both thermal comfort and energy efficiency. In [22], Cigler et al. presented an MPC to minimize the energy consumption and the value of predicted mean vote (PMV) index simultaneously. The simulation results showed that it would save 10–15% energy while keeping the comfort temperature within a level defined by standards. A hierarchical control method was proposed to improve the energy efficiency while maintaining the indoor temperature equal to a value such that the PMV index will be equal to zero reported in [23]. The results showed that it would reduce more energy consumption in comparison with previous work [24]. An economic model predictive control (MPC) method for optimising the building demand and energy cost under a TOU price policy under given bounded comfort temperature is studied in [25]. It demonstrated that this strategy is capable of reducing more energy cost and shifting the peak demand to off-peak hours while keeping the temperature at comfort bounded. In [26,27], the authors presented an MPC that minimises the expected energy cost and bounds of temperature comfort violations. One can note that all the above contributions focus on improving the energy efficiency of buildings by heating, ventilation and air conditioning (HVAC) temperature control. However, ensuring the indoor humidity at an appropriate level is also a crucial problem since it directly affects building occupants' thermal comfort and the operating efficiency of building A/C installations [28]. In fact, in cities with high humid climates, such as Cape Town or Hongkong, high humidity may still adversely impact indoor thermal comfort level and energy efficiency of building A/C systems even when indoor air temperature has been maintained at a desired value.

In recent years, a model-based predictive control algorithm proposed for HVAC system to control indoor temperature and humidity simultaneously taking into account energy efficiency was reported in [29]. In the study, the indoor air temperature and humidity are considered in two separate control loops. However, the control method remained inadequate fundamentally. A multi-input-multi-output (MIMO) control strategy is proposed for controlling the indoor air temperature and humidity simultaneously by varying the speeds of the compressor and the supply fan in an experimental direct expansion (DX) A/C system in [30]. In the research, the authors considered the coupling effect between indoor air temperature and humidity; so that the control accuracy and sensitivity can be improved. However, the control strategy was carried out based on the linearized system around a particular operational point, i.e., fixing the supply air temperature and moisture content. For a DX A/C system, its inlet air temperature and humidity affect its output cooling capacity directly [31]. The development of a physical model-based controller for a variable speed DX A/C system, aiming at controlling indoor air temperature and humidity simultaneously should be within its entire possible working range. An artificial neural network (ANN)-based modeling and control for an experimental variable speed DX A/C system was proposed to control the indoor air temperature and humidity simultaneously [32]. A real-time neural inverse optimal control for the simultaneous control of indoor air temperature and humidity using a DX A/C system was reported in [33]. A three-evaporator air conditioning system for simultaneous indoor air temperature and humidity control was studied in [34]. In [35], a fuzzy logic controller was developed for temperature and humidity control. The results demonstrated that the fuzzy logic controller developed can achieve the simultaneous control over indoor air temperature and humidity, with a reasonable control accuracy and sensitivity.

Nowadays, the indoor air quality (IAQ) is also an important issue for users, especially in office buildings, since a poor IAQ has a direct effect on work efficiency. In [36,37], Zhu et. al., studied indoor air temperature, humidity and CO_2 concentration control simultaneously without considering their coupling effects. However, these coupling effects cannot be ignored in many cases. In fact, the experimental investigation [38] suggested that the indoor CO_2 concentration affected indoor air temperature. Furthermore, indoor humidity was correlated with CO_2 concentration according to measurement results reported in [39]. Indoor air temperature, relative humidity and CO_2 levels assessment in academic buildings with different HVAC systems was studied in [40]. In [41], this study aimed to establish an optimal occupant behavior that can reduce total energy consumption and improve the thermal comfort, IAQ and visual comfort simultaneously by an energy simulation and optimization tool. In [42], an energy-optimised open loop controller and a closed-loop regulation of the multi-input-multi-output (MIMO) MPC schemes for a DX A/C system were proposed to improve both thermal comfort and IAQ, while minimizing energy consumption. The results showed that the energy savings were achieved and thermal comfort and IAQ were improved. However, the setpoints of thermostats are constant over a 24-h period. This strategy is simple but not optimal in the sense of energy efficiency or cost-effectiveness. On the other hand, the outside temperature and humidity are also constant over a 24-h period in the study, while the outdoor air temperature and humidity vary over a 24-h period actually. Besides, a ventilation fan with an independent pressure swing absorption box was added to improve IAQ, which would increase the complexity and the cost of hardware.

Reduction of energy consumption and cost is important to promote economic and environmental development. Therefore, it is of great interest to develop advanced control technologies for building A/C systems to reduce energy consumption and cost. However, several control methods were proposed recently to reduce energy consumption and cost of building A/C systems while maintaining thermal comfort and IAQ at required levels. In this paper, an autonomous hierarchical control method is proposed to ensure occupants' thermal comfort and IAQ in a certain environment, and at the same time, tries to reduce the energy consumption and cost for a DX A/C system. The use of the DX A/C system has many advantages. When compared to central chilled water-based A/C systems, DX A/C systems are simpler in system configuration, more energy efficient [43] and cost less to own and maintain. Therefore, DX A/C systems have been widely used over recent decades in buildings, especially in small to medium scaled buildings. The proposed control strategy will further enhance the performance of DX A/C system. The proposed autonomous hierarchical controller is formed by two layers. (i) The upper layer consists of a nonlinear optimizer, which provides trajectory references of indoor air temperature, humidity and CO_2 concentration within acceptable ranges. This controller uses an open loop controller to optimise the energy cost of the DX A/C system and the value of the PMV index under a TOU price policy. (ii) Meanwhile, the lower layer contains a closed-loop MPC controller to track adaptively and automatically the trajectory references of indoor air temperature, humidity and CO_2 concentration calculated by the upper layer. To demonstrate the advantage of the proposed control method, we will compare the proposed control and a baseline control strategy.

The contributions of this paper are listed below. The references of indoor air temperature, humidity and CO_2 concentration are not needed. We present a method to autonomously and adaptively optimise and generate all steady states on required levels of thermal comfort and IAQ which could vary during the day. The volume of outside air entering into the system is fixed in [37,42]. In our study, the volumes of fresh air entering the DX A/C system are considered to vary with the environment over a 24-h period and are optimised by the proposed method. Moreover, a supply fan to drive the pressure swing absorption with a built-in PI controller is proposed to reduce indoor CO_2 concentration in this paper. Hence, it has the potential of reducing the complexity of computation and the cost of hardware. The PMV index is traditionally used as an indicator of indoor thermal comfort. In this study, it is used as an indicator of thermal comfort and that of IAQ when the indoor air CO_2 concentration is at its steady state.

The remainder of this paper includes five parts. The nonlinear reduced order dynamical system models, the energy consumption models of the DX A/C system and the indoor cooling load models are presented in Section 2. The proposed control method is presented in Section 3. Results are given in Section 4, and conclusions are drawn in Section 5.

2. System model

2.1. DX A/C system

A DX A/C system is mainly composed of two parts, which are the DX refrigeration plant (refrigerant side) and air-distribution sub-system (air side). Fig. 1 is the simplified schematic diagram of the DX A/C system. The DX refrigeration side mainly consists of the following components: a variable speed rotor compressor, an electronic expansion valve (EEV), a high-efficiency tube-louver-finned DX evaporator and an air-cooled tube-plant-finned condenser. The evaporator is placed inside the supply air duct on the air side to work as a DX air cooling coil which is located in the room. The air side includes an air-distribution ductwork with return air dampers, a variable speed centrifugal supply fan, a pressure swing absorption (PSA) box, a conditioned space and a damper position which is used to control the proportion of return air to outside air. The PSA box absorbs the CO_2 contaminant concentration to maintain IAQ. The allowed fresh air is also used to improve indoor fresh air ratio.

2.2. DX A/C models

The dynamic model of the DX A/C system is mainly derived from the principles of energy and mass balance. The model is highly nonlinear with respect to temperature, moisture content and CO_2

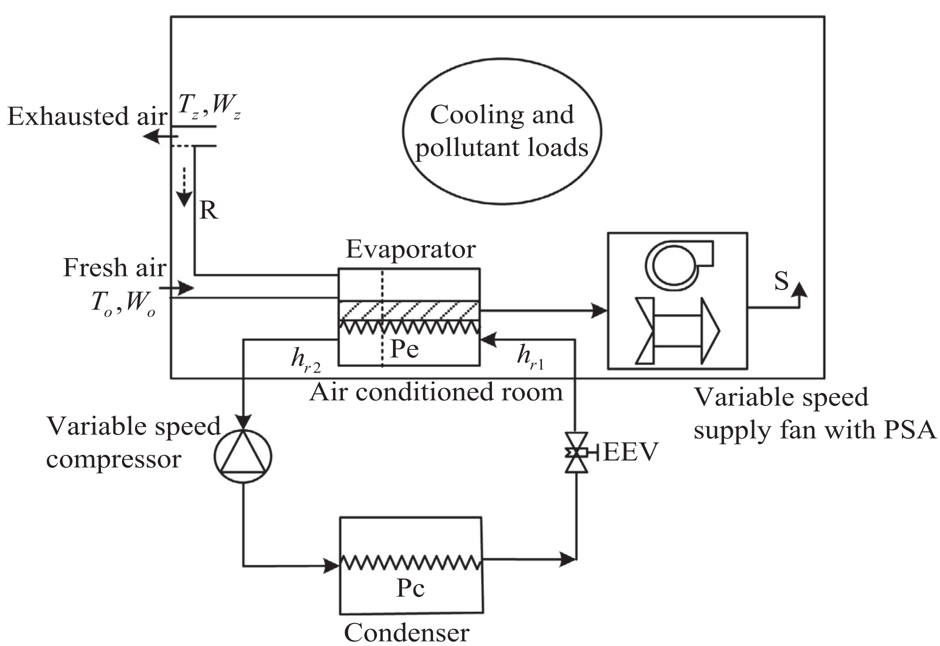


Fig. 1. Simplified diagram of DX air conditioning system.

concentration. In this paper, the system is assumed to operate in the cooling mode. The basic operation and assumptions of the system on the cooling mode are given for the purpose of simplicity as below: (i) It is assumed that p% of outside air enters into the system and gets mixed with $(100-p)\%$ of recirculated air entering into the system. (ii) Sufficient air mixing occurs inside the heat exchangers where the air gets conditioned. (iii) Two regions on the air side of the DX evaporator are shown in Fig. 2, i.e., dry-cooling region (sensible heat transfer only) and wet-cooling region (sensible and latent heat transfer region). The coupling between both regions that the outlet air properties of the first one (dry-cooling region) are the inlet air conditions entering the wet-cooling region. Therefore, the area from zero to A_1 pertains to the dry-cooling region and the rest of the total surface area is the wet-cooling region A_2. The boundary between the dry surface and the wet surface within a DX evaporator can be determined by the distribution of the surface temperature. It is a time-varying parameter under different conditions. (iv) Thermal losses in air ducts are neglected. (v) The supply air enters into the air-conditioned space to offset the cooling and pollutant loads acting upon the system. (vi) The air in the conditioned room exhausts through a fan, where $(100-p)\%$ of the air is recirculated and the rest is exhausted from the system through the fan.

Based on the above assumptions, the dynamic mathematical model for the DX A/C system for controlling indoor air temperature, moisture content and CO_2 concentration is developed based on the energy and mass conservation principles, which can be described by the following equations:

$$C_z\rho V\frac{dT_z}{dt} = C_z\rho v_f(T_s-T_z) + Q_{load}, \tag{1}$$

$$\rho V\frac{dW_z}{dt} = \rho v_f(W_s-W_z) + M_{load}, \tag{2}$$

$$C_z\rho V_{h1}\frac{dT_d}{dt} = C_z\rho v_f((1-p\%)T_z + p\%T_0-T_d) + \alpha_1A_1\left(T_w-\frac{(1-p\%)T_z + p\%T_0 + T_d}{2}\right), \tag{3}$$

$$C_z\rho V_{h2}\frac{dT_s}{dt} + \rho V_{h2}h_{fg}\frac{dW_s}{dt} = C_z\rho v_f(T_d-T_s) + h_{fg}\rho v_f((1-p\%)W_z + p\%W_0-W_s) + \alpha_2A_2\left(T_w-\frac{T_d + T_s}{2}\right), \tag{4}$$

$$C_w\rho_w V_w\frac{dT_w}{dt} = \alpha_1A_1\left(\frac{(1-p\%)T_z + p\%T_0 + T_d}{2}-T_w\right) + \alpha_2A_2\left(\frac{T_d + T_s}{2}-T_w\right) -(h_{r2}-h_{r1})m_r, \tag{5}$$

$$V\frac{dC_c}{dt} = v_s(C_s-C_c) + C_{load}. \tag{6}$$

More details for the system models (1)–(6) can be found in [42,44]. Note that the system models (1)–(5) without outside air entering into the system have been reported and validated by the experimental demonstrated in [44]. The model (6) has been verified in [45] by using an online learning and estimation approach for model parameter identification with acceptable accuracy.

We assume that the CO_2 concentration absorption rate v_s is a PI controller designed by

$$v_s = k_P v_f + k_I\int_0^{T_I} v_f\,ds. \tag{7}$$

The relationship among air enthalpy, temperature and the moisture content leaving the evaporator can be described by:

$$h_s = C_zT_s + h_{fg}W_s. \tag{8}$$

Then, Eqs. (2) and (4) can be rewritten by

$$\rho V\frac{dW_z}{dt} = \rho v_f\left(\frac{h_s-C_zT_s}{h_{fg}}-W_z\right) + M_{load}, \tag{9}$$

$$\rho V_{h2}\frac{dh_s}{dt} = C_z\rho v_f(T_d-T_s) + h_{fg}\rho v_f\left((1-p\%)W_z + p\%W_0-\frac{h_s-C_zT_s}{h_{fg}}\right) + \alpha_2A_2\left(T_w-\frac{T_d + T_s}{2}\right). \tag{10}$$

The air side convective heat transfer coefficients for the louver finned evaporator under both dry-cooling and wet-cooling regions are calculated as follows [46]:

$$\alpha_1 = j_{e1}\rho v_a\frac{C_z}{Pr^{\frac{2}{3}}},\quad \alpha_2 = j_{e2}\rho v_a\frac{C_z}{Pr^{\frac{2}{3}}}, \tag{11}$$

where Pr is Prandtl number, j_{e1} and j_{e2} are the Colburn factors. The air velocity v_a is described as follows:

$$v_f = dv_a + \varepsilon,$$

where $d\,(m^2)$ is the cross-sectional area of the conditioned space, ε is the error vector since the air enters or exits through the door or window.

The left-hand side of (1) and (2) is the heat flow into the conditioned space. On the right-hand side of (1), the first term denotes the heat transfer from the DX A/C system to the conditioned space, which is positive if $T_s > T_z$ for the heat mode and negative if $T_s < T_z$ for the cooling mode; the other terms mean the sensible heat load needs to be removed by the DX A/C system. Similarly, on the right-hand side of (2), the first term represents the wet-bulb temperature transferred to the conditioned space, which is positive if $W_s > W_z$ for the humidification mode and negative if $W_s < W_z$ for dehumidification mode; the second term denotes the moisture load to be removed by the DX A/C system. Eqs. (3), (5) and (10) mean that the heat transfer takes place in the inside DX A/C system. In Eq. (3), the first term of the right-hand side represents the heat transfer between the mixed air and the air side at the evaporator; the second term means the heat transfer between the mixed air and the evaporator wall. Eq. (6) represents a dynamic balance of indoor CO_2 concentration.

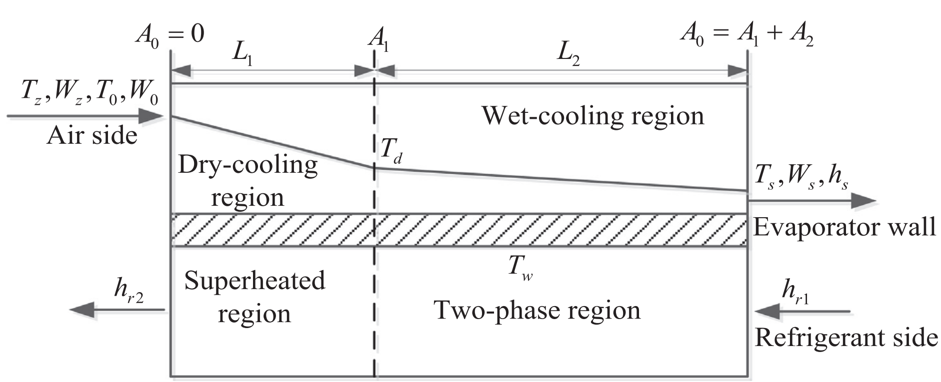


Fig. 2. Simplified diagram of DX evaporator [44].

Remark 1. In this paper, the relationship between the moisture content and temperature at the evaporator outlet [44], $W_s = \frac{0.0198T_s^2 + 0.085T_s + 4.4984}{1000}$, has been released since it may not be feasible under different operating conditions. The proportion of outside air entering into the system is not fixed according to the changing environment during the day. In our previous work [42], we used a variable air volume (VAV) ventilation fan with an independent PSA to reduce indoor CO_2 concentration. In this paper, we use the supply fan to drive the PSA with a built-in PI controller. This results in one less independent control input.

2.3. Load models

Thermal comfort and IAQ are influenced by a set of disturbances, such as external air, solar radiation through opaque and transparent surfaces and internal heat gains due to appliances, lights, occupants, etc. Therefore, good performance for controlling indoor air temperature, humidity and CO_2 concentration is required to deal with the disturbances. When the disturbances are neglected, a large error occurs. Nevertheless, a perfect prediction of disturbances in the future is inadequate in practice. Some disturbances can be measured, such as outside temperature, humidity and CO_2 concentration, and others, such as solar radiation and internal gains, cannot but may be estimated. Next, we will provide more details on the sensible heat load Q_{load}, moisture load M_{load}, pollutant load C_{load}.

The indoor sensible heat load is usually related to the internal loads, including occupants, lighting, equipment, fresh air entering inside and applications and the external loads, including heat transfer conduction through the building walls, roof, floor, doors and heat transfer by radiation through fenestration such as windows and skylights. In this paper, we consider the external load including heat loads by radiation through windows and the fresh air by ventilation. The moisture load is relevant to occupants, equipment, fresh air entering inside and applications. The CO_2 pollutant load is relevant to occupants' respiration. The sensible heat and moisture loads from lighting, equipment and applications are easy to identify, based on their electrical characteristics; the main uncertainties in identifying the sensible heat load and latent heat loads are from the loads associated with the occupants in the conditioned space. The sensible heat and moisture loads by occupants are determined through the current CO_2 emission. To estimate the sensible heat, moisture and indoor pollutant loads, a method is proposed as follows:

$$Q_{load}(t) = Q_{r,load} + Q_{spl} + \mu C_c + \nu + Q_{air}, \tag{12a}$$

$$M_{load}(t) = \phi C_c + \gamma + M_{air}, \tag{12b}$$

$$C_{load}(t) = G\cdot Occp, \tag{12c}$$

where μ and ϕ are the sensible heat and moisture gain coefficients, respectively, ν and γ are the certainties sensible heat and moisture loads, respectively. The heat gain of the supply fan Q_{spl} increases with the air volumetric flow rate of supply air as follows:

$$Q_{spl} = k_{spl} v_f. \tag{13}$$

The external heat load by radiation $Q_{r,load}$ through windows is described by the following equation:

$$Q_{r,load} = n_{win}\varepsilon_{win}A_{win}Q_{rad}, \tag{14}$$

where n_{win} denotes whether the conditioned space has a window, i.e., when $n_{win} = 1$, if it has a window, while if $n_{win} = 0$, it does not. The fresh air of the sensible heat load Q_{air} and the moisture load M_{air} in conditioned space are expressed as follows:

$$Q_{air} = p\% C_z \rho v_f (T_0 - T_z), \tag{15a}$$

$$M_{air} = p\% \rho v_f (W_0 - W_z). \tag{15b}$$

Remark 2. In this section, a simple method is given to estimate the indoor sensible heat and moisture loads, and CO_2 pollutant load. An alternative method to estimate cooling load has been reported in [26]. Besides, the weather forecast data from the weather station in Cape Town are qualified for this research, because: (1) the current weather station is precisely predicted and (2) the weather conditions and solar radiation in this area are relatively stable, indicating that the profiles of the predicted outside temperature, relative humidity and CO_2 concentration are representative.

2.4. PMV index

The PMV index is used as a human thermal comfort requirement indicator. This indicator was first proposed by Fanger [47] to predict the average vote of a large group of persons on the thermal sensation scale. This sensation is expressed by relating the integer range [−3, +3] to the qualitative words cold, cool, slightly cool, neutral, slightly warm, warm, and hot. PMV is defined by six variables, namely metabolic rate M (W/m^2), clothing insulating I_{cl} (m$^{2\circ}$C/W), air temperature T_z, air humidity H_z, air velocity v_a (m/s), and mean radiant temperature T_r. The PMV index can be described by the following equation [47]:

$$\begin{aligned} PMV = {} & (0.303e^{-0.036M} + 0.028)\{(M-W) - 3.05\times10^{-3}[5733 - 6.99(M-W) \\ & -P_a] - 0.42[(M-W) - 58.15] - 1.7\times10^{-5}M(5867 - P_a) \\ & -0.0014M(34 - T_z) - 3.96\times10^{-8}f_{cl}[(T_{cl}+273)^4 - (T_r+273)^4] \\ & -f_{cl}h_c\cdot(T_{cl} - T_z)\}, \end{aligned} \tag{16}$$

where W (W/m^2) is the external work; P_a is the partial water vapor pressure in Pascal. The surface temperature of clothing T_{cl} is given by:

$$\begin{aligned} T_{cl} = {} & 35.7 - 0.028(M-W) - I_{cl}\{3.96\times10^{-8}f_{cl}[(T_{cl}+273)^4 - (T_r+273)^4] \\ & + f_{cl}h_c(T_{cl} - T_z)\}, \end{aligned} \tag{17}$$

and the convective heat transfer coefficient h_c is defined as:

$$h_c = \begin{cases} h_c^*, & \text{if } h_c^* > 12.1\sqrt{v_a}, \\ 12.1\sqrt{v_a}, & \text{if } h_c^* < 12.1\sqrt{v_a}, \end{cases} \tag{18}$$

where $h_c^* = 2.38\cdot(T_{cl} - T_z)^{0.25} f_{cl}$ is the ratio of body surface area covered by clothes to the naked surface area, can be defined as:

$$f_{cl} = \begin{cases} 1.00 + 1.290 I_{cl} & \text{if } I_{cl} \leqslant 0.078, \\ 1.05 + 0.645 I_{cl} & \text{if } I_{cl} > 0.078. \end{cases} \tag{19}$$

The mean radiant temperature T_r is determined as [48]:

$$T_r = [(T_g + 273)^4 + \frac{1.10\times10^8 v_a^{0.6}}{\epsilon D^{0.4}}(T_g - T_z)]^{1/4} - 273, \tag{20}$$

where T_g is the globe temperature; D and ϵ are the globe diameter in meters and the globe emissivity coefficient, respectively. P_a is related to the relative humidity of the air H_z by means of Antoine's equation [49]:

$$P_a = 10H_z e^{(16.6536 - 4030.183/(T_z + 235))}, \tag{21}$$

where $H_z = 100W_z/A_{conv}$, A_{conv} is the unit transfer coefficient. The metabolic rate M is determined by [50]:

$$M = \lambda G,$$

where the coefficient λ is a constant. Then the metabolic rate M under a steady state of the indoor CO_2 concentration can be rewritten as follows:

$$M = \frac{\lambda}{Occp}\left(k_P v_f + k_I \int_0^{T_I} v_f\, ds\right)(C_c - C_s).$$

The PMV can be written as a function of the following variables:

$$\text{PMV} = g(T_z, W_z, C_c, v_f, T_r, I_{cl}, T_{cl}). \tag{22}$$

Remark 3. There are several existing metrics to measure human (dis) comfort, e.g., temperature constraint violations [26], comfort penalty [17], predicted percentage dissatisfied (PPD) [51], and PMV index [23,29]. The PMV index has been used as an indicator to maintain indoor comfort temperature [23] and to control indoor temperature and humidity [29]. In this paper, the Eq. (22) implies that the modified PMV index can estimate not only the indoor thermal comfort but also IAQ under a steady state of the indoor CO_2 concentration and keep them in a certain range. This is subjective and can be considered as perfect when PMV = 0.

2.5. Energy models for the DX A/C system

The DX A/C system components that consume energy include the power input of the evaporator fan, compressor fan, and condenser. The power to drive the dampers is assumed to be negligible. The total power consumption P_{tot} of the DX A/C system at time t then is calculated as [52]:

$$P_{tot} = P_e + P_c + P_f, \tag{23}$$

where the fan power input of the evaporator P_e, the fan power of the compressor P_f and the power input of the condenser P_c are given below:

$$P_e = a_0 + a_1 v_f, + a_2 v_f^2 + a_3 T_s + a_4 T_s^2 + a_5 Q_c + a_6 Q_c^2 + a_7 v_f T_s + a_8 v_f Q_c + a_9 T_s Q_c, \tag{24}$$

$$P_f = b_0 + b_1 T_d + b_2 T_s + b_3 T_d^2 + b_4 T_d T_s + b_5 T_s^2 + b_6 T_d^3 + b_7 T_d^2 T_s + b_8 T_d T_s^2 + b_9 T_s^3, \tag{25}$$

$$P_c(t) = c_0 + c_1 m_r + c_2 m_r^2, \tag{26}$$

where the coefficients $a_i (i = 0,1,...,9), b_i (i = 0,1,...,9), c_i (i = 0,1,2)$ are constant and can be determined by curve-fitting of experimental data in [52]. The indoor cooling load Q_c is the summation of the sensible and latent heat loads.

2.6. Constraints

The DX A/C system is subject to thermal comfort, IAQ and operational constraints defined as below.

(C1) PMV $\in [\underline{\text{PMV}}, \overline{\text{PMV}}]$. The limit of the PMV value means thermal comfort and IAQ are within the required levels for human.
(C2) $T_z \in [\underline{T}_z, \overline{T}_z], W_z \in [\underline{W}_z, \overline{W}_z], C_c \in [\underline{C}_c, \overline{C}_c]$. The indoor air temperature, moisture content and CO_2 concentration are within the required ranges for occupants in the conditioned space.
(C3) $T_s \in [\underline{T}_s, \overline{T}_s], W_s \in [\underline{W}_s, \overline{W}_s]$. The bounds of the supply air temperature and moisture content are constrained because of the physical characteristics of the coils and the air cooling coils inside the evaporator. Besides, the upper bounds $\overline{T}_s$ and $\overline{W}_s$ are less than T_z and W_z, respectively, since the DX A/C system is operating in the cooling mode. The bound of air enthalpy h_s satisfies: $h_s \in [C_z \underline{T}_s + h_{fg} \underline{W}_s, C_z \overline{T}_s + h_{fg} \overline{W}_s]$ due to (8).
(C4) $T_w \leqslant T_d$. The air temperature after the surface of the DX cooling coil cannot be warm.
(C5) $v_f \in [\underline{v}_f, \overline{v}_f], m_r \in [\underline{m}_r, \overline{m}_r]$. The upper bounds of the air volumetric flow rate $\overline{v}_f$ and mass flow rate of refrigerant $\overline{m}_r$ are limited by the physical characteristics of the DX A/C system. The lower bounds $\underline{v}_f > 0$ and $\underline{m}_r > 0$ match minimum operation and ventilation demands.
(C6) $p\% \in [\underline{p}\%, \overline{p}\%)$. The upper and lower bounds limit the ratio of the fresh air entering indoor.
(C7) $T_d \leqslant (1-p\%) T_z + p\% T_0, W_s \leqslant (1-p\%) W_z + p\% W_0$. The mixed temperature and moisture content between the fresh air and return air after the DX dry-cooling region and wet-cooling region can only decrease, respectively.

By collecting the system dynamic Eqs. (1), (3), (5), (6), (9) and (10), we reach the following:

$$\dot{x}(t) = f(x(t), u(t), w(t)), \tag{27}$$

where the *state vector* of the system is denoted by

$$x = [h_s, T_z, T_d, T_w, W_z, C_c]^T,$$

the *control vector* is denoted by

$$u = [v_f, m_r]^T,$$

the *load vector* is denoted by

$$w = [Q_{load}, M_{load}, C_{load}]^T,$$

the *output vector* is denoted by

$$y = [T_z, W_z, C_c]^T.$$

The constraints in (C1)-(C7) are compactly written as

$$x \in \mathbb{X},\ u \in \mathbb{U},\ PMV \in \mathbb{F},\ p \in \mathbb{P},\ T_s \in \mathbb{T}_s,\ W_s \in \mathbb{W}_s, \text{ and } h(x) \leqslant 0, \tag{28}$$

where $\mathbb{X}, \mathbb{U}, \mathbb{P}, \mathbb{T}_s$ and $\mathbb{W}_s$ are bounded sets, and $h(x)$ is a function of state variables.

2.7. TOU Strategy

In this paper, the energy charge is determined based on the TOU strategy. The TOU electricity tariff is a typical program of demand-side management, in which the electricity price changes over different periods based on the electricity supply cost; for example, a high price σ_h for peak periods $\mathscr{T}_h$, medium price σ_m for standard periods $\mathscr{T}_m$ and low price σ_l for off-peak periods $\mathscr{T}_l$. In this study, the daily TOU electricity price can be described as

$$\sigma(l) = \begin{cases} \sigma_h = 0.20538\text{"\$"/kW h},\ l \in \mathscr{T}_h, \\ \sigma_m = 0.05948\text{"\$"/kW h},\ l \in \mathscr{T}_m, \\ \sigma_l = 0.03558\text{"\$"/kW h},\ l \in \mathscr{T}_l, \end{cases} \tag{29}$$

where $\mathscr{T}_h = (8,11] \bigcup (19,21], \mathscr{T}_l = (0,7] \bigcup (23,24]$ and $\mathscr{T}_m = (7,8] \bigcup (11,19] \bigcup (21,23]$. "\$" is the United States dollar and time $\mathscr{T}$ is the whole period of the day with $l = 1,...,24$. Since there is a big difference in energy prices between the peak and off-peak hours, cost savings can be expected if significant amount of peak power consumption is shifted to off-peak hours. To minimize the energy cost, some previous optimization control strategies are reported in [18,25]. In this paper, we propose an alternative optimisation control scheme to minimize not only the energy cost but also the energy consumption.

3. Hierarchical control

Hierarchical control can be interpreted as an attempt to handle complex problems by decomposing them into smaller subproblems and reassembling their solutions in a hierarchical structure. The idea is to establish a hierarchical control structure composed of two layers. The two layers are adopted by using a control schedule, the simplified scheme of which is described in Fig. 3. The main principle of hierarchical control is as follows. At the upper layer, the objective is performed to compute the optimal conditions with respect to a performance index representing an economic and environmental criterion over a long-term scale horizon H_L with a sampling period T_L. At this stage, a detailed, a physical nonlinear model of the system although static is used. At the lower layer, a simple linear dynamic model is used to design an MPC controller, guaranteeing that the target values transmitted from the upper layer are obtained over a short time horizon $h_l = T_L$ with a smaller sampling period $t_l = T_L/n_l$. Fig. 3 implies that the upper layer sends information to the lower layer at the sampling instant $mT_L (m = 0,1,...,\infty)$; meanwhile, the lower layer receives the information as a task, and then completes the task within the sampling intervals

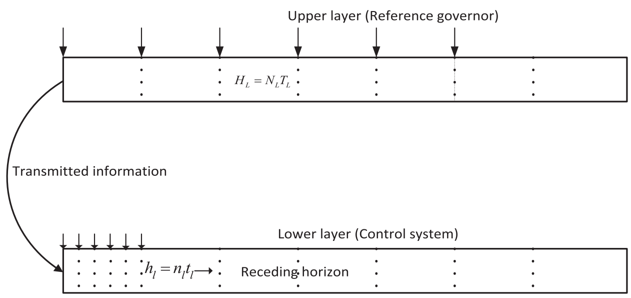


Fig. 3. Simplified schematic of two-layer hierarchical structure scheme.

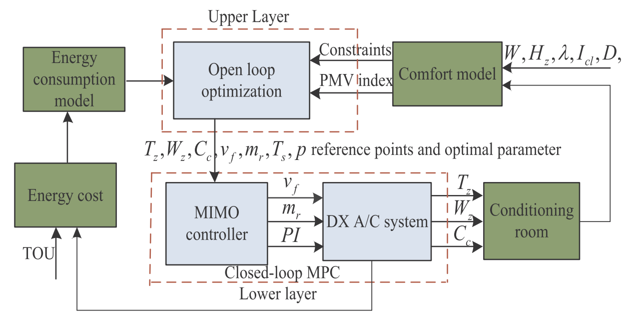


Fig. 4. Conceptual framework of the proposed hierarchical control approach.

$[mT_L + qt_l, mT_L + (q+1)t_l)(q = 0,1,\ldots,n_l-1)$.

This paper presents an autonomous hierarchical control approach to obtain a real-time optimisation scheduling strategy for the DX A/C system to minimise the total energy cost while maintaining the indoor thermal comfort and IAQ within acceptable ranges. The control method is based on a traditional control scheme with a reference governor in the upper layer, named the optimization layer, which, by means of a nonlinear optimizer, is able to generate the steady states and the optimal volume of air entering the system by optimising the energy cost of the DX A/C system and the value of the PMV index under the TOU strategy. Then, the lower layer receives the steady states as input, and the closed-loop MPC controller is designed to track the trajectory references of indoor air temperature, moisture content and CO_2 concentration. The conceptual framework of the proposed autonomous hierarchical control approach is shown in Fig. 4. The details are provided in the following subsections.

3.1. Optimization layer (Upper layer)

At the upper layer, the reference governor has been defined according to the optimization problem described by (30). Note that the PMV index (22) and the energy consumption model (23) are the optimization objectives. At the upper layer, the optimisation problem is considered as an open loop optimal control framework. Considering the DX A/C system (27) and its constraints (28), we formulate the following optimal controller to generate the steady states.

$$\min(\alpha|PMV(t_{m0})| + (1-\alpha)P_{tot}(t_{m0})\sigma(t_{m0})), \tag{30}$$

subject to the following constraints:

$$f(x(t_{m0}),u(t_{m0}),T_s(t_{m0}),p(t_{m0})) = 0, \tag{31}$$

$$x(t_{m0}) \in \mathbb{X},\ u(t_{m0}) \in \mathbb{U},\ PMV(t_{m0}) \in \mathbb{F},\ p(t_{m0}) \in \mathbb{P},\ T_s(t_{m0}) \in \mathbb{T}_s,\ W_s(t_{m0}) \in \mathbb{W}_s,\ h(x(t_{m0})) \leqslant 0, \tag{32}$$

where α is a weighting factor ($0 < \alpha < 1$), x,u and $f(\cdot)$ are denoted in (27). $x(t_{m0}),u(t_{m0}),v(t_{m0})$ are the optimization variables for $m = 0,\ldots,N_L-1$, where $v = [p,T_s,T_r,T_{cl}]$.

Assuming that all the variables are within the bounded sets, feasible solutions exist for the optimization problem (30) by using an open loop controller. Among all the feasible solutions, let $x_s(t_{m0}),u_s(t_{m0}),v_s(t_{m0})$ be the optimal solution of optimization problem (30), and $x_s(t_{m0}) \in \boldsymbol{X}_s, u_s(t_{m0}) \in \boldsymbol{U}_s, v_s(t_{m0}) \in \boldsymbol{V}_s$ for $m = 0,\ldots,N_L-1$. $\boldsymbol{X}_s,\boldsymbol{U}_s,\boldsymbol{V}_s$ are the optimal sequence points of the state, input and parameter variables. In this paper, the optimal sequence points are the steady states of Eq. (31).

Remark 4. The weighting factor α is chosen to balance the tradeoff between the two objectives, which are energy cost and comfort levels. Specifically, a relative large α gives better comfort level but worse cost savings. In the case that α is relatively large, more effort is put into optimizing the most comfortable indoor air temperature, humidity and CO_2, which may result in a loss of balancing capacity. The parameter α can be adjusted by utilities to achieve different goals.

The above nonlinear steady state optimization algorithm is provided as below.

Algorithm 1. Nonlinear Programming algorithm to the DX A/C system static optimization problem.

Initialization: Given initial state values $x(0)$ and $u(0)$. The initial state values are selected within their bounds.
1: Input the data of the outside temperature, relative humidity, sensible heat load, latent heat load and pollutant load.
2: The objective function (30) and constraints in (31) and (32) are converted into the following standard nonlinear programming so that it can be conveniently solved by the Matlab built-in function *fmincon*:

$$\min f_c^T \cdot z\ s.\ t.\ \begin{cases} c(z) \leqslant 0 \\ ceq(z) = 0 \\ A\cdot z \leqslant b \\ A_{eq}\cdot z = beq \\ lb \leqslant z \leqslant ub \end{cases} \tag{33}$$

3: Solve the above procedure (33).

3.2. Control layer (Lower layer)

As discussed above, for each every sample period T_L, the upper layer controller computes the optimal steady state point and delivers it into the lower layer. The task of the lower layer receives the steady state as the trajectory reference and includes a control algorithm trying to drive the system to track the trajectory reference. Therefore, in this case, this layer consists of a discrete-time MPC controller with a sampling time of $t_{mq} \in [mT_L + qt_l, MT_L + (q+1)t_l), m = 0,\ldots,N_L-1, q = 0,1,\ldots,t_l-1$, which is designed to track the reference point of indoor air temperature, moisture content and CO_2 concentration.

In the sequel, we make a commensurate quantization assumption: all variables are quantised in the two sampling schemes, i.e., they are represented by the starting values and remain these values in the same sampling interval, and the objective functions $PMV(t),P_{tot}(t)$, the TOU function $\sigma(t)$, and the constraints in (C1)-(C7) are coarsely quantised, i.e., they take their corresponding values at mT_L, for all $t \in [mT_L,(m+1)T_L)$. This assumption ensures that if the steady state $(x_s(t_{mq}),u_s(t_{mq}))$ would be obtained from the optimisation (30)–(32), then one would have $(x_s(t_{mq}),u_s(t_{mq})) = (x_s(t_{m0}),u_s(t_{m0}))$.

The lower layer receives the reference points of state vector and input vector, which are defined as $x_s(t_{mq}) \triangleq [h_{s,s}(t_{mq}),T_{z,s}(t_{mq}),T_{d,s}(t_{mq}),T_{w,s}(t_{mq}),W_{z,s}(t_{mq}),C_{c,s}(t_{mq})]^T$ and $u_s(t_{mq}) = [v_{f,s}(t_{mq}),m_{r,s}(t_{mq})]^T$. Define $\delta T_z(t_{mq}) = T_z(t_{mq})-T_{z,s}$

$(t_{m_q}),\delta W_z(t_{m_q}) = W_z(t_{m_q})-W_{z,s}(t_{m_q}),\delta C_c(t_{m_q}) = C_c(t_{m_q})-C_{c,s}(t_{m_q}),\delta h_s(t_{m_q}),$
$= h_s(t_{m_q})-h_{s,s}(t_{m_q}),\delta T_d(t_{m_q}) = T_d(t_{m_q})-T_{d,s}(t_{m_q}),\delta T_w(t_{m_q})$
$= T_w(t_{m_q})-T_{w,s}(t_{m_q}),\delta v_f(t_{m_q}) = v_f(t_{m_q})-v_{f,s}(t_{m_q}),\delta m_r(t_{m_q})$
$= m_r(t_{m_q})-m_{r,s}(t_{m_q})$
as the deviations of states and inputs from their trajectory references at sampling period $[mT_L + qt_l, mT_L + (q+1)t_l)$. Therefore, the dynamical mathematical equation of the DX A/C system at time t_{m_q} can be linearized and written in a linear state-space representation:

$$\begin{cases}\delta\dot{x}(t_{m_q}) = A_c(x_s(t_{m0}),u_s(t_{m0}))\delta x(t_{m_q}) + B_c(x_s(t_{m0}),u_s(t_{m0}))\delta u(t_{m_q}),\\ y(t_{m_q}) = C\delta x(t_{m_q}) + y_s(t_{m0}),\end{cases} \tag{34}$$

where the state variables $\delta x(t_{m_q}) = x(t_{m_q})-x_s(t_{m0}) = [\delta h_s(t_{m_q}),\delta T_z(t_{m_q}),\delta T_d(t_{m_q}),\delta T_w(t_{m_q}),\delta W_z(t_{m_q}),\delta C_c(t_{m_q})]^T$, the input variables $\delta u(t_{m_q}) = u(t_{m_q})-u_s(t_{m_q}) = [\delta v_f(t_{m_q}),\delta m_r(t_{m_q})]^T, y_s(t_{m0}) = [T_{z,s}(t_{m0}),W_z(t_{m0}),C_{c,s}(t_{m0})]^T$ and $y(t_{m_q}) = [T_z(t_{m_q}),W_{z,s}(t_{m_q}),C_c(t_{m_q})]^T$ are the original output variables.

$A(x_s(t_{m0}),u_s(t_{m0})),B(x_s(t_{m0}),u_s(t_{m0})),C$ are the system state matrix, input matrix and output matrix at the sampling time t_{m_q}, respectively, which can be calculated by:

$$A_c(x_s(t_{m0}),u_s(t_{m0})) = \frac{\partial f(x(t_{m0}),u(t_{m0}))}{\partial x(t_{m0})}\bigg|_{\substack{x(t_{m0})=x_s(t_{m0}),\\ u(t_{m0})=u_s(t_{m0})}}$$

$$B_c(x_s(t_{m0}),u_s(t_{m0})) = \frac{\partial f(x(t_{m0}),u(t_{m0}))}{\partial u(t_{m0})}\bigg|_{\substack{x(t_{m0})=x_s(t_{m0}),\\ u(t_{m0})=u_s(t_{m0})}}$$

and

$$C = \begin{bmatrix} 0&1&0&0&0&0\\ 0&0&0&0&1&0\\ 0&0&0&0&0&1\end{bmatrix}.$$

Consider the discrete-time version of (34):

$$\begin{cases}\delta x(t_{m_{q+1}}) = A_d(x_s(t_{m0}),u_s(t_{m0}))\delta x(t_{m_q}) + B_d(x_s(t_{m0}),u_s(t_{m0}))\delta u(t_{m_q}),\\ y(t_{m_q}) = C\delta x(t_{m_q}) + y_s(t_{m0}),\end{cases} \tag{35}$$

where $x(t_{m_q}),u(t_{m_q})$ and $y(t_{m_q})$ are the state vector, input vector and output vector at sampling instant $mT_L + qt_l, m = 0,1,\ldots,N_L-1, q = 0,1,\ldots,n_l-1$. $A_d(x_s(t_{m0}),u_s(t_{m0})) = e^{A_c(x_s(t_{m0}),u_s(t_{m0}))tl}, B_d(x_s(t_{m0}),u_s(t_{m0})) = (\int_0^{tl} e^{A_c(x_s(t_{m0}),u_s(t_{m0}))\tau}d\tau)B_c(x_s(t_{m0}),u_s(t_{m0}))$ are the system state matrix and input matrix, respectively.

The objective of the proposed MPC controller is to maintain the indoor air temperature, moisture content and CO_2 concentration at the required levels with low energy cost. To achieve this aim, the cost function to be minimised can be chosen as

$$\min_{\delta u} J(t_{m_q}) = \underbrace{\sum_{j=1}^{n_p} \|y(t_{m_{q+j}}|t_{m_q})-r(t_{m_{q+j}})\|^2}_{(a)} + \underbrace{R_{\delta u}\sum_{j=0}^{n_c-1} \|\delta u(t_{m_{q+j}})\|^2}_{(b)}, \tag{36}$$

subject to:

$$\begin{cases}\delta x(t_{m_{l_1}}|t_{m_{q+1}}) = A_d(x_s(t_{m0}),u_s(t_{m0}))\delta x(t_{m_{l_1-1}}|t_{m_q})\\ \qquad + B_d(x_s(t_{m0}),u_s(t_{m0}))\delta u(t_{m_{l_1-1}}|t_{m_q}),\\ y(t_{m_{l_1-1}}|t_{m_q}) = C\delta x(t_{m_{l_1-1}}|t_{m_q}) + y_s(t_{m0}),\end{cases} \tag{37}$$

$$\begin{aligned}&\delta x(t_{m_{l_1}}|t_{m_q}) + x_s(t_{m0}) \in \mathbb{X}, \delta u(t_{m_{l_2}}) + u_s(t_{m0}) \in \mathbb{U},\\ &l_1 = q+1,\ldots,q+n_p, l_2 = q,\ldots,q+n_c-1,\\ &q = 0,1,\ldots,n_l-1, m = 0,1,\ldots,N_L-1.\end{aligned} \tag{38}$$

where (a) penalizes the indoor air temperature, moisture content and CO_2 concentration tracking error and (b) penalizes the balancing signal tracking error in quadratic forms. The current time index t_{m_q} denotes the current time $mT_L + qt_l$; $|t_{m_q}$ means that the predicted value is based on the information up to $t = mT_L + qt_l$; $n_p = T_L/t_l$ is the prediction horizon; $n_c = T_L/t_l$ is the control horizon; $r(t_{m_{q+j}})$ is the reference vector at step $t_{m_{q+j}}$; $y(t_{m_{q+j}}|t_{m_q})$ is the predicted output vector at step $t_{m_{q+j}}$; $\delta u(t_{m_{q+j}})$ is the predicted control vector at step $t_{m_{q+j}}$; $R_{\delta u}$ is used as a tuning parameter for the desired closed-loop performance.

Remark 5. The system matrices of the system (34) are updated to $A_d(x_s(t_{(m+1)0}),u_s(t_{(m+1)0}))$ and $B_d(x_s(t_{(m+1)0}),u_s(t_{(m+1)0}))$ when the system transiting from the sampling interval $[mT_L + (n_l-1)t_l,(m+1)T_L)$ to $[(m+1)T_L,(m+1)T_L + t_l)$. On the other hand, on sampling interval $[(m+1)T_L,(m+1)T_L + t_l)$, the variables $\delta x(t_{m_{n_l-1}})$ and $\delta u(t_{m_{n_l-1}})$ as the initial points are fed to the system (37), and the references are updated in (36). The convergence for this periodic MPC for an optimisation problem over an infinite time horizon has been proven in [53,54].

The proposed MPC algorithm is as below:

Algorithm 2. MPC algorithm to the DX A/C tracking control problem.

Initialization: Given initial state value $x(0)$ and let $t_{m_q} = 0 (m = 0, q = 0)$.
1: Compute the optimal solution $\overline{U}(t_{m0}) = [\overline{u}(t_{m0}),\overline{u}(t_{m1}),\ldots,\overline{u}(t_{m_{n_l-1}})]^T$ of the problem formulated in (36) and (38).
2: Apply the MPC control $u_{mpc}(t_{m0}) = \overline{u}(t_{m0})$ to the system in the sampling interval $[t_{m0},t_{m0} + t_l)$; the rest of the solutions $\overline{u}(t_{m_q}), q = 1,\ldots,n_l-1$ are discarded. $x(t_{m_{q+1}})$ is calculated by $x(t_{m_{q+1}}) = f(x(t_{m_q}),u_{mpc}(t_{m_q}))$.
3: Set $t_{m_q} := t_{m_{q+1}}$, and update system states, inputs and outputs with control $u_{mpc}(t_{m0})$ and state equation $x(t_{m_{q+1}}) = f(x(t_{m_q}),u_{mpc}(t_{m_q}))$.
4: Until $t_{m_q} := t_{m_{n_l-1}}$, and update system states, inputs and outputs; repeat the steps 1 and 2, we have obtain that $u_{mpc}(t_{m_{n_l-1}}) = \overline{u}(t_{m_{n_l-1}})$. Apply the MPC control $u_{mpc}(t_{m_{n_l-1}})$ to the system in the sampling interval $[t_{m_{n_l-1}},t_{(m+1)0})$.
5: Set $t_{m_q} := t_{(m+1)0}$, measure the state value $x(t_{m_{n_l-1}})$ by the step $t_{m_q} = t_{m_{n_l-1}}$, and $u_{mpc}(t_{m_{n_l-1}})$ to the system $x(t_{(m+1)0}) = f(x(t_{m_{n_l-1}}),u_{mpc}(t_{m_{n_l-1}}))$, and update reference $r(t_{m0}) := r(t_{(m+1)0})$ in (36).
6: Compute the optimal solution $\overline{U}(t_{(m+1)0}) = [\overline{u}(t_{(m+1)0}),\overline{u}(t_{(m+1)1}),\ldots,\overline{u}(t_{(m+1)_{n_l-1}})]^T$ of the problem formulated in (36) and (38). Then the MPC control $u_{mpc}(t_{(m+1)0}) = \overline{u}(t_{(m+1)0})$ (the remaining $\overline{u}(t_{(m+1)_q}), q = 1,\ldots,n_l-1$ are discarded) is applied to the system in the sampling interval $[t_{(m+1)0},t_{(m+1)0} + t_l)$ to obtain the closed-loop MPC solution $x(t_{(m+1)1}) = f(x(t_{(m+1)0}),u_{mpc}(t_{(m+1)0}))$ over the period $[t_{(m+1)0} + t_l,t_{(m+1)0} + 2t_l)$.
7: Set $t_{m_q} := t_{(m+1)_1}$ and go to step 1.

Generally, the above MPC algorithm never stops, and it updates the controller at each time interval $[t_{m_q},t_{m_{q+1}})$ to include feedback information.

4. Results

Here, a case study is presented to demonstrate the performance of the closed-loop system with the proposed hierarchical control for the DX A/C system. The proposed hierarchical control strategy is compared with a baseline strategy through simulations over a 24-h period.

4.1. System setup

In the case study, an office room is taken as the conditioned space. The volume of the DX conditioned space is 77 m^3. The parameters of the DX A/C system are listed in Table 1. For the proposed hierarchical control strategy, the values of the system dynamic variable constraints are listed in Table 2, and we constrain the value of the PMV in the range of [−0.5,0.5] to ensure that the DX A/C system is able to control indoor thermal comfort and IAQ at acceptable levels. The coefficients of the energy consumption models (23) of the DX A/C system are taken from [52], which are summarized in Table 3.

In this paper, data of the outside temperature and relative humidity in a single summer are given, as shown in Fig. 5(a). The data is obtained from a meteorological station located in Cape Town, South Africa. The predicted solar radiative heat flux density profile of Cape Town is shown in Fig. 5(b). The certainty internal sensible and latent heat loads, the external sensible heat load and pollutant load in the conditioned space are predicted in Fig. 6. The values in Figs. 5 and 6 at every hour are commensurately quantised. It is assumed that the PI controller can absorb the CO_2 concentration in the air supply, where $C_s = 360$ ppm is used in this paper.

The TOU schedule for summer hours is summarized in (29); for simplicity, only the TOU energy charge is used in the cost function. The unit of Relative Humidity (RH) is percent (%). $\frac{11.35}{1000}$ kg/kg of moisture content is equivalent to 60% RH in the conditioned space. In addition, the original nonlinear system (27) is used as the system to be controlled in the simulation.

4.2. Two scheduling strategies

Here, we consider two strategies to schedule the operation of the DX A/C system in the conditioned space. One is energy optimised open loop controller and the closed-loop regulation of the MIMO MPC approach, which serves as a baseline strategy [42], and the other is the proposed energy and comfort optimised open loop controller and the closed-loop tracking of the MIMO MPC strategy. To simplify the comparison, the predicted load profiles are the same in both control strategies.

(1) Baseline: The baseline can be described as follows: We first select a setpoint for indoor air temperature and relative humidity based on a comfort zone within the psychrometric chart and a setpoint for CO_2 concentration based on the required level of occupants. The ASHRAE comfort zone is shown in [55]. Its details are omitted here because of space limitations. Under the given setpoint, we can obtain a unique steady state of the DX A/C system by solving the Eqs. (1), (3), (5), (6), (9) and (10) at every hour over a 24-h period. The nonlinear model is then linearised around its steady state. An MPC is designed for the linearised model. The proposed MPC with sampling period 2 min is applied to achieve better performance on thermal comfort and IAQ with superior energy efficiency simultaneously.

(2) Proposed method: For the proposed control strategy, the details are also given as below: We first consider the open loop controller to solve the optimization problem (30) to obtain steady states at every hour. The open loop controller and closed-loop MPC are employed to track the references of temperature, humidity and CO_2 concentration. In the proposed control method, the volume of the outside air entering indoor is optimized. The optimal volume of the outside air is used in the DX A/C system for closed-loop MPC controller. The sampling period is set to $T_L = 1$ h; the sampling interval is set to $N_L = 24$ hour; the sampling period for MPC design is $t_l = 2$ min, the prediction horizon and control horizon are taken as $n_p = n_c = 30$ in the lower layer. At each time step, the open loop controller is employed to solve the optimization problem (30) and the steady states obtained are sent to the lower layer. In Section 4.3, we will compare the energy consumption and energy cost for the baseline and the proposed strategies next.

Table 1
Parameters of system model.

Notations	Values	Notations	Values
ρ	1.2 kg/m^3	h_{fg}	2450 kJ/kg
V	77 m^3	ε_{win}	0.45
V_{h1}	0.04 m^3	V_{h2}	0.16 m^3
k_{spl}	0.0251 kJ/m^3	C_z	1.005 kJ kg^{-1} °C^{-1}
A_0	22.07m^2		

Table 2
Constraints of system variables.

Notations	Values	Notations	Values
$\overline{T}_s$	22 °C	$\underline{T}_s$	8 °C
$\overline{T}_z$	26 °C	$\underline{T}_z$	22 °C
$\overline{T}_d$	22 °C	$\underline{T}_d$	10 °C
$\overline{T}_w$	22 °C	$\underline{T}_w$	10 °C
$\overline{W}_z$	12.3/1000 kg/kg	$\underline{W}_z$	9.85/1000 kg/kg
$\overline{C}_c$	800 × 10^{-6} ppm	$\underline{C}_c$	650 × 10^{-6} ppm
$\overline{W}_s$	9.85/1000 kg/kg	$\underline{W}_s$	7.85/1000 kg/kg
$\overline{h}_s$	46.3 kJ/kg	$\underline{h}_s$	27.3 kJ/kg
$\overline{v}_f$	0.8 m^3/s	$\underline{v}_f$	0 m^3/s
$\overline{m}_r$	0.11 kg/s	$\underline{m}_r$	0 kg/s

Table 3
Coefficients of energy consumption models.

Notations	Values	Notations	Values
a_0	900.5	a_1	−8.1
a_2	6.18	a_3	−0.15
a_4	−4.61	a_5	0.02
a_6	−0.2	a_7	0.01
a_8	0.12	a_9	0.09
b_0	−6942	b_1	82
b_2	−0.7	b_3	2.4
b_4	−2.5	b_5	2.68
b_6	0.03	b_7	−0.02
b_8	0.04	b_9	0.0001
c_0	138.1	c_1	0.52
c_2	−2.3		

4.3. Comparison of two strategies

The performance of both strategies is compared with historical weather data of a specific day in Cape Town. The total simulation time is $K = 24$ h. The predicted indoor cooling loads profile is depicted in Fig. 7 overlaid with an electricity rate for summer hours. We duplicate the indoor cooling loads profile for the next day to simulate the MPC scheme. The temperature profile of the air leaving the DX evaporator and p% of the outside air entering into the system over a 24-h period are shown in Fig. 8(a) and (b). The data is used in the DX A/C system for closed-loop tracking control.

The controls computed from two strategies are applied to the DX A/C system. The tracking reference points of indoor air temperature in the conditioned space for the proposed strategy and the setpoint regulation of indoor air temperature for the baseline strategy are depicted in Fig. 9(a). The tracking reference points of indoor air relative humidity in the conditioned space for the proposed strategy and the setpoint regulation of indoor air relative humidity for the baseline strategy are depicted in Fig. 9(b). The tracking reference points of indoor CO_2 concentration in the conditioned space for the proposed strategy and the setpoint regulation of indoor CO_2 concentration for the baseline strategy are depicted in Fig. 9(c). We observe that the indoor temperature, humidity and CO_2 concentration for the proposed strategy can track their reference points well. We also observe that for the proposed

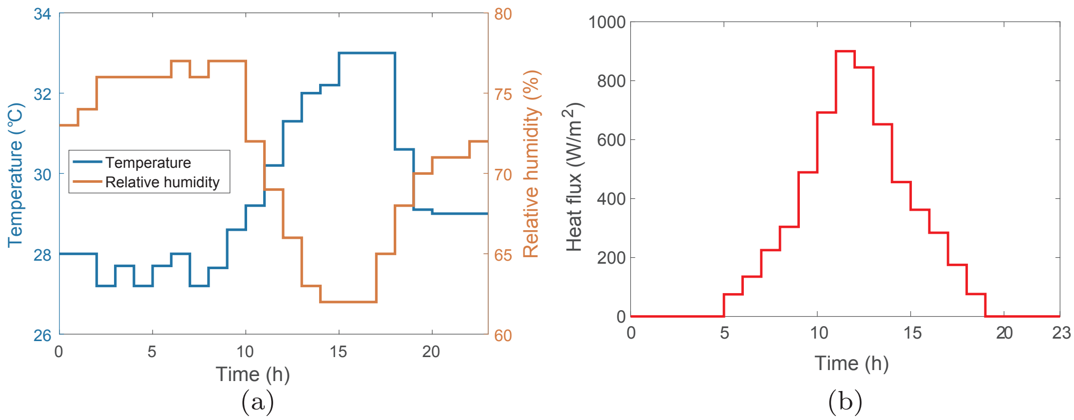


Fig. 5. (a) Profiles of outside temperature and relative humidity over a 24-h period. (b) Profiles of radiative heat flux over a 24-h period.

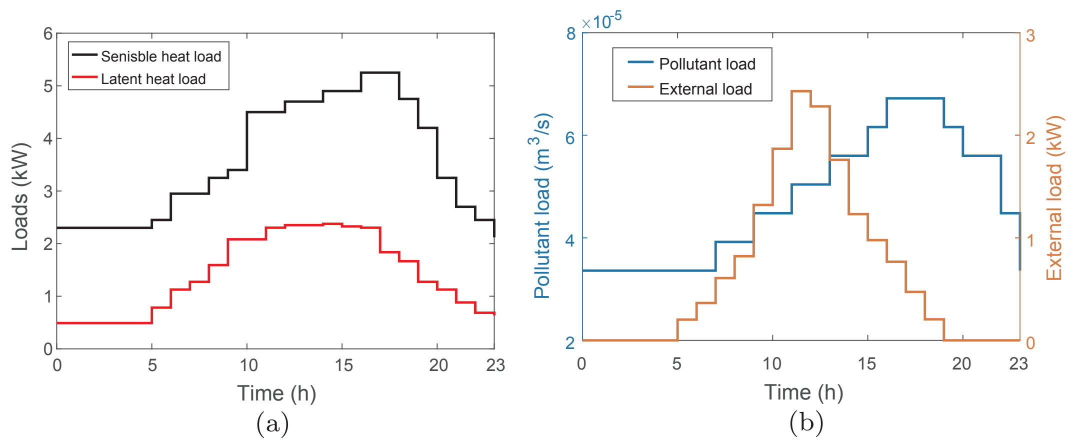


Fig. 6. (a) Certainty internal sensible and latent heat loads. (b) pollutant load and external sensible heat load.

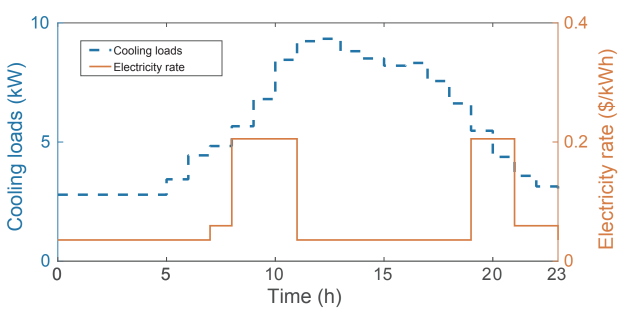


Fig. 7. Cooling loads and electricity rates over a 24-h period.

strategy the reference points are tallish during peak hours for temperature and humidity tracking. This is because the proposed controller can automatically adjust the reference points upward during peak hours such that the energy cost and energy consumption are minimized while both the thermal comfort and IAQ still maintain in the acceptable ranges. We further observe that with the baseline strategy under the varying loads, the MPC controller always maintains the indoor temperature, humidity and CO_2 concentration at their setpoint by regulating the control inputs. From the local zooming out of Fig. 9, the reference points of indoor air temperature, humidity and CO_2 concentration are reached after a transient process of 18 min. After reaching their reference points, the proposed controller maintains the

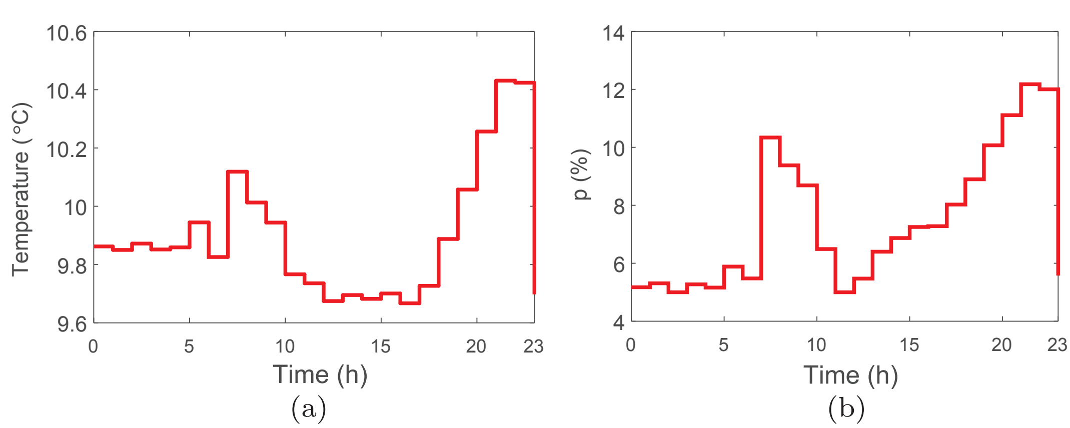


Fig. 8. (a) Supply air temperature over a 24-h period. (b) p% of fresh air entering into the system over a 24-h period.

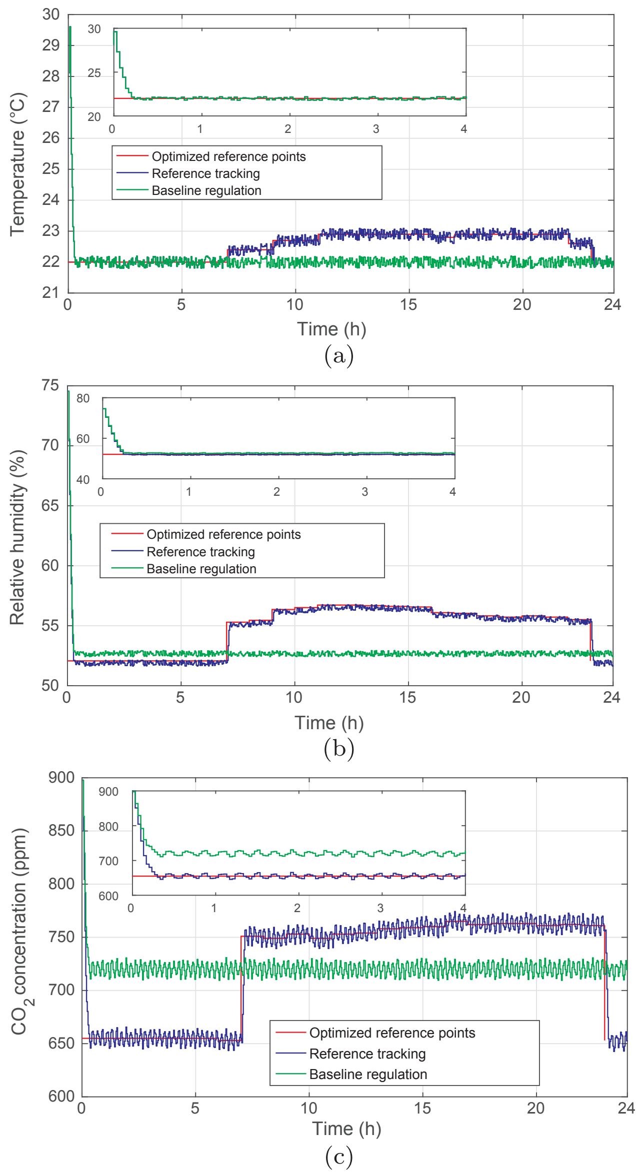


Fig. 9. (a) Temperature tracking. (b) Relative humidity tracking. (c) CO_2 concentration tracking.

reference points with small variation ranges. Fig. 10 shows the air volumetric flow rate and mass flow rate of refrigerant over a 24-h period. The two input variables vary to drive the indoor air temperature,

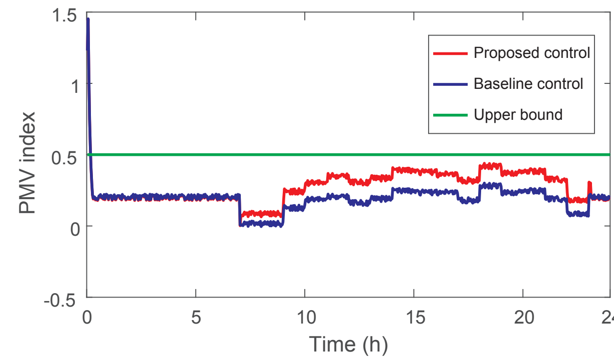


Fig. 11. Profile of the value of the PMV index over a 24-h period.

humidity and CO_2 concentration to track their trajectory references according to the changing environment during the day. In Fig. 11, it can be observed that the values of the PMV index for the two control methods lie within the expected range $[-0.5,0.5]$.

Fig. 12(a) and (b) illustrate the energy consumption and cost of the DX A/C system operation for the proposed strategy and the baseline strategy. We observe from Fig. 12(a) and (b) that both strategies consume almost the same energy cost from 0:00 to 7:00. The indoor temperature, humidity and CO_2 concentration reference points stay at the lower bound of the PMV index during off-peak hours without more cost. After 8:00, the energy costs of the baseline and proposed strategies start to increase since the increased cooling loads are required to be removed and the electric power price is increased. Compared to the baseline strategy, Fig. 12(a) shows that the proposed method consumes less energy. Comparing the two strategies, we observe that under the proposed method, more energy costs are reduced during peak hours. The reason is that the proposed method automatically adjusts upward the reference points such that the energy consumption and the energy costs are minimized during peak hours while maintaining both thermal comfort and IAQ at the required levels. From the simulation, it is verified that the major energy consumption and costs have been reduced effectively during peak hours. We summarize the total energy consumption, energy cost and comfort levels in Table 4. According to it, the proposed hierarchical control strategy performs better than the baseline by around 31.38% in terms of total energy consumption, and by around 33.85% in terms of total energy cost. It can be seen from Table 4 that the proposed control strategy presents a lower energy consumption and costs compared to the baseline control strategy. The table also shows that the total values of the PMV index for the proposed control strategy is higher than that of the baseline control strategy. It is expected that the proposed control strategy reduces energy consumption and cost at the expenses of the comfort level, which is still reasonably and optimally regulated to acceptable levels. Therefore, the utilities can choose

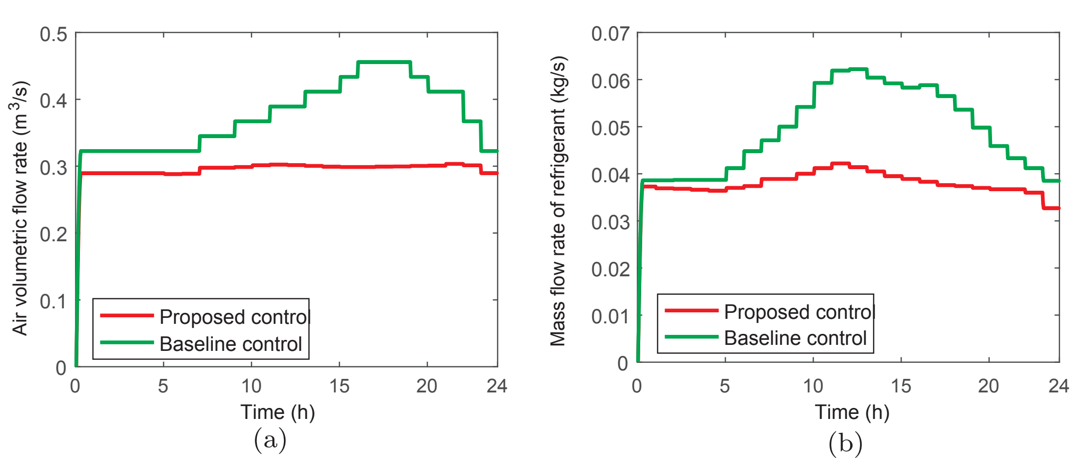


Fig. 10. (a) Air volumetric flow rate over a 24-h period. (b) Mass flow rate of refrigerant over a 24-h period.

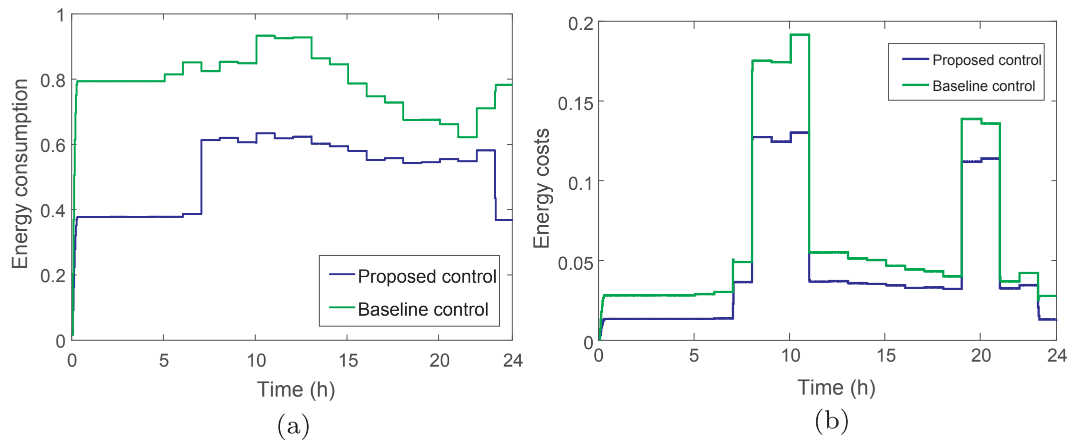


Fig. 12. (a) Energy consumption by two strategies over a 24-h period. (b) Energy cost by two strategies over a 24-h period.

Table 4
Comparison of baseline and proposed strategies.

Control strategy	Energy consumption (kW h)	Energy cost ($)	$\sum$ |PMV|
Baseline control	20.52	1.734	103.33
Proposed control	14.08	1.147	162.32
Saving (%)	31.38	33.85	

Table 5
Weather conditions for the testing days.

Date	Control	Average T_0	Average H_0	T_0^{max}	H_0^{max}
12/30	Baseline	28.6	72.4%	33.9	80%
12/31	Proposed	29.2	71.6%	34.2	81%
01/01	Proposed	28.8	72.1%	32.2	79%
01/02	Proposed	28.9	72.4%	33.2	81%
01/03	Proposed	28.1	73.2%	32.0	82%
01/04	Baseline	28.0	72.3%	32.4	79%
01/05	Baseline	27.6	73.4%	32.0	80%

the two control strategies to implement building DX A/C systems based on their different aims.

Though it is desirable to calculate the cost savings brought by the proposed control strategy over the baseline strategy, it is an impossible task for real buildings to simply compare the cost values of two control strategies in one day because load factors and ambient temperature and humidity cannot be the same in every day. To demonstrate the effectiveness of the proposed automatic hierarchical control strategy in different conditions, the proposed testing days happened to be much warmer than the baseline days in this test. The weather conditions of the testing days are shown in Table 5. The energy consumption in the testing days is shown in Fig. 13. From this comparison, all proposed control testing days have much lower power consumption, showing successful energy efficiency improvement by the proposed control strategy.

4.4. Parameter sensitivities analysis

The simulation results presented here are obtained under the assumption that the parameters are accurate and the DX A/C system models can perfectly represent the real system. However, in reality, there usually exist uncertainties in parameters and models. In this section, a simple uncertainty analysis is carried out to demonstrate how the uncertainty parameter would affect the potential performance of the proposed autonomous hierarchical control strategies. Here, we consider the uncertainties of some major parameters of the DX A/C system, namely, the heat transfer area of the DX evaporator in the dry-cooling region A_1 and the heat transfer area of the DX evaporator in the wet-cooling region A_2. The total area $A_0 = A_1 + A_2$ is known. Hence, it is only necessary to consider the effect of the uncertainty parameter A_1 on the performance of the proposed control technique. The open loop optimal controller and the closed-loop tracking of the MPC with different values of the uncertainty parameter A_1 are verified through simulation. For the case study considered here, the simulations for indoor air temperature optimised by open loop optimal controller and the closed-loop MPC temperature tracking under different parameter values are depicted in Figs. 14 and 15, and the results for the open loop optimal controller under all different ranges of the uncertainty parameter

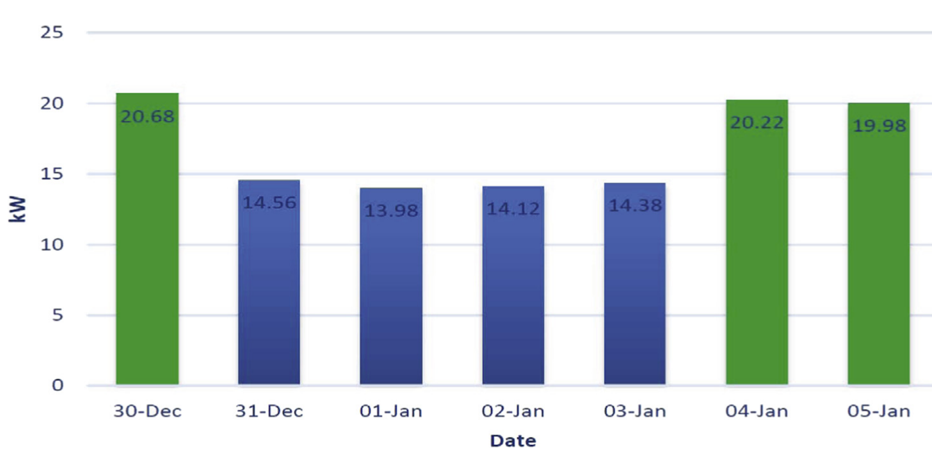


Fig. 13. Energy consumption for the proposed and baseline strategies testing days.

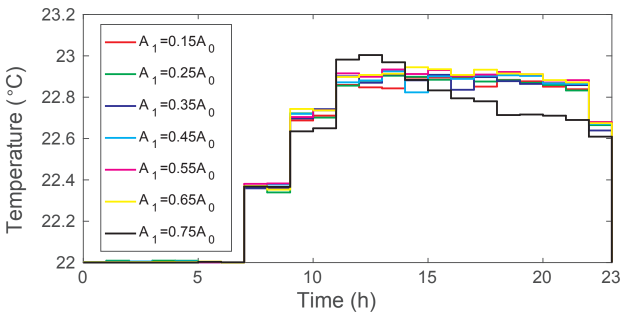


Fig. 14. Steady state of indoor air temperature optimised by the open loop controller under different values of the dry-cooling regions.

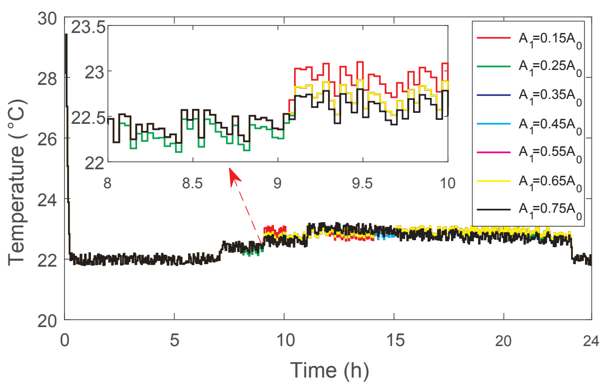


Fig. 15. Closed-loop MPC temperature tracking under different values of the dry-cooling regions.

are listed in Table 6. The standard deviations for the steady state of indoor air temperatures are less than 0.2 °C. The standard deviations for the objective function values of the open loop controller are less than 6%. The results show that the fluctuation of the control performances caused by parameter uncertainty is relatively small. Thus, the proposed autonomous hierarchical control strategy is not very sensitive to the modeling parameter A_1 specified here.

Table 6
Open loop optimization results under different values of the dry-cooling regions.

Objective function value of open loop optimization	Area of dry region	Portion of A_1 m^2	Derivation %
43.04	$(0,0.05A_0]$	$0.05A_0$	2.14%
44.23	$(0.05A_0,0.10A_0]$	$0.10A_0$	0.57%
43. 98	$\boldsymbol{(0.\ 1A_0,0.\ 15A_0]}$	$0.15A_0$	
43.00	$(0.15A_0,0.20A_0]$	$0.20A_0$	2.23%
43.20	$(0.2A_0,0.25A_0]$	$0.25A_0$	1.77%
43.48	$(0.25A_0,0.30A_0]$	$0.30A_0$	1.14%
42.87	$(0.3A_0,0.35A_0]$	$0.35A_0$	2.52%
41.57	$(0.35A_0,0.40A_0]$	$0.40A_0$	5.48%
42.54	$(0.4A_0,0.45A_0]$	$0.45A_0$	3.27%
42.47	$(0.45A_0,0.50A_0]$	$0.50A_0$	3.43%
42.21	$(0.50A_0,0.55A_0]$	$0.55A_0$	4.02%
42.26	$(0.55A_0,0.60A_0]$	$0.65A_0$	3.91%
44.00	$(0.60A_0,0.70A_0]$	$0.70A_0$	0.04%
42.22	$(0.70A_0,0.75A_0]$	$0.75A_0$	4.00%
41.91	$(0.75A_0,0.80A_0]$	$0.80A_0$	4.71%
41.60	$(0.80A_0,0.85A_0]$	$0.85A_0$	5.41%
41.52	$(0.85A_0,0.90A_0]$	$0.90A_0$	5.59%
41.70	$(0.90A_0,0.95A_0]$	$0.95A_0$	5.18%

5. Conclusions

This work formulates an autonomous hierarchical control problem to minimize energy consumption and cost while maintaining both thermal comfort and indoor air quality at the required levels for supervisory control of a direct expansion air conditioning system. It proposes an efficient control algorithm to solve the autonomous hierarchical control problem based on nonlinear programming and closed-loop model predictive control. The optimal reference points of indoor air temperature, humidity and CO_2 concentration for the direct expansion air conditioning system are obtained, and the closed-loop model predictive controller steers the direct expansion air conditioning system to reach the reference points, whereas the energy consumption and energy costs are reduced. Results show that the proposed control method could achieve a reduction of the operation energy consumption by 33.9% and cost by 33.85% with the predicted mean vote value in [-0.5,0.5], respectively. The performances of the proposed control are obtained under the assumption that the models and parameters can perfectly represent the real system. However, in reality, there usually exist uncertainties. The uncertainty analysis has been made in this paper. The results show that the proposed control method is effective because the standard deviations of energy savings are less than 5% in comparison with around 35% energy saving for normal values. The proposed control method is significant to be applied in theoretical and practical applications.

Acknowledgement

The first author would like to express my sincere appreciation to Dr. Xiangguo Xu for his valuable suggestions and academic discussion.

References

[1] Setlhaolo D, Xia X, Zhang J. Optimal scheduling of household appliances for demand response. Electric Pow Syst Res 2014;116:24–8.
[2] Setlhaolo D, Xia X. Optimal scheduling of household appliances with a battery storage system and coordination. Energy Build 2015;94(May):61–70.
[3] Setlhaolo D, Xia X. Combined residential demand side management strategies with coordination and economic analysis. Int J Electric Pow Energy Syst 2016;79:150–60.
[4] Tazvinga H, Xia X, Zhang J. Minimum cost solution of photovoltaic-diesel-battery hybrid power systems for remote consumers. Sol Energy 2013;96:292–9.
[5] Tazvinga H, Zhu B, Xia X. Energy dispatch strategy for a photovoltaic-wind-diesel-battery hybrid power system. Sol Energy 2014;108:412–20.
[6] Tazvinga H, Zhu B, Xia X. Optimal power flow management for distributed energy resources with batteries. Energy Convers Manage 2015;102:104–10.
[7] Zhu B, Tazvinga H, Xia X. Switched model predictive control for energy dispatching of a photovoltaic-diesel-battery hybrid power system. IEEE Trans Control Syst Tech 2015;23:1229–36.
[8] Wu Z, Xia X. Optimal switching renewable energy system for demand side management. Sol Energy 2015;114:278–88.
[9] Wang B, Xia X, Zhang J. A multi-objective optimization model for the life-cycle cost analysis and retrofitting planning of buildings. Energy Build 2014;77:227–35.
[10] Wang B, Xia X. Optimal maintenance planning for building energy efficiency retrofitting from optimization and control system perspectives. Energy Build 2015;96:299–308.
[11] Wu Z, Xia X, Wang B. Improving building energy efficiency by multiobjective neighborhood field optimization. Energy Build 2015;87:45–56.
[12] Wang B, Wu Z, Xia X. A multistate-based control system approach toward optimal maintenance planning. IEEE Trans Control Syst Technol 2017;25(1):374–81.
[13] Wu Z, Wang B, Xia X. Large-scale building energy efficiency retrofit: concept, model and control. Energy 2016;109:456–65.
[14] Fan Y, Xia X. A multi-objective optimization model for energy-efficiency building envelope retrofitting plan with rooftop PV system installation and maintenance. Appl Energy 2017;189:327–35.
[15] Wanjiru EM, Zhang L, Xia X. Model predictive control strategy of energy-water management in urban households. Appl Energy 2016;179:821–31.
[16] Wanjiru EM, Sichilalu SM, Xia X. Model predictive control of heat pump water heater-instantaneous shower powered with integrated renewable-grid energy systems. Appl Energy 2017;204:1333–46.
[17] Morosan P, Bourdais R, Dumur D, Buisson J. Building temperature regulation using a distributed model predictive control. Energy Build 2010;42(9):1445–52.
[18] Ma J, Qin SJ, Salsbury T, Xu P. Demand reduction in building energy systems based on economic model predictive control. Chem Eng Sci 2012;67(1):92–100.
[19] Váňa Z, Cigler J, Široký J, Žáčeková E, Ferkl L. Model-based energy efficient control

applied to an office building. J Process Control 2014;24(6):790–7.

[20] Razmara M, Maasoumy M, Shahbakhti M, Robinett RD. Optimal exergy control of building HVAC system. Appl Energy 2015;156:555–65.

[21] Zhang S, Cheng Y, Fang Z, Huan C, Lin Z. Optimization of room air temperature in stratum-ventilated rooms for both thermal comfort and energy saving. Appl Energy 2017;204:420–31.

[22] Cigler J, Prívara S, Váňa Z, Žáčeková E, Ferkl L. Optimization of predicted mean vote index within model predictive control framework: computationally tractable solution. Energy Build 2012;52:39–49.

[23] Castilla M, Álvarez JD, Normey-Rico JE, Rodríguez F. Thermal comfort control using a non-linear MPC strategy: a real case of study in a bioclimatic building. J Process Control 2014;24(6):703–13.

[24] Castilla M, Álvarez JD, Berenguel M, Rodríguez F, Guzmán JL, Pérez M. A comparison of thermal comfort predictive control strategies. Energy Build 2011;43(10):2737–46.

[25] Ma J, Qin SJ, Salsbury T. Application of economic MPC to the energy and demand minimization of a commercial building. J Process Control 2014;24(8):1282–91.

[26] Ma Y, Matuško J, Borrelli F. Stochastic model predictive control for building HVAC systems: complexity and conservation. IEEE Trans Control Syst Technol 2015;23(1):101–16.

[27] Salsbury T, Mhaskar P, Qin SJ. Predictive control methods to improve energy efficiency and reduce demand in buildings. Comput Chem Eng 2013;51:77–85.

[28] Toftum J, Fanger PO. Air humidity requirements for human comfort. ASHRAE Trans 1999;105:641–7.

[29] Freire R, Oliveira G, Mendes N. Predictive controllers for thermal comfort optimization and energy savings. Energy Build 2008;40(7):1353–65.

[30] Qi Q, Deng S. Multivariable control of indoor air temperature and humidity in a direct expansion (DX) air conditioning (A/C) system. Build Environ 2009;44(8):1659–67.

[31] Xu X, Xia L, Chan M, Deng S. Inherent correlation between the total output cooling capacity and equipment sensible heat ratio of a direct expansion air conditioning system under variable-speed operation (XXG SMD SHR DX AC unit). Appl Therm Eng 2010;30(13):1601–7.

[32] Li N, Xia L, Deng S, Xu X, Chan M. Dynamic modeling and control of a direct expansion air conditioning system using artificial neural network. Appl Energy 2012;91(1):290–300.

[33] Muñoz F, Sanchez EN, Xia Y, Deng S. Real-time neural inverse optimal control for indoor air temperature and humidity in a direct expansion (DX) air conditioning (A/C) system. Int J Refrig 2017;79:196–206.

[34] Yan H, Xia Y, Deng S. Simulation study on a three-evaporator air conditioning system for simultaneous indoor air temperature and humidity control. Appl Energy 2017;207:294–304.

[35] Yan H, Xia Y, Xu X, Deng S. Inherent operational characteristics aided fuzzy logic controller for a variable speed direct expansion air conditioning system for simultaneous indoor air temperature and humidity control. Energy Build 2018;158:558–68.

[36] Zhu Y, Jin X, Fang X, Du Z. Optimal control of combined air conditioning system with variable refrigerant flow and variable air volume for energy saving. Int J Refrig 2014;42:14–25.

[37] Zhu Y, Jin X, Du Z, Fang X, Fan B. Control and energy simulation of variable refrigerant flow air conditioning system combined with outdoor air processing unit. Appl Therm Eng 2014;64:385–95.

[38] Lin SY, Chiu SC, Chen WY. Simple automatic supervisory control system for office building based on energy-saving decoupling indoor comfort control. Energy Build 2015;86:7–15.

[39] Gladyszewska-Fiedoruk K. Correlations of air humidity and carbon dioxide concentration in the kindergarten. Energy Build 2013;62:45–50.

[40] Asif A, Zeeshan M, Jahanzaib M. Indoor temperature, relative humidity and CO_2 levels assessment in academic buildings with different heating, ventilation and air-conditioning systems. Build Environ 2018;133:83–90.

[41] Kim J, Hong T, Jeong J, Lee M, Lee M, Jeong K, et al. Establishment of an optimal occupant behavior considering the energy consumption and indoor environmental quality by region. Appl Energy 2017;204:1431–43.

[42] Mei J, Xia X. Energy-efficient predictive control of indoor thermal comfort and air quality in a direct expansion air conditioning system. Appl Energy 2017;195:439–52.

[43] Yun GY, Choi J, Kim JT. Energy performance of direct expansion air handling unit in office buildings. Energy Build 2014;77:425–31.

[44] Qi Q, Deng S. Multivariable control-oriented modeling of a direct expansion (DX) air conditioning (A/C) system. Int J Refrig 2008;31(5):841–9.

[45] Wang S, Jin X. Model-based optimal control of VAV air-conditioning system using genetic algorithm. Build Environ 2000;35:471–87.

[46] Chen W, Deng S. Development of a dynamic model for a DX VAV air conditioning system. Energy Convers Manage 2006;47(48-49):2900–24.

[47] Fanger PO. Thermal comfort: analysis and applications in environmental engineering. Copenhagen, Denmark: Danish Technical Press; 1972.

[48] Owen MS. 2009 ASHRAE handbook-fundamentals. SI ed. ASHRAE; 2009.

[49] Cena K, Clark J. Bioengineering, thermal physiology and comfort. Studies in environmental science vol. 10. Elsevier Scientific Publishing Company; 1981.

[50] Djongyang N, Tchinda R, Njomo D. Thermal comfort: a review paper. Renew Sustain Energy Rev 2010;14(9):2626–40.

[51] West SR, Ward JK, Wall J. Trial results from a model predictive control and optimisation system for commercial building HVAC. Energy Build 2014;72:271–9.

[52] Vakiloroaya V, Samali B, Pishghadam K. Investigation of energy-efficient strategy for direct expansion air-cooled air conditioning systems. Appl Therm Eng 2014;66(1–2):84–93.

[53] Xia X, Zhang J, Elaiw A. An application of model predictive control to the dynamic economic dispatch of power generation. Control Eng Pract 2011;19:638–48.

[54] Zhang J, Xia X. A model predictive control approach to the periodic implementation of the solutions of the optimal dynamic resource allocation problem. Automatica 2011;47(2):358–62.

[55] ASHRAE. Handbook of HVAC Applications. Atlanta, GA, USA: American Society of Heating, Refrigerating and Air Conditioning Engineers; 2003.

Neurocomputing 296 (2018) 100–108

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Distributed optimization of multi-agent systems with delayed sampled-data

Junxiu Yan[a,b], Hui Yu[a,b,*], Xiaohua Xia[c]

[a] College of Science, China Three Gorges University, Yichang 443002, China
[b] Three Gorges Mathematical Research Center, China Three Gorges University, Yichang 443002, China
[c] Centre of New Energy Systems, Department of Electrical, Electronics and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa

ARTICLE INFO

Article history:
Received 7 December 2017
Revised 10 March 2018
Accepted 11 March 2018
Available online 22 March 2018

Communicated by Dr xianming zhang

Keywords:
Distributed optimization
Consensus
Sampled-data
Multi-agent system

ABSTRACT

In this paper, we study the distributed optimization problem of multi-agent systems with delayed sampled-data, where the interconnected topology is directed, weighted-balanced and strongly connected, and also local cost functions are strongly convex with globally Lipschitz gradients. Based on synchronous and asynchronous sampled-data, we construct two respective algorithms. Our main results, sufficient conditions for the convergence to an optimal solution, are obtained under assumption that all design parameters are chosen properly. We also present one example to validate our theoretical results.

1. Introduction

Over the past years, distributed optimization problem has been a hot topic. As a result, there are an increasing number of studies conducted on distributed optimization in the context of control theory. Its wide range of applications can be found in various fields, such as statistical machine learning [1], smart grid [2,3], sensor networks [4], and so on. Based on multi-agent environments, objective of a distributed optimization problem is to solve an optimization problem cooperatively in a distributed way, where the objective function formed by a sum of local objective functions, and each agent can access to one local objective function only. The ultimate goal is to make states of all agents converge to the optimal solution of the optimization problem via a local computation and information exchange with its neighbors. Compared with the consensus problem of multi-agent systems, which makes all agents achieve a common state [5–12], the optimization problem of consensus does not only make all agents achieve the same state but also minimizes the optimization problem.

It is common that time-delay exists in practical systems [13–15] because of the finite speeds of information transmission and spreading as well as traffic congestions, and time-delay may result in undesirable dynamics such that the system runs out. Therefore, it is important to analyze the robustness against time-delay of a system and take time-delay into account in the algorithm design of multi-agent systems [16–18]. Meanwhile, in the real situation, agents in systems usually communicate with each other in some certain time intervals. Due to implementation of digital sensors, filters, and controllers, it is desirable that sampled-data takes place only at the discrete sampling instants but not the entire continuous process [19–21]. It is well known that effective methods to deal with sampled-data consensus problems is one of the input delay approach [22–24]. The consensus of multi-agent systems with both sampling data and time-delay is considered in [25]. The simultaneous stability problem of a finite number of linear subsystems is studied in [26] under asynchronous and aperiodic sampling, time-varying delays, and measurement errors. Furthermore, [27,28] provided overviews of recent advances in distributed sampled-data cooperative control and event-triggered consensus of multi-agent systems, respectively. Thus, the consensus problem with sampled-data and time-delay is a meaningful research topic.

In distributed optimization problems of multi-agent systems, most algorithms in earlier works were time-varying, consensus-based dynamics implemented in discrete time [29–31]. In the context of time-varying network topology, discrete time subgradient algorithms are proposed for unconstrained, separable, convex optimization problems in [29,30]. Recent works have introduced continuous-time methods whose convergence properties can be

* Corresponding author at: College of Science, China Three Gorges University, Yichang 443002, China.
E-mail addresses: Yanjunx_2017@163.com (J. Yan), yuhui@ctgu.edu.cn (H. Yu), xxia@up.ac.za (X. Xia).

https://doi.org/10.1016/j.neucom.2018.03.036
0925-2312/

analyzed via classical stability theory. Based on the gradient algorithm and integral feedback, auxiliary-variables are introduced in stability analysis of continuous-time dynamical systems [32–34]. From the control system viewpoint, a continuous-time multi-agent system is proposed with strongly connected and weight-balanced directed communication topology in [32]. A modified system is proposed in [33] with auxiliary-variables no longer need to exchange information, where centralized synchronous and distributed asynchronous event-triggered communication schemes are also considered to reduced communication bandwidth. In [34], time-delays are considered in continuous-time multi-agent systems for distributed optimization and a sampled-data communication scheme is formulated based on the results of delay systems, where conditions are derived in form of Linear Matrix Inequality(LMI). In order to avoid using auxiliary-variables, a family of Zero-Gradient-Sum algorithms are proposed over fixed communication topology in [35]. In [36,37], the continuous time Zero-Gradient-Sum algorithm, sampled-data, event-triggered communication for distributed convex optimization problem are considered over directed networks and undirected, connected networks, respectively. In [38], a periodic event-triggered consensus of first-order time-delayed multi-agent systems under switching topologies can be achieved with appropriate choices of the event-triggering parameters, sampling period, and time-delay. Moreover, output consensus problem of delayed sampled-data is considered in [39], where the data of the system is sampled at a sampling instant but can be available with a time-delay. However, so far, studies on the distributed optimization problem of multi-agent systems with delayed sampled-data are rare.

In this paper, the distributed consensus optimization problem of multi-agent systems with delayed sampled-data is considered. The interconnected graph is assumed to be directed, strongly connected and weight balanced. Only available data of the system is assumed to be sampled and delayed. Local costs are strongly convex with global Lipschitz gradients. Two control algorithms under synchronous and asynchronous sampled-data are proposed for the sampled-data multi-agent systems to reach the consensus and optimal state, respectively. A stability analysis is conducted based on Lyapunov theory and algebraic graph theory. Finally, sufficient conditions are obtained such that optimization problems can be solved in the consensus state.

The main contributions of this paper are listed as follows: Firstly, two control algorithms using sampled-data with time-delay under synchronous and asynchronous sampling are presented for the considered multi-agent systems, respectively. Secondly, sufficient conditions are obtained to guarantee the convergence to the optimal solution. In general, the multi-agent system with sampled-data is transformed into time-delay system, and then LMI conditions can be obtained such as in [34]. Other works related to this issue are based on event-triggered scheme due to the advantages of reducing communication resources such as in [37]. The main differences between this paper and previously mentioned works are that the sampled-data becomes available with a time-delay, and then some inequalities conditions are obtained such that the parameters can be chosen properly. In other words, we consider a distributed optimization problem of multi-agent systems with delayed sampled-data in this paper. To the best of our knowledge, no similar results appear in the existing literatures.

This paper is organized as follows. Some preliminaries on algebraic graph theory, useful lemmas and model formulation are presented in Section 2. The convergence results of the proposed algorithm are established and proved under a given communication condition on network topology in Section 3. An example is provided to illustrate results in this paper in Section 4. Finally, this paper concludes in Section 5.

Notations: $\mathcal{R}$ and $\mathcal{R}^n$ represent the set of real numbers and the set of $n \times 1$ real vectors, respectively; $I_n \in \mathcal{R}^{n\times n}$ is the $n \times n$ identity matrix; $\mathbf{1}_n$ (or $\mathbf{0}_n$) denotes an n dimensional column vector whose all entries being 1 (or 0); A^T represents the transpose of a matrix A; for vectors $x_1, x_2, \dots, x_n$, $\mathrm{col}(x_1, x_2, \dots, x_n) = [x_1^T, x_2^T, \dots, x_n^T]^T$; for a vector w, $\|w\| = \sqrt{w^T w}$ represents the standard Euclidean norm; for a matrix P, $\lambda_{\min}(P)$ and $\lambda_{\max}(P)$ denote the smallest and largest eigenvalue.

2. Preliminaries and problem statement

2.1. Preliminaries

For a multi-agent system, the information exchange among N agents can be modeled by a weighted digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ with the finite set of nodes $\mathcal{V} = \{1, 2, \dots, N\}$ and edge set $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. An edge starts from i and ends on j, which means that agent j can obtain information from agent i. The weighted adjacency matrix $\mathcal{A} = [a_{ij}] \in R^{N\times N}$ with $a_{ij} > 0$ if $(j, i) \in \mathcal{E}$ and $a_{ij} = 0$ otherwise. If $\sum_{j=1}^N a_{ij} = \sum_{j=1}^N a_{ji}$ for all $i \in \mathcal{V}$, the digraph $\mathcal{G}$ is called weighted-balanced. A path is a sequence of connected edges in a graph. If for every pair of nodes there is a directed path connecting them, the digraph $\mathcal{G}$ is said to be strongly connected, otherwise disconnected. The Laplacian $L = [\ell_{ij}] \in R^{N\times N}$ of graph $\mathcal{G}$ is defined by

$$\ell_{ij} = \begin{cases} \sum_{k=1,k\neq i}^N a_{ik} & j = i \\ -a_{ij} & j \neq i \end{cases}.$$

The next lemmas related to the important properties of Laplace L and provide useful mathematical tools.

Lemma 1 [40]**.** *Laplace matrix L has least one zero eigenvalue with $\mathbf{1}_N = [1, 1, \dots, 1] \in \mathcal{R}^N$ as its eigenvector, and all the non-zero eigenvalues of L have positive real parts. Laplacian L has a simple zero eigenvalue if and only if $\mathcal{G}$ is strongly connected.*

Lemma 2. *For matrices A, B, C and D with appropriate dimensions, the Kronecker product $\otimes$ satisfies $(1)(A\otimes B)(C\otimes D) = (AC)\otimes(BD)$; $(2)(A\otimes B)^T = A^T \otimes B^T$; $(3)(A\otimes B)^{-1} = A^{-1}\otimes B^{-1}$.*

Lemma 3 [41]**.** *For a given $n\times n-$matrix $G > 0$ and for all continuous functions ω in $[a, b] \to \mathcal{R}^n$, the following inequality holds:*

$$\left[\int_a^b \omega(s)ds\right]^T G\left[\int_a^b \omega(s)ds\right] \leq (b-a)\int_a^b \omega^T(s)G\omega(s)ds.$$

2.2. Problem statement

We consider a multi-agent system consisting of N agents. The dynamics of the ith agent, $i \in \mathcal{V}$, are described by

$$\dot{x}_i(t) = u_i(t), \tag{1}$$

where $x_i \in \mathcal{R}^m$ denotes the state of agent i, $u_i \in \mathcal{R}^m$ is the control input.

Consider the multi-agent optimization problem, in which the goal is to minimize the sum of local cost functions associated to the individual agent. More specially, it can be expressed as

$$\text{minimize} \quad f(x) = \sum_{i=1}^N f_i(x), x \in \mathcal{R}^m. \tag{2}$$

Let $\boldsymbol{x} = \mathrm{col}(x_1, x_2, \dots, x_N) \in \mathcal{R}^{Nm}$. Next, we provide an alternative formulation of (2), i.e.,

$$\begin{aligned} &\text{minimize} \quad f(\boldsymbol{x}) = \sum_{i=1}^N f_i(x_i), x_i \in \mathcal{R}^m, \\ &\text{subject to} \quad (L\otimes I_m)\boldsymbol{x} = 0_{Nm}. \end{aligned} \tag{3}$$

We can see problem (2) on $\mathcal{R}^m$ is equivalent to problem (3) on $\mathcal{R}^{Nm}$.

In this paper, our goal is to design a distributed controller for each agent such that the states of all agents converge to a optimal solution of the optimization problem (2) via local communication.

Before proceed, we give the following assumption on the local cost function f_i based on the convex analysis [42].

Assumption 1. (a) For each $i\in\mathcal{V}$, f_i is differentiable and its gradient is Lipschitz with constant $\rho_i>0$ in $\mathcal{R}^m$:

$$\|\nabla f_i(x)-\nabla f_i(y)\|\leq\rho_i\|x-y\|,\quad\forall x,y\in\mathcal{R}^m. \tag{4}$$

(b) For $i\in\mathcal{V}$, f_i is m_i-strongly convex with constant $m_i>0$:

$$(x-y)^T(\nabla f_i(x)-\nabla f_i(y))\geq m_i\|x-y\|^2,\quad\forall x,y\in\mathcal{R}^m. \tag{5}$$

Remark 1. Under Assumption 1(b), we can note that f is strictly convex, then the optimization problem (3) has an unique optimal solution.

Assumption 2. The directed graph $\mathcal{G}$ is weighted-balanced and strongly connected.

Remark 2. From Assumption 2, zero is a simple eigenvalue of matrix L and $1_N^TL=0$. Moreover, there exists a matrix $Q\in\mathcal{R}^{N\times(N-1)}$ with

$$\mathbf{1}_N^TQ=0,\quad Q^TQ=I_{N-1},\quad QQ^T=I_N-\frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T, \tag{6}$$

such that the matrix $Q^TLQ=H$, where the real parts of all eigenvalues of H are positive, and $H+H^T$ is a positive definite matrix.

3. Main results

3.1. Synchronous sampling

The state $x_i(t)$ of system (1) and (3) is assumed to be sampled at time instants t_k and available at $t_k+\tau_k$. $\{t_k\}$, $k=0,1,\ldots,\infty$, is a strictly increasing sequence such that $\lim_{k\to\infty}t_k=\infty$ and $\tau_k\geq 0$, that is the sampled-data $x_i(t_k)$ is available with a time-delay τ_k. The sampling interval $[t_{k-1},t_k)$ satisfy $0<T_{\min}\leq t_k-t_{k-1}=T_k\leq T_{\max}$ for all $k=0,1,\ldots,\infty$, where T_k is the length of the kth sampling interval, $T_{\min}=\min\{T_k\}$ and $T_{\max}=\max\{T_k\}$, and when $t\in[t_k+\tau_k,t_{k+1}+\tau_{k+1})$, $x_i(t_{k+1}+\tau_{k+1})=\lim_{t\to(t_{k+1}+\tau_{k+1})^-}x_i(t)$. We assume that τ is the upper bound of τ_k, that is $\tau_k\leq\tau$, and satisfy $\tau<T_{\min}$, which means that the sampled-data at time t_k can be used before next sampling time instant.

We use the following sampled-data based control algorithm to achieve consensus and optimum:

$$\begin{aligned}
u_i(t)&=-k\sum_{j=1}^N a_{ij}[x_i(t_k)-x_j(t_k)]-w_i(t)-\gamma\nabla f_i(x_i(t)),\\
\dot{w}_i(t)&=\alpha\sum_{j=1}^N a_{ij}[x_i(t_k)-x_j(t_k)],\\
w_i(0)&=0,\quad t\in[t_k+\tau_k,t_{k+1}+\tau_{k+1}),\quad k\geq 0,
\end{aligned} \tag{7}$$

where $w_i(t)$ is an auxiliary state of agent i and k, α, γ are the scalar tuning positive parameter. It can be seen from (7) that each agent only uses the information at time t_k when $t\in[t_k+\tau_k,t_{k+1}+\tau_{k+1}),k\geq 0$.

Noticed that $x_i(t_k)(i\in\mathcal{V})$ are constant in all of the time intervals $[t_k+\tau_k,t_{k+1}+\tau_{k+1}),k\geq 0$. From the second equation of (7), we can know that $w_i(t)$ is continuous in $[t_k+\tau_k,t_{k+1}+\tau_{k+1})$. From Assumption 1, $\nabla f_i(x_i(t))$ is Lipschitz and then continuous in $[t_k+\tau_k,t_{k+1}+\tau_{k+1})$. Therefore, $\dot{x}_i(t)$ and then $x_i(t)$ is continuous in time intervals $[t_k+\tau_k,t_{k+1}+\tau_{k+1}),k\geq 0$. According to the definition $x_i(t_{k+1}+\tau_{k+1})=\lim_{t\to(t_{k+1}+\tau_{k+1})^-}x_i(t)$ in the beginning of this section, we have $\lim_{t\to(t_{k+1}+\tau_{k+1})^-}x_i(t)=\lim_{t\to(t_{k+1}+\tau_{k+1})^+}x_i(t)=x_i(t_{k+1}+\tau_{k+1})$, which means that $x_i(t)$ is continuous in the time instant $t_{k+1}+\tau_{k+1}$. Thus, $x_i(t)(i\in\mathcal{V})$ are continuous in the time interval $[t_0,\infty)$.

Let

$$\boldsymbol{w}(t)=\mathrm{col}(w_1(t),w_2(t),\ldots,w_N(t)),$$

and

$$\nabla\overline{f}(x(t))=\mathrm{col}(\nabla f_1(x_1(t)),\nabla f_2(x_2(t)),\ldots,\nabla f_N(x_N(t))).$$

Then the closed-loop systems of (1) and (7) can be expressed as a compact form:

$$\begin{aligned}
\dot{\boldsymbol{x}}(t)&=-k(L\otimes I_m)\boldsymbol{x}(t_k)-\boldsymbol{w}(t)-\gamma\nabla\overline{f}(\boldsymbol{x}(t)),\\
\dot{\boldsymbol{w}}(t)&=\alpha(L\otimes I_m)\boldsymbol{x}(t_k).
\end{aligned} \tag{8}$$

Let the right-side of the closed-loop system (8) equal to 0, then we can get a equilibrium point $(\boldsymbol{x}^*,\boldsymbol{w}^*)$, i.e.

$$\begin{aligned}
-k(L\otimes I_m)\boldsymbol{x}^*-\boldsymbol{w}^*-\gamma\nabla\overline{f}(\boldsymbol{x}^*)&=0,\\
\alpha(L\otimes I_m)\boldsymbol{x}^*&=0.
\end{aligned} \tag{9}$$

According to the properties of Laplacian matrix, and from (9), one can obtain

$$\begin{aligned}
\boldsymbol{x}^*&=\mathbf{1}_N\otimes\pi,\ \pi\in\mathcal{R}^m,\\
\boldsymbol{w}^*&=-\gamma\nabla\overline{f}(\boldsymbol{x}^*).
\end{aligned} \tag{10}$$

Under Assumption 2, we have $\mathbf{1}_N^TL=0$. Left multiplying the second equation of (8) by $\mathbf{1}_N^T\otimes I_m$, we obtain $\sum_{j=1}^N\dot{w}_j(t)=0$, and using initial condition $w_i(0)=0$, then

$$\sum_{j=1}^N w_j(t)=\sum_{j=1}^N w_j(0)=0,\quad\forall t\geq 0. \tag{11}$$

Using $\mathbf{1}_N^T\otimes I_m$ left multiply the second equation of (10) again results in

$$0=\sum_{j=1}^N w_j^*=-\gamma(\mathbf{1}_N^T\otimes I_m)\nabla\overline{f}(\boldsymbol{x}^*)=-\gamma\sum_{j=1}^N\nabla f_i(\pi)=-\gamma\nabla f(\boldsymbol{x}^*).$$

Thus, the optimal condition $\nabla f(\boldsymbol{x}^*)=0$ is satisfied, which means that $\boldsymbol{x}^*=\mathbf{1}_N\otimes x^*,x^*\in\mathcal{R}^m$ is the optimal solution of the optimization problem (3).

Using the transformation

$$\overline{\boldsymbol{x}}(t)=\boldsymbol{x}(t)-\boldsymbol{x}^*,\overline{\boldsymbol{w}}(t)=\boldsymbol{w}(t)-\boldsymbol{w}^*, \tag{12}$$

one can shift the equilibrium point into the origin, then system (8) can be transformed into the following form:

$$\begin{aligned}
\dot{\overline{\boldsymbol{x}}}(t)&=-k(L\otimes I_m)\overline{\boldsymbol{x}}(t_k)-\overline{\boldsymbol{w}}(t)-\gamma\Psi(\overline{\boldsymbol{x}}(t)),\\
\dot{\overline{\boldsymbol{w}}}(t)&=\alpha(L\otimes I_m)\overline{\boldsymbol{x}}(t_k),
\end{aligned} \tag{13}$$

where $\Psi(\overline{\boldsymbol{x}}(t))=\nabla\overline{f}(\boldsymbol{x}(t))-\nabla\overline{f}(\boldsymbol{x}^*)$.

Let

$$e(t)=(T^T\otimes I_m)\overline{\boldsymbol{x}}(t),\quad\vartheta(t)=(T^T\otimes I_m)\overline{\boldsymbol{w}}(t),\quad T=\begin{bmatrix}\frac{\mathbf{1}_N}{\sqrt{N}} & Q\end{bmatrix}.$$

Denote $e=\mathrm{col}(e_1,e_2)$, and $\vartheta=\mathrm{col}(\vartheta_1,\vartheta_2)$ with $e_1,\vartheta_1\in\mathcal{R}^m$, and $e_2,\vartheta_2\in\mathcal{R}^{m(N-1)}$. By the structure of T and (6), we can know T is an orthogonal matrix. Then system (13) can be rewritten as:

$$\begin{aligned}
\dot{e_1}(t)&=-\gamma\left(\frac{\mathbf{1}_N^T}{\sqrt{N}}\otimes I_m\right)\Psi(\overline{\boldsymbol{x}}(t)),\\
\dot{e_2}(t)&=-k(H\otimes I_m)e_2(t_k)-\vartheta_2(t)-\gamma(Q^T\otimes I_m)\Psi(\overline{\boldsymbol{x}}(t)),\\
\dot{\vartheta}_1(t)&=0,\\
\dot{\vartheta}_2(t)&=\alpha(H\otimes I_m)e_2(t_k).
\end{aligned} \tag{14}$$

Let $\varepsilon(t)=\mathrm{col}(e_2(t),\vartheta_2(t))$, then

$$\dot{\varepsilon}(t)=C\varepsilon(t)-E\int_{t_k}^{t}\dot{\varepsilon}(s)ds+F(t), \tag{15}$$

with

$$C=\begin{pmatrix}-kH & -I_{N-1}\\ \alpha H & 0\end{pmatrix}\otimes I_m,\quad E=\begin{pmatrix}-kH & 0\\ \alpha H & 0\end{pmatrix}\otimes I_m,$$

and

$$F(t)=\begin{pmatrix}-\gamma(Q^T\otimes I_m)\Psi(\bar{\boldsymbol{x}}(t))\\ 0\end{pmatrix}.$$

Then the main results can be obtained as follows.

Theorem 1. *Suppose Assumptions 1 and 2 hold, the optimization problem (3) for multi-agent system (1) can be solved by the optimization control (7), if the following conditions are satisfied:*

$$\underline{m}-(3m+1)\gamma\overline{\rho}^2>0, \tag{16}$$

$$\underline{\lambda}_1-6m[(\alpha^2+k^2)\overline{\lambda}_2+1]-\frac{1}{2}>0, \tag{17}$$

and

$$T_{\max}+\tau<\sqrt{\frac{m}{\overline{\lambda}_2[(k-\alpha)^2+3m(\alpha^2+k^2)]}}, \tag{18}$$

where $m>0$ is a constant, $\underline{\lambda}_1=\lambda_{\min}(R)$, $\overline{\lambda}_2=\lambda_{\max}(H^TH)$, and $R=\begin{pmatrix}(k-\alpha)(H+H^T) & I_{N-1}\\ I_{N-1} & 2I_{N-1}\end{pmatrix}\otimes I_m$.

Proof. Consider the following Lyapunov function:

$$V_1(t)=\frac{1}{2}e_1^T(t)e_1(t)+\frac{1}{2}\boldsymbol{\varepsilon}^T(t)\Omega\boldsymbol{\varepsilon}(t),$$

where $\Omega=\begin{pmatrix}I_{N-1} & I_{N-1}\\ I_{N-1} & \frac{k}{\alpha}I_{N-1}\end{pmatrix}\otimes I_m$ is positive definite for $k>\alpha$, the condition $k>\alpha$ will be proved later. The derivation of V_1 along the first equality of (14) and - system (15) yields:

$$\begin{aligned}\dot{V}_1 &= e_1^T(t)\dot{e}_1(t)+\frac{1}{2}\varepsilon^T(t)(\Omega C+C^T\Omega)\varepsilon(t)\\ &\quad-\varepsilon^T(t)\Omega E\int_{t_k}^{t}\dot{\varepsilon}(s)ds+\varepsilon^T(t)\Omega F(t).\end{aligned} \tag{19}$$

Due to $e(t)=(T^T\otimes I_m)\bar{\boldsymbol{x}}(t)$ and from Assumption 1, we have

$$\begin{aligned}&e_1^T(t)\dot{e}_1(t)+\varepsilon^T(t)\Omega F(t)\\ &=-\gamma e_1^T(t)\left(\frac{\mathbf{1}_N^T}{\sqrt{N}}\otimes I_m\right)\Psi(\bar{\boldsymbol{x}}(t))+\varepsilon^T(t)\Omega F(t)\\ &=-\gamma\bar{\boldsymbol{x}}^T(t)\Psi(\bar{\boldsymbol{x}}(t))+\gamma e_2^T(t)(Q^T\otimes I_m)\Psi(\bar{\boldsymbol{x}}(t))\\ &\quad-\gamma e_2^T(t)(Q^T\otimes I_m)\Psi(\bar{\boldsymbol{x}}(t))-\gamma\vartheta_2^T(t)(Q^T\otimes I_m)\Psi(\bar{\boldsymbol{x}}(t))\\ &=-\gamma\bar{\boldsymbol{x}}^T(t)\Psi(\bar{\boldsymbol{x}}(t))-\gamma\vartheta_2^T(t)(Q^T\otimes I_m)\Psi(\bar{\boldsymbol{x}}(t))\\ &\le-\gamma\underline{m}\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t)+\frac{1}{4}\vartheta_2^T(t)\vartheta_2(t)+\gamma^2\|(Q^T\otimes I_m)\Psi(\bar{\boldsymbol{x}}(t))\|^2\\ &\le-\gamma\underline{m}\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t)+\frac{1}{4}\vartheta_2^T(t)\vartheta_2(t)+\gamma^2\overline{\rho}^2\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t),\end{aligned} \tag{20}$$

where $\underline{m}=\min\{m_1,m_2,\ldots,m_N\}$, $\overline{\rho}=\max\{\rho_1,\rho_2,\ldots,\rho_N\}$.

Let $R=-(\Omega C+C^T\Omega)$, according to condition (17), we can know R is positive definite, and due to $H+H^T$ is positive definite, then $k>\alpha$. Thus

$$\begin{aligned}&-\varepsilon^T(t)\Omega E\int_{t_k}^{t}\dot{\varepsilon}(s)ds\\ &=(k-\alpha)e_2^T(t)(H\otimes I_m)\int_{t_k}^{t}\dot{e}_2(s)ds\\ &\le\frac{1}{4}e_2^T(t)e_2(t)+(k-\alpha)^2\overline{\lambda}_2\left(\int_{t_k}^{t}\dot{e}_2(s)ds\right)^T\left(\int_{t_k}^{t}\dot{e}_2(s)ds\right)\\ &\le\frac{1}{4}e_2^T(t)e_2(t)+(k-\alpha)^2\overline{\lambda}_2\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)^T\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right).\end{aligned} \tag{21}$$

From (19)–(21) and $\boldsymbol{\varepsilon}^T(t)R\boldsymbol{\varepsilon}(t)\ge\underline{\lambda}_1\boldsymbol{\varepsilon}^T(t)\boldsymbol{\varepsilon}(t)$, we have

$$\begin{aligned}\dot{V}_1(t)&\le-\gamma\underline{m}\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t)+\frac{1}{4}\vartheta_2^T(t)\vartheta_2(t)+\gamma^2\overline{\rho}^2\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t)\\ &\quad-\frac{1}{2}\varepsilon^T(t)R\varepsilon(t)\\ &\quad+\frac{1}{4}e_2^T(t)e_2(t)+(k-\alpha)^2\overline{\lambda}_2\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)^T\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)\\ &\le-\gamma\underline{m}\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t)+\frac{1}{4}\varepsilon^T(t)\varepsilon(t)+\gamma^2\overline{\rho}^2\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t)\\ &\quad-\frac{1}{2}\underline{\lambda}_1\varepsilon^T(t)\varepsilon(t)\\ &\quad+(k-\alpha)^2\overline{\lambda}_2\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)^T\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right).\end{aligned} \tag{22}$$

Construct the following auxiliary integral function

$$V_2(t)=\int_{t-T_{\max}-\tau}^{t}\int_{\theta}^{t}\dot{\varepsilon}^T(s)\dot{\varepsilon}(s)dsd\sigma,$$

we can obtain

$$\dot{V}_2(t)=(T_{\max}+\tau)\dot{\varepsilon}^T(t)\dot{\varepsilon}(t)-\int_{t-T_{\max}-\tau}^{t}\dot{\varepsilon}^T(s)\dot{\varepsilon}(s)ds.$$

By calculation, we have

$$\begin{aligned}\dot{\varepsilon}^T(t)\dot{\varepsilon}(t)&=\varepsilon^T(t)C^TC\varepsilon(t)+\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)^TE^TE\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)\\ &\quad+F^T(t)F(t)\\ &\quad-2\varepsilon^T(t)C^TE\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)-2\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)^TE^TF(t)\\ &\quad+2\varepsilon^T(t)C^TF(t).\end{aligned} \tag{23}$$

Due to $2a^Tb\le a^TXa+b^TX^{-1}b$, we have

$$\begin{aligned}-2\varepsilon^T(t)C^TE\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)&\le\varepsilon^T(t)C^TC\varepsilon(t)+\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)^T\\ &\quad\times E^TE\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right),\end{aligned} \tag{24}$$

$$\begin{aligned}-2\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)^TE^TF(t)&\le\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)^TE^TE\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)\\ &\quad+F^T(t)F(t),\end{aligned} \tag{25}$$

and

$$2\varepsilon^T(t)C^TF(t)\le\varepsilon^T(t)C^TC\varepsilon(t)+F^T(t)F(t). \tag{26}$$

Based on Assumption 1(a), the following result can be obtained:

$$F^T(t)F(t)=\gamma^2\|(Q^T\otimes I_m)\Psi(\bar{\boldsymbol{x}}(t))\|^2\le\gamma^2\overline{\rho}^2\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t). \tag{27}$$

From (23)–(27), we have

$$\begin{aligned}\dot{\varepsilon}^T(t)\dot{\varepsilon}(t)&\le3[\varepsilon^T(t)C^TC\varepsilon(t)+\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)^TE^TE\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)\\ &\quad+F^T(t)F(t)]\\ &\le3[(\alpha^2+k^2)\overline{\lambda}_2+1]\varepsilon^T(t)\varepsilon(t)+3\gamma^2\overline{\rho}^2\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t)\end{aligned}$$

$$+3(\alpha^2+k^2)\overline{\lambda}_2\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)^T\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right). \tag{28}$$

Let $V(t)=V_1(t)+\frac{m}{T_{\max}+\tau}V_2(t)$, we have

$$\begin{aligned}\dot{V}(t) \le & -\gamma\underline{m}\overline{\boldsymbol{x}}^T(t)\overline{\boldsymbol{x}}(t)+\frac{1}{4}\varepsilon^T(t)\varepsilon(t)+\gamma^2\overline{\rho}^2\overline{\boldsymbol{x}}^T(t)\overline{\boldsymbol{x}}(t)\\ & -\frac{1}{2}\underline{\lambda}_1\varepsilon^T(t)\varepsilon(t)\\ & +(k-\alpha)^2\overline{\lambda}_2\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)^T\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)\\ & +3m[(\alpha^2+k^2)\overline{\lambda}_2+1]\varepsilon^T(t)\varepsilon(t)+3m\gamma^2\overline{\rho}^2\overline{\boldsymbol{x}}^T(t)\overline{\boldsymbol{x}}(t)\\ & +3m(\alpha^2+k^2)\overline{\lambda}_2\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)^T\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)\\ & -\frac{m}{T_{\max}+\tau}\int_{t-T_{\max}-\tau}^{t}\dot{\varepsilon}^T(s)\dot{\varepsilon}(s)ds. \end{aligned}\tag{29}$$

From Lemma 3, we have

$$\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)^T\left(\int_{t_k}^{t}\dot{\varepsilon}(s)ds\right)\le (t-t_k)\int_{t_k}^{t}\dot{\varepsilon}^T(s)\dot{\varepsilon}(s)ds,$$

where $t\in[t_k+\tau_k, t_{k+1}+\tau_{k+1})$, $t-t_k\le T_{\max}+\tau$, that is $t-T_{\max}-\tau\le t_k$. Then, we have

$$\begin{aligned}\dot{V}(t) \le & -\gamma\underline{m}\overline{\boldsymbol{x}}^T(t)\overline{\boldsymbol{x}}(t)+\frac{1}{4}\varepsilon^T(t)\varepsilon(t)+\gamma^2\overline{\rho}^2\overline{\boldsymbol{x}}^T(t)\overline{\boldsymbol{x}}(t)\\ & -\frac{1}{2}\underline{\lambda}_1\varepsilon^T(t)\varepsilon(t)\\ & +(k-\alpha)^2\overline{\lambda}_2(t-t_k)\int_{t_k}^{t}\dot{\varepsilon}^T(s)\dot{\varepsilon}(s)ds\\ & +3m[(\alpha^2+k^2)\overline{\lambda}_2+1]\varepsilon^T(t)\varepsilon(t)+3m\gamma^2\overline{\rho}^2\overline{\boldsymbol{x}}^T(t)\overline{\boldsymbol{x}}(t)\\ & +3m(\alpha^2+k^2)\overline{\lambda}_2(t-t_k)\int_{t_k}^{t}\dot{\varepsilon}^T(s)\dot{\varepsilon}(s)ds\\ & -\frac{m}{T_{\max}+\tau}\int_{t-T_{\max}-\tau}^{t}\dot{\varepsilon}^T(s)\dot{\varepsilon}(s)ds. \end{aligned}\tag{30}$$

Note that

$$(t-t_k)\int_{t_k}^{t}\dot{\varepsilon}^T(s)\dot{\varepsilon}(s)ds\le (T_{\max}+\tau)\int_{t-T_{\max}-\tau}^{t}\dot{\varepsilon}^T(s)\dot{\varepsilon}(s)ds,$$

we have

$$\begin{aligned}\dot{V}(t) \le & -\gamma\underline{m}\overline{\boldsymbol{x}}^T(t)\overline{\boldsymbol{x}}(t)+\frac{1}{4}\varepsilon^T(t)\varepsilon(t)+\gamma^2\overline{\rho}^2\overline{\boldsymbol{x}}^T(t)\overline{\boldsymbol{x}}(t)\\ & -\frac{1}{2}\underline{\lambda}_1\varepsilon^T(t)\varepsilon(t)\\ & +(k-\alpha)^2\overline{\lambda}_2(T_{\max}+\tau)\int_{t-T_{\max}-\tau}^{t}\dot{\varepsilon}^T(s)\dot{\varepsilon}(s)ds\\ & +3m[(\alpha^2+k^2)\overline{\lambda}_2+1]\varepsilon^T(t)\varepsilon(t)+3m\gamma^2\overline{\rho}^2\overline{\boldsymbol{x}}^T(t)\overline{\boldsymbol{x}}(t)\\ & +3m(\alpha^2+k^2)\overline{\lambda}_2(T_{\max}+\tau)\int_{t-T_{\max}-\tau}^{t}\dot{\varepsilon}^T(s)\dot{\varepsilon}(s)ds\\ & -\frac{m}{T_{\max}+\tau}\int_{t-T_{\max}-\tau}^{t}\dot{\varepsilon}^T(s)\dot{\varepsilon}(s)ds, \end{aligned}\tag{31}$$

and then

$$\begin{aligned}\dot{V}(t) \le & -[\gamma\underline{m}-(3m+1)\gamma^2\overline{\rho}^2]\overline{\boldsymbol{x}}^T(t)\overline{\boldsymbol{x}}(t)\\ & -\left\{\frac{1}{2}\underline{\lambda}_1-3m[(\alpha^2+k^2)\overline{\lambda}_2+1]-\frac{1}{4}\right\}\varepsilon^T(t)\varepsilon(t)\\ & -\Big[\frac{m}{T_{\max}+\tau}-(k-\alpha)^2\overline{\lambda}_2(T_{\max}+\tau)\\ & -3m(\alpha^2+k^2)\overline{\lambda}_2(T_{\max}+\tau)\Big]\int_{t-T_{\max}-\tau}^{t}\dot{\varepsilon}^T(s)\dot{\varepsilon}(s)ds. \end{aligned}\tag{32}$$

Hence, conditions (16)–(18) guarantee that $\dot{V}(t)<0$. Based on Lyapunov stability theory, we can conclude that $e_1(t)\to 0$ and $\boldsymbol{\varepsilon}(t)\to 0$, that is $e(t)\to 0_{mN}$, $\vartheta(t)\to 0_{mN}$ as $t\to\infty$.

With the transformation $\overline{\boldsymbol{x}}(t)=(T\otimes I_m)e(t)$ and $\overline{\boldsymbol{w}}(t)=(T\otimes I_m)\vartheta(t)$ and T is a orthogonal matrix, we can obtain $\overline{\boldsymbol{x}}(t)\to 0_{mN}$, $\overline{\boldsymbol{w}}(t)\to 0_{mN}$, which means $\boldsymbol{x}(t)\to\boldsymbol{x}^*$, $\boldsymbol{w}(t)\to\boldsymbol{w}^*$ as $t\to\infty$. As a result, this proof is completed. □

3.2. Asynchronous sampling

Based on the dynamics (1) and the optimization problem (3), we assume that each agent i independently samples its own state at sampling instant t_k^i and the sampled-data is available at $t_k^i+\tau_k^i$, $i\in\mathcal{V}, k=0,1,\ldots,\infty$, which is determined by its own clock. $\{t_k^i\}$ is a strictly increasing sequence such that $\lim_{k\to\infty}t_k^i=\infty$. The sampling interval $[t_k^i, t_{k+1}^i)$ satisfies $0<T_{\min}\le t_{k+1}^i-t_k^i\le T_{\max}$ for all $k\ge 0$. $\tau_k^i>0$ denote the transmission delay with an upper bound $\tau\ge\max\{\tau_k^i\}$, and satisfy $\tau<T_{\min}$, which means that the sampled-data at time t_k^i can be used before next sampling time instant. When $t\in[t_k^i+\tau_k^i, t_{k+1}^i+\tau_{k+1}^i)$, $x_i(t_{k+1}^i+\tau_{k+1}^i)=\lim_{t\to(t_{k+1}^i+\tau_{k+1}^i)^-}x_i(t)$.

The following asynchronous sampled-data control algorithm is proposed:

$$\begin{aligned} u_i(t) &= -k\sum_{j=1}^{N}a_{ij}[x_i(t_k^i)-x_j(t_k^j)]-w_i(t)-\gamma\nabla f_i(x_i(t)),\\ \dot{w}_i(t) &= \alpha\sum_{j=1}^{N}a_{ij}[x_i(t_k^i)-x_j(t_k^j)],\\ w_i(0) &= 0,\quad t\in[t_k^i+\tau_k^i, t_{k+1}^i+\tau_{k+1}^i),\quad k\ge 0. \end{aligned}\tag{33}$$

Similar mechanism of asynchronous sampling for consensus problem of multi-agent systems can be found in [26].

By the similar discussion as in the Section 3.1, we can also conclude that $\lim_{t\to(t_{k+1}^i+\tau_{k+1}^i)^-}x_i(t)=\lim_{t\to(t_{k+1}^i+\tau_{k+1}^i)^+}x_i(t)=x_i(t_{k+1}^i+\tau_{k+1}^i)$, which means that $x_i(t)$ is continuous in time instant $t_{k+1}^i+\tau_{k+1}^i$. Thus, $x_i(t)$ is continuous in the time interval $[t_0,\infty)$.

According to the definition of Laplacian matrix L, (1) and (33) can be rewritten as

$$\begin{aligned} \dot{x}_i(t) &= -k\sum_{j=1}^{N}\ell_{ij}x_j(t_k^j)-w_i(t)-\gamma\nabla f_i(x_i(t)),\\ \dot{w}_i(t) &= \alpha\sum_{j=1}^{N}\ell_{ij}x_j(t_k^j),\quad w_i(0)=0,\\ & \quad t\in[t_k^i+\tau_k^i, t_{k+1}^i+\tau_{k+1}^i),\quad k\ge 0. \end{aligned}\tag{34}$$

Let $\hat{\boldsymbol{x}}(t)=\mathrm{col}(x_1(t_k^1),x_2(t_k^2),\ldots,x_N(t_k^N))$, then system (34) can be expressed as the following compact form:

$$\begin{aligned} \dot{\boldsymbol{x}}(t) &= -k(L\otimes I_m)\hat{\boldsymbol{x}}(t)-\boldsymbol{w}(t)-\gamma\nabla\overline{f}(\boldsymbol{x}(t)),\\ \dot{\boldsymbol{w}}(t) &= \alpha(L\otimes I_m)\hat{\boldsymbol{x}}(t). \end{aligned}\tag{35}$$

Similarly, we can obtain the equilibrium point $\boldsymbol{x}^*=\mathbf{1}_N\otimes x^*$, $\boldsymbol{w}^*=-\gamma\nabla\overline{f}(\boldsymbol{x}^*)$, where $x^*\in\mathcal{R}^m$ is the optimal solution of the optimization problem (3).

Using transformation (12) and $\hat{\overline{\boldsymbol{x}}}(t)=\hat{\boldsymbol{x}}(t)-\boldsymbol{x}^*$, one can shift the equilibrium point into the origin, then system (35) can be transformed into the following form:

$$\begin{aligned}\dot{\hat{\bar{\boldsymbol{x}}}}(t) &= -k(L\otimes I_m)\hat{\bar{\boldsymbol{x}}}(t) - \overline{\boldsymbol{w}}(t) - \gamma\Psi(\bar{\boldsymbol{x}}(t)),\\ \dot{\overline{\boldsymbol{w}}}(t) &= \alpha(L\otimes I_m)\hat{\bar{\boldsymbol{x}}}(t),\end{aligned}\tag{36}$$

Let

$$e(t) = (T^T\otimes I_m)\bar{\boldsymbol{x}}(t),\quad \hat{e}(t) = (T^T\otimes I_m)\hat{\bar{\boldsymbol{x}}}(t),$$

$$\vartheta(t) = (T^T\otimes I_m)\overline{\boldsymbol{w}}(t),\quad T = \left[\frac{\mathbf{1}_N}{\sqrt{N}}\quad Q\right].$$

Denote $e = \mathrm{col}(e_1, e_2)$, $\hat{e} = \mathrm{col}(\hat{e}_1, \hat{e}_2)$ and $\vartheta = \mathrm{col}(\vartheta_1, \vartheta_2)$ with $e_1, \hat{e}_1, \vartheta_1 \in \mathcal{R}^m$, and $e_2, \hat{e}_2, \vartheta_2 \in \mathcal{R}^{m(N-1)}$. By the structure of T and (6), we can know T is an orthogonal matrix. Then system (36) can be rewritten as:

$$\begin{aligned}\dot{e}_1(t) &= -\gamma\left(\frac{\mathbf{1}_N^T}{\sqrt{N}}\otimes I_m\right)\Psi(\bar{\boldsymbol{x}}(t)),\\ \dot{e}_2(t) &= -k(H\otimes I_m)\hat{e}_2(t) - \vartheta_2(t) - \gamma(Q^T\otimes I_m)\Psi(\bar{\boldsymbol{x}}(t)),\\ \dot{\vartheta}_1(t) &= 0,\\ \dot{\vartheta}_2(t) &= \alpha(H\otimes I_m)\hat{e}_2(t).\end{aligned}\tag{37}$$

Let $\varepsilon(t) = \mathrm{col}(e_2(t), \vartheta_2(t)), \hat{\varepsilon}(t) = \mathrm{col}(\hat{e}_2(t), \hat{\vartheta}_2(t))$, and $\tilde{\varepsilon}(t) = \varepsilon(t) - \hat{\varepsilon}(t)$, then

$$\dot{\varepsilon}(t) = C\varepsilon(t) - E\tilde{\varepsilon}(t) + F(t),\tag{38}$$

with

$$C = \begin{pmatrix}-kH & -I_{N-1}\\ \alpha H & 0\end{pmatrix}\otimes I_m,\quad E = \begin{pmatrix}-kH & 0\\ \alpha H & 0\end{pmatrix}\otimes I_m,$$

and

$$F(t) = \begin{pmatrix}-\gamma(Q^T\otimes I_m)\Psi(\bar{\boldsymbol{x}}(t))\\ 0\end{pmatrix}.$$

Theorem 2. *Suppose Assumptions 1 and 2 hold, the optimization problem (3) for multi-agent system (1) can be solved by the optimization control (33), if the following conditions are satisfied:*

$$\underline{m} - (4m+1)\gamma\overline{\rho}^2 > 0,\tag{39}$$

$$\underline{\lambda}_1 - 6m(1 + 2k^2\overline{\lambda}_2) - \frac{1}{2} > 0,\tag{40}$$

and

$$T_{\max} + \tau < \sqrt{\frac{m}{\overline{\lambda}_2[(k-\alpha)^2 + 6mk^2]}},\tag{41}$$

where $m > 0$ is a constant, $\underline{\lambda}_1 = \lambda_{\min}(R)$, $\overline{\lambda}_2 = \lambda_{\max}(H^TH)$, and $R = \begin{pmatrix}(k-\alpha)(H^T+H) & I_{N-1}\\ I_{N-1} & 2I_{N-1}\end{pmatrix}\otimes I_m$.

Proof. Consider Lyapunov function $V_1(t)$ given in Theorem 1, the derivation of V_1 along the first equality of (37) and system (38) yields:

$$\begin{aligned}\dot{V}_1 &= e_1^T(t)\dot{e}_1(t) + \frac{1}{2}\varepsilon^T(t)(\Omega C + C^T\Omega)\varepsilon(t)\\ &\quad - \varepsilon^T(t)\Omega E\tilde{\varepsilon}(t) + \varepsilon^T(t)\Omega F(t),\end{aligned}\tag{42}$$

Let $R = -(\Omega C + C^T\Omega)$, we obtain that R is positive definite and $k > \alpha$. Then, we have

$$\begin{aligned}-\varepsilon^T(t)\Omega E\tilde{\varepsilon}(t) &= (k-\alpha)e_2^T(t)(H\otimes I_m)\tilde{e}_2(t)\\ &\le \frac{1}{4}e_2^T(t)e_2(t) + (k-\alpha)^2\tilde{e}_2^T(t)(H^TH\otimes I_m)\tilde{e}_2(t)\\ &\le \frac{1}{4}e_2^T(t)e_2(t) + (k-\alpha)^2\overline{\lambda}_2\tilde{e}_2^T(t)\tilde{e}_2(t),\end{aligned}\tag{43}$$

where $\tilde{e}_2(t) = e_2(t) - \hat{e}_2(t) = (Q^T\otimes I_m)(\bar{\boldsymbol{x}}(t) - \hat{\bar{\boldsymbol{x}}}(t))$, and

$$\begin{aligned}\bar{\boldsymbol{x}}(t) - \hat{\bar{\boldsymbol{x}}}(t) &= \mathrm{col}[\bar{x}_1(t) - \bar{x}_1(t_k^1), \bar{x}_2(t) - \bar{x}_2(t_k^2), \ldots, \bar{x}_N(t) - \bar{x}_N(t_k^N)]\\ &= \mathrm{col}\left[\int_{t_k^1}^t \dot{\bar{x}}_1(s)ds, \int_{t_k^2}^t \dot{\bar{x}}_2(s)ds, \ldots, \int_{t_k^N}^t \dot{\bar{x}}_N(s)ds\right].\end{aligned}\tag{44}$$

Combining (20), (42) and (43), we have

$$\begin{aligned}\dot{V}_1(t) &\le -\gamma\underline{m}\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t) + \frac{1}{4}\vartheta_2^T(t)\vartheta_2(t) + \gamma^2\overline{\rho}^2\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t)\\ &\quad - \frac{1}{2}\varepsilon^T(t)R\varepsilon(t)\\ &\quad + \frac{1}{4}e_2^T(t)e_2(t) + (k-\alpha)^2\overline{\lambda}_2\tilde{e}_2^T(t)\tilde{e}_2(t)\\ &\le -\gamma\underline{m}\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t) + \frac{1}{4}\varepsilon^T(t)\varepsilon(t) + \gamma^2\overline{\rho}^2\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t)\\ &\quad - \frac{1}{2}\underline{\lambda}_1\varepsilon^T(t)\varepsilon(t) + (k-\alpha)^2\overline{\lambda}_2\tilde{e}_2^T(t)\tilde{e}_2(t).\end{aligned}\tag{45}$$

Construct the following auxiliary integral function

$$V_2(t) = \int_{t-T_{\max}-\tau}^t\int_\theta^t \dot{e}^T(s)\dot{e}(s)dsd\sigma,$$

we can obtain

$$\dot{V}_2(t) = (T_{\max}+\tau)\dot{e}^T(t)\dot{e}(t) - \int_{t-T_{\max}-\tau}^t \dot{e}^T(s)\dot{e}(s)ds,$$

where

$$\dot{e}^T(t)\dot{e}(t) = \dot{e}_1^T(t)\dot{e}_1(t) + \dot{e}_2^T(t)\dot{e}_2(t),$$

with

$$\dot{e}_1^T(t)\dot{e}_1(t) = \gamma^2\left\|\left(\frac{\mathbf{1}_N^T}{\sqrt{N}}\otimes I_m\right)\Psi(\bar{\boldsymbol{x}}(t))\right\|^2 \le \gamma^2\overline{\rho}^2\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t),$$

and

$$\begin{aligned}\dot{e}_2^T(t)\dot{e}_2(t) &= k^2\hat{e}_2^T(t)(H^TH\otimes I_m)\hat{e}_2(t) + \vartheta_2^T(t)\vartheta_2(t)\\ &\quad + \gamma^2\|(Q^T\otimes I_m)\Psi(\bar{\boldsymbol{x}}(t))\|^2\\ &\quad + 2k\hat{e}_2^T(t)(H^T\otimes I_m)\vartheta_2(t)\\ &\quad + 2k\gamma\hat{e}_2^T(t)(H^T\otimes I_m)(Q^T\otimes I_m)\Psi(\bar{\boldsymbol{x}}(t))\\ &\quad + 2\gamma\vartheta_2^T(t)(Q^T\otimes I_m)\Psi(\bar{\boldsymbol{x}}(t)).\end{aligned}\tag{46}$$

Similarly to (24)–(27), and due to the fact that

$$\begin{aligned}\hat{e}_2^T(t)\hat{e}_2(t) &= (e_2(t) - \tilde{e}_2(t))^T(e_2(t) - \tilde{e}_2(t)) \le 2(e_2^T(t)e_2(t)\\ &\quad + \tilde{e}_2^T(t)\tilde{e}_2(t)),\end{aligned}$$

we have

$$\begin{aligned}\dot{e}_2^T(t)\dot{e}_2(t) &\le 3[k^2\hat{e}_2^T(t)(H^TH\otimes I_m)\hat{e}_2(t) + \vartheta_2^T(t)\vartheta_2(t)\\ &\quad + \gamma^2\|(Q^T\otimes I_m)\Psi(\bar{\boldsymbol{x}}(t))\|^2]\\ &\le 3[k^2\overline{\lambda}_2\hat{e}_2^T(t)\hat{e}_2(t) + \vartheta_2^T(t)\vartheta_2(t) + \gamma^2\overline{\rho}^2\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t)]\\ &\le 6k^2\overline{\lambda}_2\tilde{e}_2^T(t)\tilde{e}_2(t) + 6k^2\overline{\lambda}_2e_2^T(t)e_2(t) + 3\vartheta_2^T(t)\vartheta_2(t)\\ &\quad + 3\gamma^2\overline{\rho}^2\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t)\\ &\le 6k^2\overline{\lambda}_2\tilde{e}_2^T(t)\tilde{e}_2(t) + 6k^2\overline{\lambda}_2\varepsilon^T(t)\varepsilon(t) + 3\varepsilon^T(t)\varepsilon(t)\\ &\quad + 3\gamma^2\overline{\rho}^2\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t).\end{aligned}\tag{47}$$

Let $V(t) = V_1(t) + \frac{m}{T_{\max}+\tau}V_2(t)$, then

$$\begin{aligned}\dot{V}(t) &\le -\gamma\underline{m}\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t) + \frac{1}{4}\varepsilon^T(t)\varepsilon(t) + \gamma^2\overline{\rho}^2\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t) - \frac{1}{2}\underline{\lambda}_1\varepsilon^T(t)\varepsilon(t)\\ &\quad + [(k-\alpha)^2 + 6mk^2]\overline{\lambda}_2\tilde{e}_2^T(t)\tilde{e}_2(t) + 6mk^2\overline{\lambda}_2\varepsilon^T(t)\varepsilon(t)\\ &\quad + 3m\varepsilon^T(t)\varepsilon(t) + 4m\gamma^2\overline{\rho}^2\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t)\\ &\quad - \frac{m}{T_{\max}+\tau}\int_{t-T_{\max}-\tau}^t \dot{e}^T(s)\dot{e}(s)ds.\end{aligned}\tag{48}$$

Recalling that $\tilde{e}_2(t) = (Q^T \otimes I_m)(\bar{\boldsymbol{x}}(t) - \hat{\bar{\boldsymbol{x}}}(t))$ and (44), it follows that

$$\begin{aligned}\tilde{e}_2^T(t)\tilde{e}_2(t) &= \|(Q^T \otimes I_m)(\bar{\boldsymbol{x}}(t) - \hat{\bar{\boldsymbol{x}}}(t))\|^2 \\ &\le \|\bar{\boldsymbol{x}}(t) - \hat{\bar{\boldsymbol{x}}}(t)\|^2 \\ &= \sum_{i=1}^{N}\left(\int_{t_k^i}^{t}\dot{\bar{x}}_i(s)ds\right)^T\left(\int_{t_k^i}^{t}\dot{\bar{x}}_i(s)ds\right). \end{aligned} \tag{49}$$

From Lemma 3, we have

$$\left(\int_{t_k^i}^{t}\dot{\bar{x}}_i(s)ds\right)^T\left(\int_{t_k^i}^{t}\dot{\bar{x}}_i(s)ds\right) \le (t - t_k^i)\int_{t_k^i}^{t}\dot{\bar{x}}_i^T(s)\dot{\bar{x}}_i(s)ds,$$

where $t \in [t_k^i + \tau_k^i, t_{k+1}^i + \tau_{k+1}^i)$, $t - t_k^i \le T_{\max} + \tau$, that is $t - T_{\max} - \tau \le t_k^i$, and $\dot{\bar{x}}^T(t)\dot{\bar{x}}(t) = \dot{e}^T(t)\dot{e}(t)$.

Then, we have

$$\begin{aligned}\tilde{e}_2^T(t)\tilde{e}_2(t) &\le \sum_{i=1}^{N}(t - t_k^i)\int_{t_k^i}^{t}\dot{\bar{x}}_i^T(s)\dot{\bar{x}}_i(s)ds \\ &\le \sum_{i=1}^{N}(T_{\max} + \tau)\int_{t-T_{\max}-\tau}^{t}\dot{\bar{x}}_i^T(s)\dot{\bar{x}}_i(s)ds \\ &= (T_{\max} + \tau)\int_{t-T_{\max}-\tau}^{t}\sum_{i=1}^{N}\dot{\bar{x}}_i^T(s)\dot{\bar{x}}_i(s)ds \\ &= (T_{\max} + \tau)\int_{t-T_{\max}-\tau}^{t}\dot{\bar{x}}^T(s)\dot{\bar{x}}(s)ds \\ &= (T_{\max} + \tau)\int_{t-T_{\max}-\tau}^{t}\dot{e}^T(s)\dot{e}(s)ds. \end{aligned} \tag{50}$$

From (48) and (50), we have

$$\begin{aligned}\dot{V}(t) \le& -\gamma\underline{m}\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t) + \frac{1}{4}\varepsilon^T(t)\varepsilon(t) + \gamma^2\bar{\rho}^2\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t) \\ &- \frac{1}{2}\underline{\lambda}_1\varepsilon^T(t)\varepsilon(t) \\ &+ [(k-\alpha)^2 + 6mk^2]\bar{\lambda}_2(T_{\max} + \tau)\int_{t-T_{\max}-\tau}^{t}\dot{e}^T(s)\dot{e}(s)ds \\ &+ 6mk^2\bar{\lambda}_2\varepsilon^T(t)\varepsilon(t) + 3m\varepsilon^T(t)\varepsilon(t) + 4m\gamma^2\bar{\rho}^2\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t) \\ &- \frac{m}{T_{\max} + \tau}\int_{t-T_{\max}-\tau}^{t}\dot{e}^T(s)\dot{e}(s)ds, \end{aligned} \tag{51}$$

and then

$$\begin{aligned}\dot{V}(t) \le& -[\gamma\underline{m} - (4m+1)\gamma^2\bar{\rho}^2]\bar{\boldsymbol{x}}^T(t)\bar{\boldsymbol{x}}(t) \\ &- \left[\frac{1}{2}\underline{\lambda}_1 - 3m(1 + 2k^2\bar{\lambda}_2) - \frac{1}{4}\right]\varepsilon^T(t)\varepsilon(t) \\ &- \left[\frac{m}{T_{\max} + \tau} - ((k-\alpha)^2 + 6mk^2)\bar{\lambda}_2(T_{\max} + \tau)\right] \\ &\times \int_{t-T_{\max}-\tau}^{t}\dot{e}^T(s)\dot{e}(s)ds. \end{aligned} \tag{52}$$

By the similar analysis as the proof of Theorem 1, we can conclude that $\dot{V}(t) < 0$, which completes the proof. □

4. Simulations

In this section, we give an example to validate our theoretical results. In this example, we consider a multi-agent system consisting of five agents. Supposes that the interconnected topology is described as in Fig. 1. The weight of every edge is 1.

Consider the following optimization problem

$$\text{minimize} \quad f(x) = \sum_{i=1}^{N} f_i(x), x \in \mathcal{R},$$

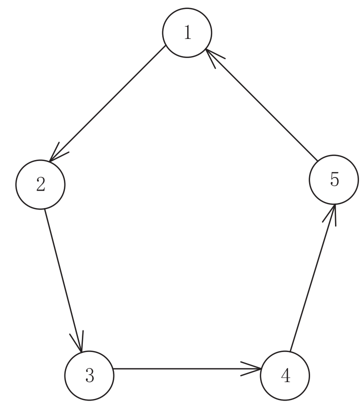


Fig. 1. Connected graph.

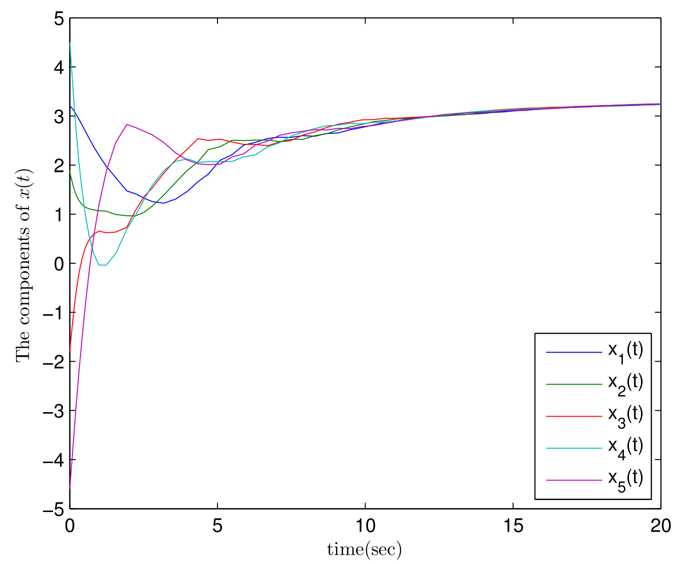


Fig. 2. The trajectories of $x_i(t)$ with $T_k = 0.01$, $\tau_k = 0.07$.

where the local objective function is given as following

$$\begin{aligned} f_1(x) &= 0.7(x-6)^2, \\ f_2(x) &= (x-4)^2, \\ f_3(x) &= \frac{x^2}{\ln(x^2+2)}, \\ f_4(x) &= \sin\frac{x}{2} + \frac{x^2}{4}, \\ f_5(x) &= \frac{x^2}{\sqrt{x^2+1}} + 0.2x^2. \end{aligned} \tag{53}$$

Obviously, for $i = 1, 2, \ldots, 5$, f_i is differentiable and satisfies Assumption 1. Choosing $\alpha = 0.6, k = 1.0, \gamma = 0.2$, we can obtain $\bar{\rho} = 2, \underline{m} = 1, \underline{\lambda}_1 = 0.6861, \bar{\lambda}_2 = 3.6180$. Let the initial values

$$x(0) = [x_1(0), x_2(0), x_3(0), x_4(0), x_5(0)]^T = [-3.2, 1.9, -1.8, 4.5, -4.6]^T,$$

and

$$w(0) = [w_1(0), w_2(0), w_3(0), w_4(0), w_5(0)]^T = [0, 0, 0, 0, 0]^T.$$

(1) synchronous sampling: The sampling interval T_k and the time-delay τ_k are given as $T_k = 0.01s$, $\tau_k = 0.07$. The simulation results are shown in Figs. 2 and 3.

(2) asynchronous sampling: The sampling interval is given as $T = 0.01s$, time-delay $\tau_k^i (i = 1, 2, \ldots, 5)$ are simulated by random numbers in the interval $[0, 0.3T]$. The simulation results are shown in Figs. 4 and 5.

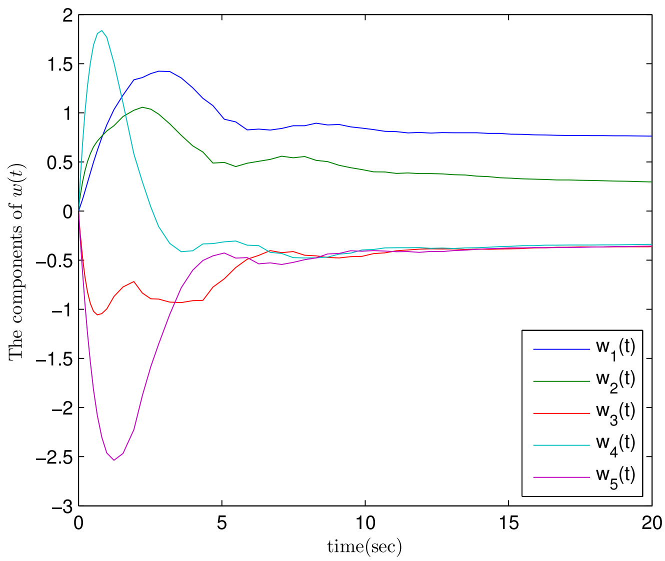


Fig. 3. The trajectories of $w_i(t)$ with $T_k = 0.01$, $\tau_k = 0.07$.

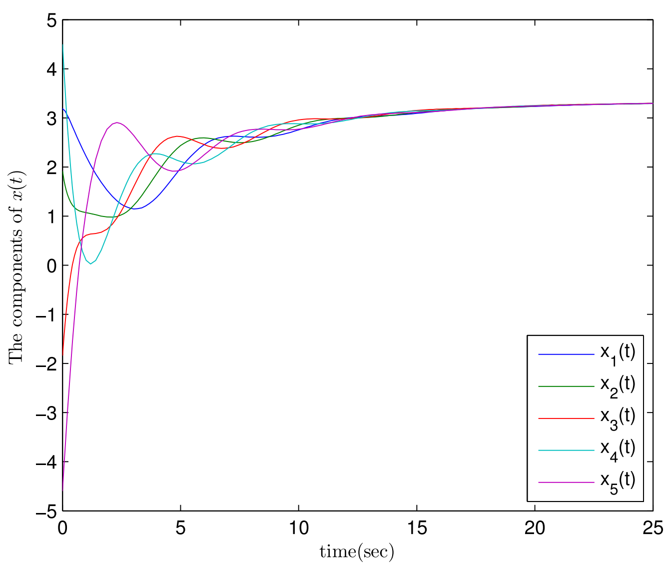


Fig. 4. The trajectories of $x_i(t)$ with $T = 0.01$, $\tau_k^i \in [0, 0.3T]$.

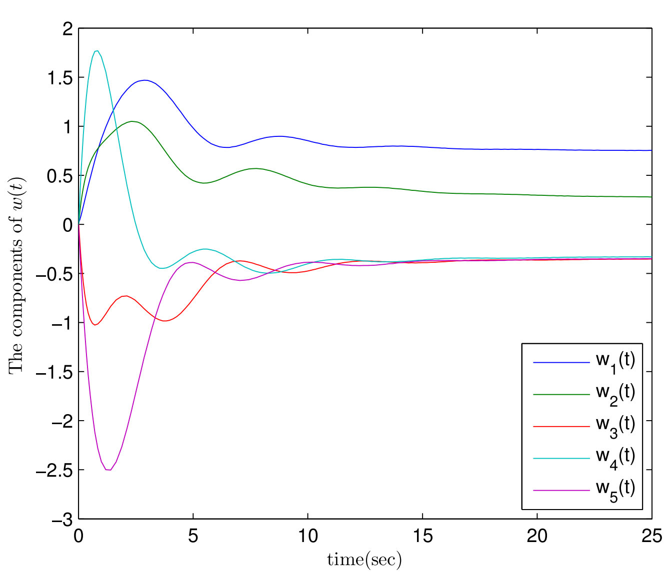


Fig. 5. The trajectories of $w_i(t)$ with $T = 0.01$, $\tau_k^i \in [0, 0.3T]$.

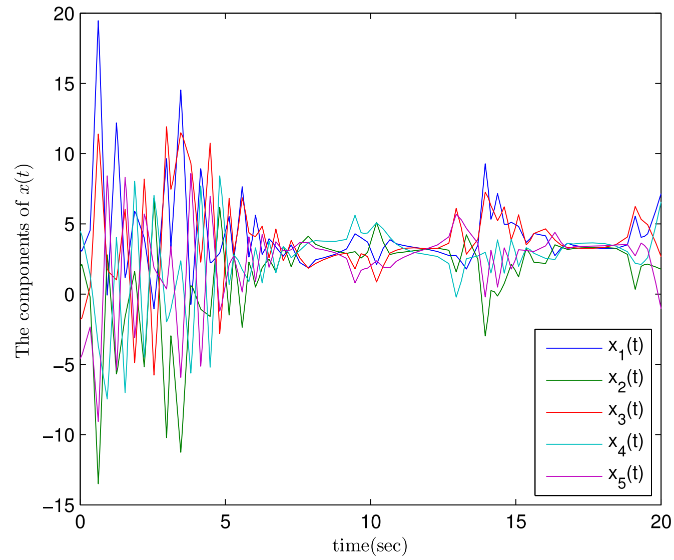


Fig. 6. The trajectories of $x_i(t)$ with $T_0 = 0.09$.

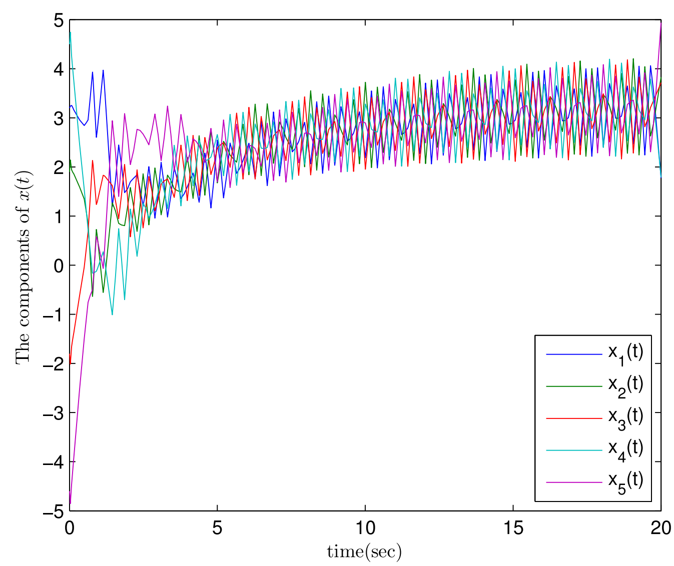


Fig. 7. The trajectories of $x_i(t)$ with $T_1 = 0.07$.

We can see that the trajectories x_i of each agent i converge to the global optimal solution $x^* = 3.1798$ of the objective function $f(x) = \sum_{i=1}^{N} f_i(x)$ and all the trajectories w_i converge to a constant, respectively, for $i = 1, 2, \ldots, 5$. The optimal value of $f(x)$ is 18.8773.

The simulation result for synchronous sampling with $T_0 = 0.09$ is depicted in Fig. 6 and asynchronous sampling with $T_1 = 0.07$ is depicted in Fig. 7, respectively, where T_0 and T_1 are larger than the upper bound $T_{\max}$ in Theorems 1 and2. Then $x_i(t)$ is not convergent.

5. Conclusion

In this paper, a distributed optimization problem of multi-agent systems with delayed sampled-data is considered. The interconnected topology is assumed to be directed, weighted-balanced and strongly connected, and the local costs are strongly convex with globally Lipschitz gradients. Two control algorithms using sampled-data with time-delay under synchronous and asynchronous sampling are presented for the multi-agent systems to reach consensus and optimal state. Based on Lyapunov theory and algebraic graph theory. Sufficient conditions are obtained to make all the agents converge to the optimal solution of the system if the design parameters are chosen properly. Finally, numerical example are given to illustrate the theoretical results.

Acknowledgments

This work is supported in part by National Natural Science Foundation (NNSF) of China (61273183, 61374028).

References

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learn. 3 (2010) 1–122.
[2] T.H. Chang, A. Nedic, A. Scaglione, Distributed constrained optimization by consensus-based primal-dual perturbation method, IEEE Trans. Autom. Control 59 (2014) 1524–1538.
[3] P. Yi, Y. Hong, F. Liu, Distributed gradient algorithm for constrained optimization with application to load sharing in power systems, Syst. Control Lett. 83 (2015) 45–52.
[4] M. Rabbat, R. Nowak, Distributed optimization in sensor networks, in: Proceedings of the Third International Symposium on Information Processing in Sensor Networks, 2004, pp. 20–27.
[5] A. Jadbabaie, J. Lin, A.S. Morse, Coordination of groups of mobile autonomous agents using nearest neighbor rules, IEEE Trans. Autom. Control 48 (2003) 988–1001.
[6] R. Olfati-Saber, R.M. Murray, Consensus problems in networks of agents with switching topology and time-delays, IEEE Trans. Autom. Control 49 (2004) 1520–1533.
[7] Y. Hatano, M. Mesbahi, Agreement over random networks, IEEE Trans. Autom. Control 50 (2005) 1867–1872.
[8] W. Ren, R.W. Beard, Consensus seeking in multiagent systems under dynamically changing interaction topologies, IEEE Trans. Autom. Control 50 (2005) 655–661.
[9] Y. Hong, J. Hu, L. Gao, Tracking control for multi-agent consensus with an active leader and variable topology, Automatica 42 (2006) 1177–1182.
[10] L. Wang, F. Xiao, Finite-time consensus problems for networks of dynamic agents, IEEE Trans. Autom. Control 55 (2010) 950–955.
[11] H. Yu, X. Xia, Adaptive consensus of multi-agents in networks with jointly connected topologies, Automatica 48 (2012) 1783–1790.
[12] F. Xiao, L. Wang, T. Chen, Finite-time consensus in networks of integrator-like dynamic agents with directional link failure, IEEE Trans. Autom. Control 59 (2014) 756–762.
[13] X. Zhang, Q. Han, Robust H_∞ filtering for a class of uncertain linear systems with time-varying delay, Automatica 44 (2008) 157–166.
[14] X. Zhang, Q. Han, Network-based H_∞ filtering using a logic jumping-like trigger, Automatica 49 (2013) 1428–1435.
[15] X. Zhang, Q. Han, Novel delay-derivative-dependent stability criteria using new bounding techniques, Int. J. Robust Nonlinear Control 23 (2013) 1419–1432.
[16] P. Lin, Y. Jia, Consensus of second-order discrete-time multi-agent systems with nonuniform time-delays and dynamically changing topologies, Automatica 45 (2009) 2154–2158.
[17] W. Zhu, D. Cheng, Leader-following consensus of second-order agents with multiple time-varying delays, Automatica 46 (2010) 1994–1999.
[18] Y. Qian, X. Wu, Consensus of second-order multi-agent systems with nonlinear dynamics and time delay, Nonlinear Dyn 78 (2014) 495–503.
[19] W. He, F. Qian, J. Lam, G. Chen, Q. Han, J. Kurths, Quasi-synchronization of heterogeneous dynamic networks via distributed impulsive control: error estimation, optimization and design, Automatica 62 (2015) 249–262.
[20] W. He, G. Chen, Q. Han, F. Qian, Network-based leader-following consensus of nonlinear multi-agent systems via distributed impulsive control, Inf. Sci. 380 (2017) 145–158.
[21] W. He, B. Zhang, Q.L. Han, F. Qian, J. Kurths, J. Cao, Leader-following consensus of nonlinear multiagent systems with stochastic sampling, IEEE Trans. Cybern. 47 (2017) 327–338.
[22] E. Fridman, A refined input delay approach to sampled-data control, Automatica 46 (2010) 421–427.
[23] G. Wen, Z. Duan, W. Yu, G. Chen, Consensus of multi-agent systems with nonlinear dynamics and sampled-data information: a delayed-input approach, Int. J. Robust Nonlinear Control 23 (2013) 602–619.
[24] Y. Wang, H. Zhou, Z. Wang, Z.-W. Liu, W. Hu, Sampled containment control for multi-agent systems with nonlinear dynamics, Neurocomputing 219 (2017) 242–247.
[25] H. Su, Z. Wang, Z. Song, X. Chen, Event-triggered consensus of non-linear multi-agent systems with sampling data and time delay, IET Control Theory Appl. 11 (2017) 1715–1725.
[26] F. Xiao, Y. Shi, W. Ren, Robustness analysis of asynchronous sampled-data multi-agent networks with time-varying delays, IEEE Trans. Autom. Control (2017) 1, doi:10.1109/TAC.2017.2756860.
[27] X. Ge, Q. Han, D. Ding, X. Zhang, B. Ning, A survey on recent advances in distributed sampled-data cooperative control of multi-agent systems, Neurocomputing 275 (2018) 1684–2312.
[28] L. Ding, Q. Han, X. Ge, X. Zhang, An overview of recent advances in event-triggered consensus of multiagent systems, IEEE Trans. Cybern. (2017) 1–14, doi:10.1109/TCYB.2017.2771560.
[29] A. Nedic, A. Ozdaglar, Distributed subgradient methods for multi-agent optimization, IEEE Trans. Autom. Control 54 (2009) 48–61.
[30] I. Lobel, A. Ozdaglar, Distributed subgradient methods for convex optimization over random networks, IEEE Trans. Autom. Control 56 (2011) 1291–1306.
[31] M. Zhu, S. Martinez, On distributed convex optimization under inequality and equality constraints, IEEE Trans. Autom. Control 57 (2012) 151–164.
[32] B. Gharesifard, J. Corts, Distributed continuous-time convex optimization on weight-balanced digraphs, IEEE Trans. Autom. Control 59 (2014) 781–786.
[33] S.S. Kia, J. Cortés, S. Martínez, Distributed convex optimization via continuous–time coordination algorithms with discrete-time communication, Automatica 55 (2015) 254–264.
[34] S. Yang, Q. Liu, J. Wang, Distributed optimization based on a multiagent system in the presence of communication delays, IEEE Trans. Syst. Man Cybern. Syst. 47 (2017) 717–728.
[35] J. Lu, C.Y. Tang, Zero-gradient-sum algorithms for distributed convex optimization: the continuous-time case, IEEE Trans. Autom. Control 57 (2012) 2348–2354.
[36] W. Chen, W. Ren, Event-triggered zero-gradient-sum distributed consensus optimization over directed networks, Automatica 65 (2016) 90–97.
[37] J. Liu, W. Chen, H. Dai, Sampled-data based distributed convex optimization with event-triggered communication, Int. J. Control Autom. Syst. 14 (2016) 1421–1429.
[38] K. Liu, Z. Ji, Periodic event-triggered consensus of time-delayed multi-agent systems under switching topologies, in: Proceedings of the Thirty Sixth Chinese Control Conference (CCC), 2017, pp. 8230–8235.
[39] F. Chen, H. Yu, X. Xia, Output consensus of multi-agent systems with delayed and sampled-data, IET Control Theory Appl. 11 (2017) 632–639.
[40] C.D. Godsil, G. Royle, Algebraic Graph Theory, Springer, New York, 2001.
[41] K. Gu, V.L. Kharitonov, J. Chen, Stability of Time-Delay Systems, Springer, New York, NY, USA, 2003.
[42] R.T. Rockafellar, Convex Analysis, Princeton University, New Jersey, 1972.

Junxiu Yan received her B.S. degree in Mathematics from Changzhi University, in 2011, and now she is a master degree candidate in Mathematics in China Three Gorges University. Her research interests include complex dynamical network, distributed optimization, and multi-agent systems.

Hui Yu received his bachelor's degree in mathematics from Central China Normal University, Wuhan, China in 1991, his master's degree in system analysis and integration from East China Normal University, Shanghai, China in 1999, and his Ph.D. degree in Automatic control from Huazhong University of Science and Technology, Wuhan, China in 2007. From 2011 to 2012, he was visiting the Centre of New Energy Systems, Department of Electrical, Electronics and Computer Engineering, University of Pretoria, South Africa. Since 1991, he has been with School of Science, China Three Gorges University, where he is currently a professor. His research interests include nonlinear control theory with applications to robots, multi-agent systems, energy efficiency.

Xiaohua Xia is a professor in the Electrical, Electronic and Computer Engineering at the University of Pretoria, South Africa, director of the Centre of New Energy Systems, and the director of the National Hub for the Postgraduate Programme in Energy Efficiency and Demand-side Management. He was academically affiliated with the University of Stuttgart, Germany, the Ecole Centrale de Nantes, France, and the National University of Singapore before joining the University of Pretoria in 1998. Prof. Xia is a fellow of the Institute for Electronic and Electrical Engineers (IEEE), a fellow of the South African Academy of Engineering (SAAE), a member of the Academy of Science of South Africa (ASSAF), and he has an A rating from the South African National Research Foundation (NRF). He has been an associate editor of Automatica, IEEE Transactions on Circuits and Systems II, IEEE Transactions on Automatic Control, and specialist editor (control) of the SAIEE Africa Research Journal. His research interests are modeling and control of nonlinear systems, energy systems and optimization.

Applied Energy 222 (2018) 244–256

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Optimized response to electricity time-of-use tariff of a compressed natural gas fuelling station ☆

Charles Kagiri, Evan M. Wanjiru, Lijun Zhang*, Xiaohua Xia

Centre of New Energy Systems, Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa

HIGHLIGHTS

- A model for optimization of compressor scheduling based on demand is presented.
- 59.3% cost reduction for a studied gas station fuelling 143 vehicles over a high demand season.
- 25% cost reduction a studied gas station fuelling 146 vehicles over a low demand season.
- Compressor switching frequency is minimized by different strategies.

ARTICLE INFO

Keywords:
Compressed natural gas
Fast-fill
Demand response
Optimization

ABSTRACT

Compressed natural gas propulsion of vehicles has been shown to have advantages over petrol and diesel propulsion due to lower carbon dioxide emissions as well as the increased durability of vehicle engines. The growth of compressed natural gas as an alternative fuel to petrol and diesel can be accelerated by implementing strategies that result in the economical operation of the distribution infrastructure. Economic scheduling of power consumption is a useful strategy for reducing the cost of energy for both industrial and domestic consumers who operate in time-of-use based electricity pricing environments. In this paper, an optimal energy management strategy is proposed for the operation of a compressed natural gas fuelling station. The compressor energy consumption, being the main component of the total operating cost of the fuelling station, presents a cost saving opportunity through which optimal scheduling of operation can be used to lower cost of operation of the station. The developed model shows potential average savings of 59.3% in daily electricity costs while maximizing compressor life through minimization of compressor cycling.

1. Introduction

Compressed Natural Gas (CNG) is one of the alternatives to liquid hydrocarbon fuels that have been promoted to address the challenges of air pollution, energy dependence and climate change [1–4]. CNG, which is largely made up of methane and small quantities of other hydrocarbons such as propane butane and ethane could be considered a clean fuel in comparison with gasoline and diesel because it has the lowest emissions among hydrocarbon fuels [5,6]. Furthermore, CNG vehicles have been shown to have lower total cost of ownership (TCO) than gasoline or diesel fuelled cars [7–9]. In recent years, the use of CNG for vehicle propulsion has been increasing worldwide in both developed and developing countries, especially in countries that have suffered severe air pollution from rapid industrialization in the past three decades such as India and China [10,11]. The expanding adoption of CNG has corresponded to a simultaneous growth of CNG distribution infrastructure for vehicular end users [12]. In South Africa for example, the Department of Energy recognizes compressed natural gas as one of the possible energy options for transportation that will contribute to the reduction of the country's carbon footprint [13]. In view of the long way to large-scale adoption of electric vehicles, CNG is viewed as an appropriate transition fuel towards a greener transportation sector [14]. Introduction of public service vehicles powered by CNG as well as growth in the number of CNG fleet customers has resulted in the growth in number of vehicle fuelling station in the city of Johannesburg and Pretoria. Being consumers of electric power, CNG fuelling stations are subject to the availability and pricing conditions of the electricity environment in which they operate [15]. While the expansion of CNG distribution infrastructure is a sign of investor confidence in the future of the industry, the distribution infrastructure is subject to challenges

☆ The short version of the paper was presented at ICAE2017, Aug 21–24, Cardiff, UK. This paper is a substantial extension of the short version of the conference paper.

* Corresponding author.
E-mail addresses: kagiricharles@gmail.com (C. Kagiri), lijun.zhang@up.ac.za (L. Zhang).

https://doi.org/10.1016/j.apenergy.2018.04.017
Received 31 December 2017; Received in revised form 3 April 2018; Accepted 5 April 2018

Nomenclature

J	objective function (currency)
Mw_a	molecular weight of the air (g)
Mw_g	molecular weight of the gas (g)
$m_{hp}^{max}, m_{mp}^{max}, m_{lp}^{max}$	maximum mass for high pressure, medium pres sure and low pressure reservoirs (kg)
$m_{hp}^{min}, m_{mp}^{min}, m_{lp}^{min}$	minimum mass for high pressure, medium pressure and low pressure reservoirs (kg)
$m_{ohp}^{max}, m_{omp}^{max}, m_{olp}^{max}$	mass demand from high pressure, medium pres sure and low pressure reservoirs (kg)
$\dot{m}_{co}$	compressor outlet mass flow rate (kg/h)
N	total samples over the control horizon
n	gas quantity (moles)
p	pressure (bars)
P_{co}	compressor motor power rating (kW)
p_e	price of electricity under TOU tariff (currency/kW h)
$p_{hp}^{max}, p_{mp}^{max}, p_{lp}^{max}$	maximum pressure for high pressure, medium pressure and low pressure reservoirs (bars)
$p_{hp}^{min}, p_{mp}^{min}, p_{lp}^{min}$	minimum pressure for high pressure, medium pressure and low pressure reservoirs (bars)
Q_{std}	capacity of the compressor under standard conditions (N m^3/h)
R	universal gas constant (L bar/K mol)
T	absolute temperature (K)
u	state of compressor switch
u_{hp}, u_{mp}, u_{lp}	state of reservoir valves for high pressure, medium pressure and low pressure reservoirs
V	volume of cascade reservoir tanks (L)
z	compressibility factor of CNG
$\rho_{a,std}$	density of air under standard conditions (kg/m^3)

arising from the supply of electricity and must implement adaptive strategies to remain energy efficient and economically attractive.

CNG is stored under high pressure in on-board vehicle tanks, from where it flows to the combustion engine under regulation [16]. CNG powered vehicles receive their fuel from high pressure reservoirs at CNG fuelling stations. Although refuelling of vehicles with natural gas can take a long time, the fast-fill process which is used at most CNG fuelling stations has been developed to achieve fuelling times of less than five minutes, which is comparable to the fuelling time of diesel or gasoline powered automobiles [17]. Fast-fill fuelling stations use reservoir tanks in a cascaded storage system divided into low pressure, medium pressure and high pressure levels [18]. The dispenser at the fast-fill station has electronic sequencing valves that are controlled by a microprocessor algorithm as well as sensors for measuring mass flow from each of the three reservoirs [19]. A vehicle typically arrives with low pressure in its tank and the dispenser starts the filling of the tank by connecting it to the low pressure reservoir. The differential pressure causes gas to flow into the vehicle tank and as the vehicle tank fills up, the mass flow rate between the reservoir and the vehicle tank falls to a limit after which the dispenser switches the filling to the medium pressure reservoir for a higher mass flow rate [20]. The vehicle tank continues to fill up from the medium pressure reservoir which results in the mass flow rate falling until a limit is reached and the dispenser switches the filling from the medium pressure reservoir to the high pressure reservoir. The high pressure reservoir completes the filling of the vehicle tank [21]. It is possible in some scenarios for the vehicle to arrive with a high tank pressure that can only be filled from the high pressure reservoir since it is almost full or from the medium pressure reservoir followed by the high pressure reservoir. The dispenser algorithm determines from initial vehicle tank pressure which reservoir to start with [22]. It is also possible in some scenarios for the customer to request a quantity of gas that does not result in filling of the vehicle tank and therefore receive gas from the low pressure reservoir and medium pressure reservoir only, or even the low pressure reservoir alone. This means that the demand of gas from one reservoir is not always synchronized with demand from the other two reservoirs. The CNG station dispenser runs a vehicle filling algorithm that is compensated for temperature and pressure to ensure that correct quantities of gas demanded by consumer are dispensed to the vehicle tank. This isolates the consumer vehicle tank from the fluctuations in the pressure and temperature that may occur in the cascade storage as a result of vehicle gas demand itself or as a result of operation control [23–27].

A priority panel controls the filling of the three reservoirs of the cascade storage, by switching the gas flowing from the outlet of the compressor between the reservoir valves [28]. The priority panel is operated through a PLC, which runs an algorithm that controls the sequence of opening the three reservoir valves during charging of the cascade storage by the reciprocating compressor [29]. The compressor is a vital part of the fast-fill operation and is the main contributor to the CNG fueling station's operating cost through its power consumption as well as wear and tear [30]. The sizing of the station compressor and other station components is based on the expected inlet flow rate from the municipal supply line and the quantity of gas expected to be dispensed at the station [31]. Efficient operation of the compressor in a CNG fuelling station presents an opportunity for the reduction of operating costs. The savings that are realised can be passed on to consumers in the form of reduced price of CNG per unit of sale.

Energy efficiency of energy converting systems falls into four general categories of equipment efficiency, technology efficiency, performance efficiency and operation efficiency [32–35]. CNG fuelling station operators, just like other commercial electricity consumers, must make careful consideration for all the four categories of energy efficiency in order to increase the economic performance of these installations [36–39].

Research into the efficient operation of CNG fuelling stations has been greatly aided by the work of previous researchers. Kountz [40] modelled the fast-fill process based on the first law of thermodynamics for gas behaviour between a single reservoir and the on-board vehicle cylinder. Other researchers have expanded the modelling of the fast-fill CNG fueling station by considering the individual components of the station infrastructure and their interaction with the flowing gas. These include [41,42] whose work advanced the thermodynamic modelling of the fast-filling process. The research on further minimization of filling time has been studied by [43]. Using thermodynamic laws and mass balance, [44] studied the effects of initial conditions and ambient temperature on the filling of the vehicle on-board cylinder and achieving of the target pressure. The effects of the connecting pipe on the process of vehicle filling was also studied by [45]. Research on the complete filling of on-board vehicle cylinder through the development of dispenser algorithms for the fast-fill process has been conducted by [46,19]. Research has also been undertaken in relation to the thermodynamic behaviour of the reciprocating compressor in achieving different performance goals for the CNG fuelling station [47–50]. Frick et al. [51] studied the optimization of the distribution of CNG refuelling stations in Switzerland. The study applied cost benefit analysis to determine optimal location of new CNG fuelling stations among the existing petrol filling stations as well as existing CNG filling stations. CNG self-fuelling of vehicles and fuelling from homes has also been the subject of other research towards efficient delivery of the gas [16,52]. The significant effect that domestic refuelling of CNG vehicles in consumer homes could have on the electric power infrastructure has been studied by [53]. Their study recognized limited infrastructure as a major technological barrier to the market penetration of CNG vehicles in the United States of America. This limitation was also shown to result in more consumers using domestic compressor units to fuel their vehicles, with potential effects on the electric grid network.

To the best of the authors' knowledge no research has been done on the implementation of electric load shifting for CNG fuelling stations in order to reduce their electricity costs charged at a time-of-use (TOU) tariff. Further, the simultaneous optimization of the operation of priority panel valves and the compressor for the purpose of achieving electric load scheduling for a CNG fuelling station has not been reported in the literature. This study presents an attempt to apply an optimized operation strategy to CNG fuelling stations in order to secure the benefits of demand response (DR) programs implemented by power utility operators through the time-of-use tariff for both the utility and the fuelling station, by means of shifting the stations electricity load out of high demand periods. This study also implements, evaluates and compares novel strategies for minimizing of the compressor switching frequency to mitigate wear and tear of compressor components caused by frequent on/off operation. The time-of-use tariff is implemented in order to encourage change in electricity usage by end users from the normal use pattern in response to change in electricity pricing based on time [54–56]. DR programs are implemented to cause intentional modifications of electricity consumption patterns by end-use customers, to alter the timing, demand level and total electricity consumption [57–60]. The present work seeks to explore the implementation of a response to the time-of-use DR program for the CNG fuelling station by tracking the gas demand profile. An optimization strategy is used to alter the operation of the compressor in order to achieve a reduction in electricity costs thereby lowering operating costs of the station. Lowering of the operating costs of CNG fuelling stations can improve the attractiveness of CNG as a fuel through passing on of some of the accrued savings to consumers in the form of reduced prices on CNG per unit of sale. The proposed method achieves a breakthrough in scheduling the on/off switching of the compressor while considering CNG demand, the electricity TOU pricing and the protection of the compressor from damage caused by excessive on-off cycling. The proposed methods provide measures for optimal use of energy resources through optimization of energy processes in alternative fuels.

2. Operation modelling and formulation

The schematic diagram of a CNG fast-fill refuelling station unit with a cascade storage system is shown in Fig. 1. The cascade storage tanks are supplied from the municipal gas supply line via the station compressor when it is turned on at switch u. The compressor switch is activated when pressure in the three reservoir tanks of the cascade storage falls to the lower limits [61]. The gas enters the storage tanks via a priority panel which is controlled to switch the incoming gas flow between the reservoir tanks by activating valves u_{hp} for the high pressure tank, u_{mp} for the medium pressure tank and u_{lp} for the low pressure tank. Only one of the three valves is activated at a time. When the maximum pressure limits are reached, the compressor is switched off. Gas from the cascade storage tanks is supplied to vehicle tank via a dispenser with a dispensing algorithm that compensates for the variation in temperature to ensure the correct quantities are transferred to vehicle tanks [24]. Depending on vehicle tank pressure, the dispenser switches the gas supply between the three cascade storage tanks to sustain flow of gas above a minimum flow rate. Considering the metered gas demand from the three reservoirs for the cascade storage m_{ohp},m_{omp} and m_{olp} for the high pressure reservoir, medium pressure reservoir and low pressure reservoir respectively, the proposed strategy seeks to optimally schedule the on/off switching of the compressor via switch u, as well as the priority panel valves u_{hp},u_{mp} and u_{lp} in order to minimize the cost of power consumed by the compressor when

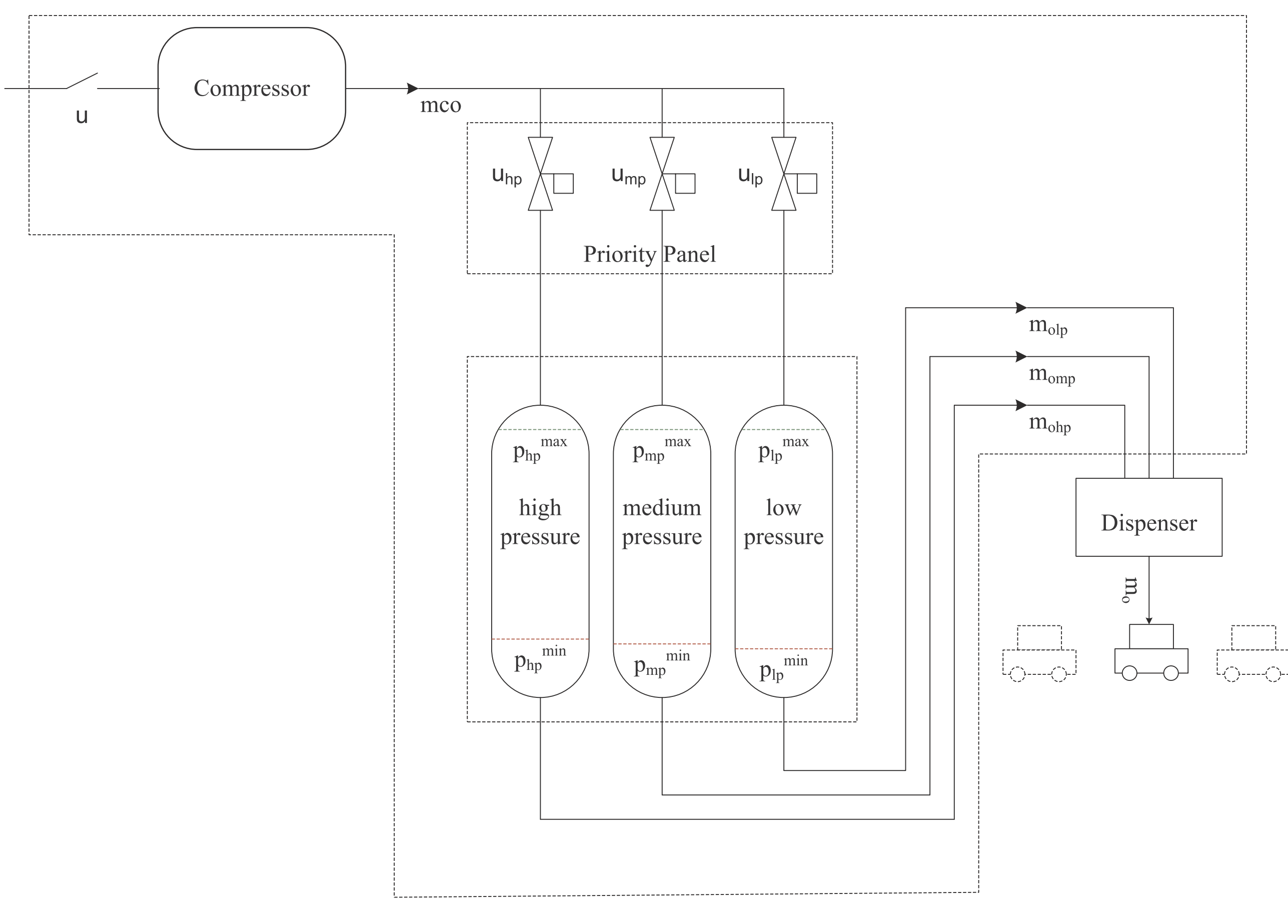


Fig. 1. Schematic of a CNG fuelling station with a cascade storage system.

electricity is purchased under a TOU tariff. The status of the compressor switch u and the status of the priority panel valves u_{hp},u_{mp} and u_{lp} are the control variables in the current problem.

2.1. The objective function

The objective is to minimize the cost of power consumed by the CNG fuelling station compressor within the limits of the cascade storage system over the control horizon [62]. The objective function is therefore expressed as

$$J = \sum_{t=1}^{N} p_{co}p_e(t)u(t)t_s \tag{1}$$

where p_{co} is the power rating of the electric motor driving the compressor, $p_e(t)$ is the price of electricity per kWh in a sampling interval, N is the total number of samples and t_s is the sampling interval. The horizon in the present work is 24 h, divided into a sampling interval of 4 min yielding a total number of samples $N = \frac{24 \times 60}{4} = 360$. The sampling time of 4 min represents the average fuelling time of vehicles visiting a typical CNG fast-fill station [29,63].

2.2. Constraints

The problem in this study is subject to the following constraints.

2.2.1. Reservoir capacity

The mass of gas in the cascade storage high pressure, medium pressure and low pressure tanks in the tth sampling interval is

$$m_{hp}(t) = m_{hp}(0) + \sum_{i=1}^{t-1} t_s\dot{m}_{cmp}u_{hp}(i) - \sum_{i=1}^{t-1} m_{ohp}(i) \tag{2}$$

$$m_{mp}(t) = m_{mp}(0) + \sum_{i=1}^{t-1} t_s\dot{m}_{cmp}u_{mp}(i) - \sum_{i=1}^{t-1} m_{omp}(i) \tag{3}$$

$$m_{lp}(t) = m_{lp}(0) + \sum_{i=1}^{t-1} t_s\dot{m}_{cmp}u_{lp}(i) - \sum_{i=1}^{t-1} m_{olp}(i) \tag{4}$$

respectively, where m_{ohp},m_{omp} and m_{ohp} are the values of mass dispensed in the ith sampling instant from the high pressure, medium pressure and low pressure reservoirs respectively. $\dot{m}_{cmp}$ is the outlet mass flow rate of the compressor obtained using the Eq. (5)[29]

$$\dot{m}_{cmp} = \rho_{std} \times Q_{std} = \left(\frac{Mw_g}{Mw_a}\right) \times \rho_{a,std} \times Q_{std} \tag{5}$$

where ρ_{std} is the density of the gas being compressed under standard conditions (0 °C) temperature and 10^5 Pa pressure),[1] Mw_g is the molecular weight of the gas, Mw_a the molecular weight of air, $\rho_{a,std}$ is the air density under standard conditions and Q_{std} is the capacity of the compressor under standard conditions.

The mass in the tanks in Eqs. (2)–(4) must be sustained at values between the masses corresponding to the maximum pressure and minimum pressure for the three reservoirs respectively such that

$$m_{hp}^{min} \leqslant m_{hp}(0) + \sum_{i=1}^{t-1} t_s\dot{m}_{cmp}u_{hp}(i) - \sum_{i=1}^{t-1} m_{ohp}(i) \leqslant m_{hp}^{max} \tag{6}$$

$$m_{mp}^{min} \leqslant m_{mp}(0) + \sum_{i=1}^{t-1} t_s\dot{m}_{cmp}u_{mp}(i) - \sum_{i=1}^{t-1} m_{omp}(i) \leqslant m_{mp}^{max} \tag{7}$$

$$m_{lp}^{min} \leqslant m_{lp}(0) + \sum_{i=1}^{t-1} t_s\dot{m}_{cmp}u_{lp}(i) - \sum_{i=1}^{t-1} m_{olp}(i) \leqslant m_{lp}^{max} \tag{8}$$

where m_{hp}^{min},m_{mp}^{min} and m_{mp}^{min} are the values of minimum quantity of gas for the high pressure, medium pressure and low pressure reservoirs respectively, while m_{hp}^{max},m_{mp}^{max} and m_{lp}^{max} are the values for the maximum quantity of gas for the high pressure, medium pressure and low pressure reservoirs respectively. These quantity limits can be obtained from the relationship between gas properties and the system rated pressure limits in the equation of state

$$pV = znRT \tag{9}$$

where p is the pressure, V the volume, T the absolute temperature, R is the universal gas constant, z the compressibility factor and n is the quantity of gas in moles [64,65].

$$n = \frac{m}{M} = \frac{pV}{zRT} \tag{10}$$

where m is the mass of the gas and M is the molar mass of the gas. Eq. (10) can be used to determine the upper and lower limits of the mass content of the reservoir tanks in the cascade storage of the CNG refuelling station. In the present work we consider the effect of the ambient temperature on the cascade storage by assuming that the cascade storage reservoirs normalize to the ambient temperature during charging and discharging. The maximum and minimum temperatures for the control horizon are taken as the day's highest and lowest temperatures respectively. When the ambient temperature rises, the pressure in the reservoirs will rise. If the upper pressure limit is breached, the safety release valve will open to expel excess gas into the atmosphere. It is therefore necessary to calculate and set the upper limit of mass that can be held in the reservoir at the highest temperature in the control horizon. Similarly, the lower limit of the reservoirs must be calculated and set at the lowest temperature to prevent the pressure of gas held in storage from falling below the lower limit when ambient temperature falls, causing a pressure drop. The limits are therefore calculated as

$$m_{hp}^{max} = \frac{MVp_{hp}^{max}}{zRT_{max}} \quad m_{hp}^{min} = \frac{MVp_{hp}^{mim}}{zRT_{min}} \tag{11}$$

$$m_{mp}^{max} = \frac{MVp_{mp}^{max}}{zRT_{max}} \quad m_{mp}^{min} = \frac{MVp_{mp}^{min}}{zRT_{min}} \tag{12}$$

$$m_{lp}^{max} = \frac{MVp_{lp}^{max}}{zRT_{max}} \quad m_{lp}^{min} = \frac{MVp_{lp}^{min}}{zRT_{min}} \tag{13}$$

where p_{hp}^{max},p_{mp}^{max} and p_{lp}^{max} are the maximum pressure limits for the respective reservoirs, p_{hp}^{min},p_{mp}^{min} and p_{lp}^{min} are the minimum pressure limits for the respective reservoirs and T_{max} and T_{min} are the maximum and minimum ambient temperatures in the control horizon respectively.

2.2.2. Switching combinations

During operation, whenever the compressor switch u is turned on, only one of the valves u_{hp},u_{mp} and u_{lp} can be on [29,66] and all the valves must be off whenever the compressor is off such that

$$u_{hp} + u_{mp} + u_{lp} - u = 0 \tag{14}$$

2.3. Compressor switching frequency

Frequent on/off switching increases mechanical stress induced in the compressor components, which causes an increase in wear and tear [67] and therefore increases the maintenance costs, while reducing the life of the compressor [68]. Transient start-up and shut-down states of the compressor have been shown to induce the highest stresses in different compressor components [69]. Additionally, there are vibrations resulting from torsional oscillations caused by loading changes throughout the compressor shaft, seals and coupled mechanisms as the motor pulls the load towards stabilization which cause further wear and tear [70]. It is therefore desirable that after the compressor is turned on,

[1] https://goldbook.iupac.org/html/S/S05910.html.

Table 1
CNG fuelling station data.

Specification	value
High pressure reservoir capacity	2000 L
Medium pressure reservoir capacity	2000 L
Low pressure reservoir capacity	2000 L
Maximum Pressure for all reservoir levels	252 bar
High pressure reservoir minimum pressure	210 bar
Medium pressure reservoir minimum pressure	150 bar
Low pressure reservoir minimum pressure	75 bar
Priority panel	3 lines
Compressor capacity	900 N m^3/h
Compressor motor rating	132 kW

it should be kept running for as long as possible, meaning it should operate in wide on or off state bands [71]. This is the motivation of presenting switch frequency minimization in our work. It would be more accurate to directly minimize the actual wear and tear caused by switching in the model instead of minimizing the switching numbers. However, there is no accurate model that captures the relationship between the switching actions and the wear and tear. Because of lack of such models in literature, the approach by many researchers is to penalise the occurrence of transient states when optimizing operation of rotary machine systems such as compressors [62,67,72]. This is still a valid approach because it is a known prior that the wear and tear increases with the increase of switching frequency.

In this study, we aim to achieve the least number of switching instances that does not raise the cost of electricity when compared to the cost achieved during optimization without penalizing of switching frequency. We implement two methods of achieving a minimum number of switching instances of the compressor. The first method is the Pretoria method proposed by [73] and used subsequently in optimization of the performance of water pumping systems [63,74]. The Pretoria method makes use of an auxiliary variable $s(t)$ which assumes a value of 1 whenever a switch start-up occurs. The auxiliary variable is then minimized in the objective function (36) which is then expressed as

$$J = \xi \sum_{t=1}^{N} p_{co}p_e(t)u(t)t_s + (1-\xi) \sum_{t=1}^{N} s(t) \tag{15}$$

where ξ is a weighting factor. Additional to the constraints in Sections 2.2.1 and 2.2.2, the problem is subject to further constraints arising from the use of the auxiliary variable. The constraints are

$$u(1)-s(1) \leqslant 0 \tag{16}$$

$$u(t)-u(t-i)-s(t) \leqslant 0 \tag{17}$$

where the inequality (16) initializes the auxiliary variable as the initial status of u while the inequality (17) favours the control that involves less switching instances [75]. The Pretoria method was shown to be superior to a method proposed by [63] that used constraints to restrict the number of on/off instances for the switch. The constraint method was determined to run at risk of infeasibility in certain control conditions.

We propose a second approach we call the non-linear objective function method which is describe in Section 2.6.

2.4. Boundaries

The condition of the switch u and the valves u_{hp},u_{mp} and u_{lp} as well as the auxiliary variable is binary such that

$$u(t),u_{hp}(t),u_{mp}(t),u_{lp}(t),s(t) \in \{0,1\} \quad (1 \leqslant t \leqslant N) \tag{18}$$

2.5. Algorithm

The generalized optimization problem in the present study is to minimize $f^T X$ subject to equality constraints ($A_{eq}X = b_{eq}$), inequality constraints ($AX \leqslant b$) and the upper and lower boundaries of the control variables ($L_B \leqslant X \leqslant U_B$) [76]. The control variables $u(t)$,$u_{hp}(t)$,$u_{mp}(t)$,$u_{lp}(t)$ and $s(j)$ are contained in vector X, while A and A_{eq} are matrices. b,L_B and U_B are vectors represented as

$$X = [u(1)\cdots u(N) \quad u_{hp}(1)\cdots u_{hp}(N) \quad u_{mp}(1)\cdots u_{mp}(N) \quad u_{lp}(1)\cdots u_{lp}(N) \quad s(1)\cdots s(N)]^T_{5N\times 1} \tag{19}$$

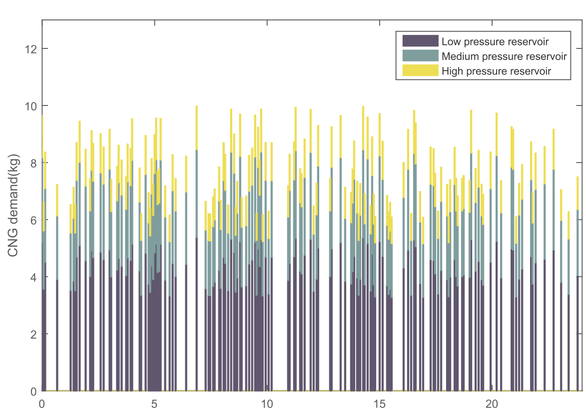
and the objective function as

$$f^T = [\xi p_{co} t_s p_e(1)\cdots \xi p_{co} t_s p_e(N) \quad 0\cdots 0 \quad 0\cdots 0 \quad 0\cdots 0 \quad (1-\xi)\cdots(1-\xi)]_{1\times 5N} \tag{20}$$

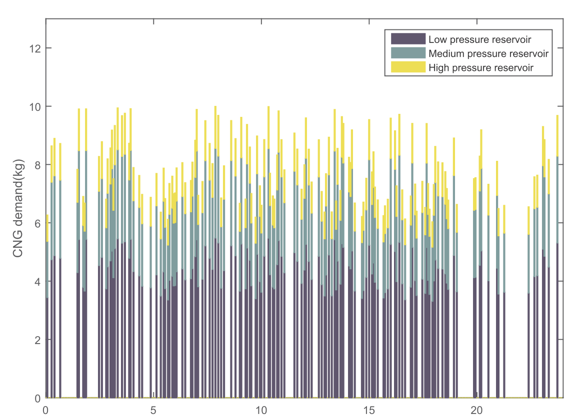
from the inequality constraints (6)–(8), (16) and (17), if we denote

$$A_c = \begin{bmatrix} -t_s\dot{m}_{cmp} & 0 & \cdots & 0 \\ -t_s\dot{m}_{cmp} & -t_s\dot{m}_{cmp} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -t_s\dot{m}_{cmp} & -t_s\dot{m}_{cmp} & \cdots & -t_s\dot{m}_{cmp} \end{bmatrix}_{N\times N} \tag{21}$$

$$A_u = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \end{bmatrix}_{N\times N} \tag{22}$$

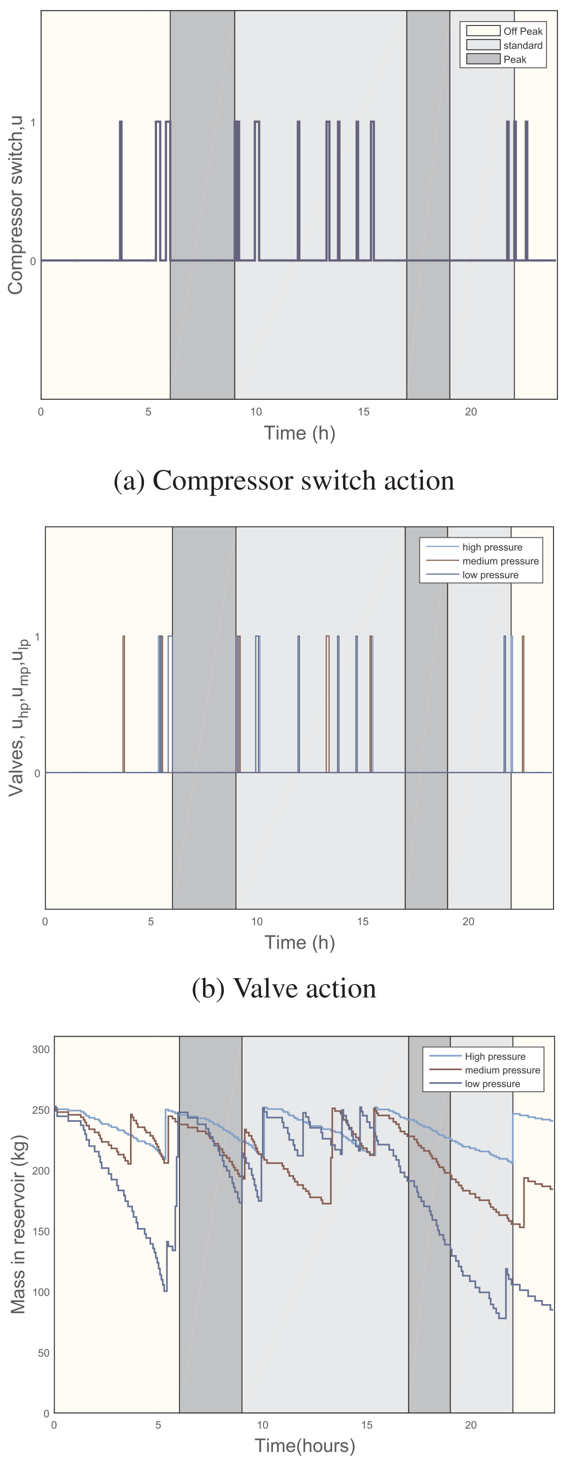


(a) High electricity demand season(143 vehicles)



(b) Low electricity demand season(146 vehicles)

Fig. 2. CNG demand from the three cascade storage levels for single day in each electricity pricing season.

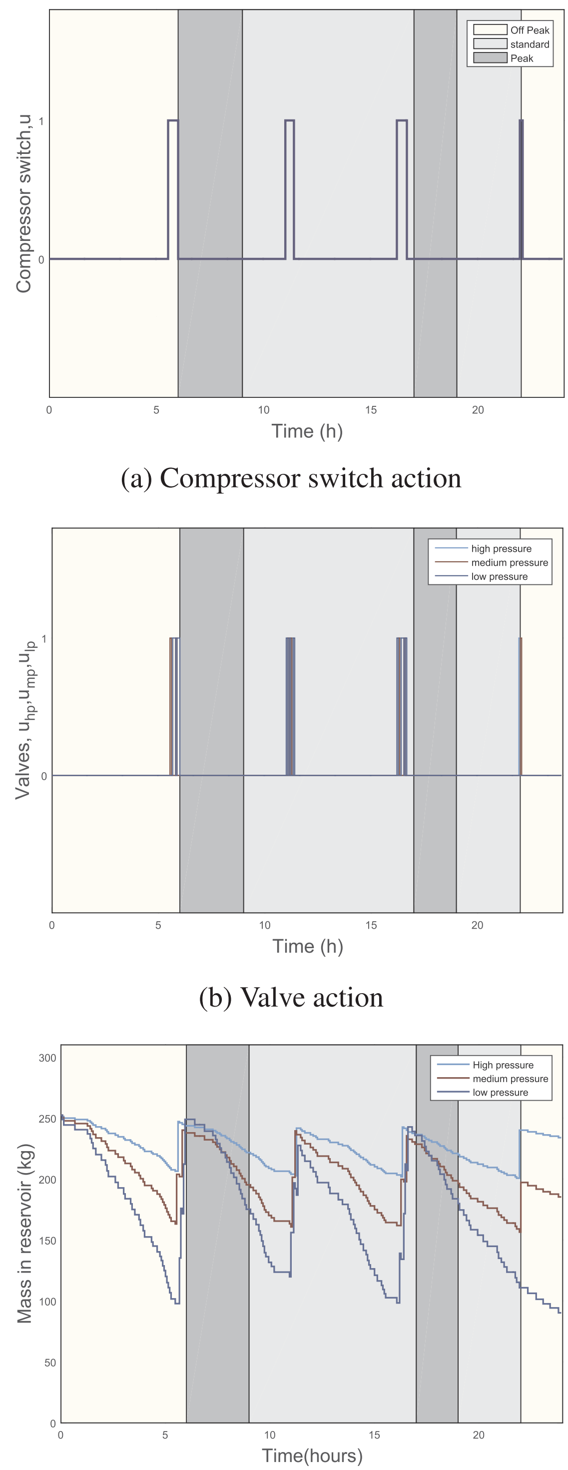


(a) Compressor switch action

(b) Valve action

(c) Mass of gas in reservoir

Fig. 3. System behaviour in the high demand electricity pricing season without consideration for compressor switching frequency.



(a) Compressor switch action

(b) Valve action

(c) Mass of gas in reservoir

Fig. 4. System behaviour in the high demand electricity pricing season using the Pretoria method of minimizing compressor switching.

then

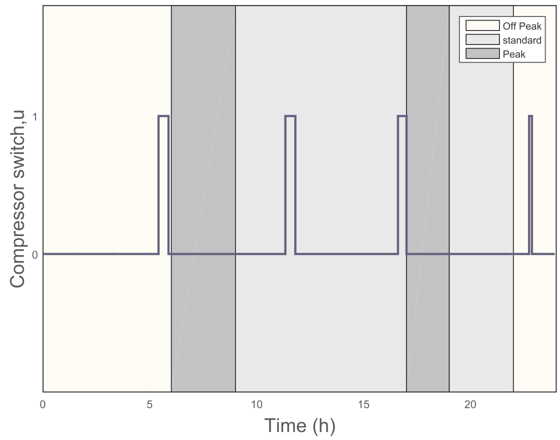
$$A_1 = [0 \quad A_c \quad 0 \quad 0 \quad 0]_{N\times 5N} \tag{23}$$

$$A_2 = [0 \quad 0 \quad A_c \quad 0 \quad 0]_{N\times 5N} \tag{24}$$

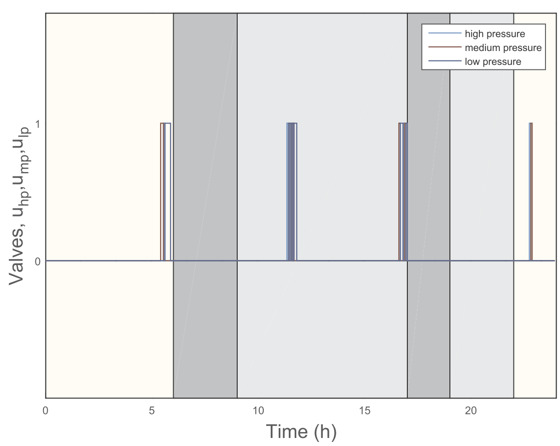
$$A_3 = [0 \quad 0 \quad 0 \quad A_c \quad 0]_{N\times 5N} \tag{25}$$

$$A_4 = [A_u \quad 0 \quad 0 \quad 0 \quad -I]_{N\times 5N} \tag{26}$$

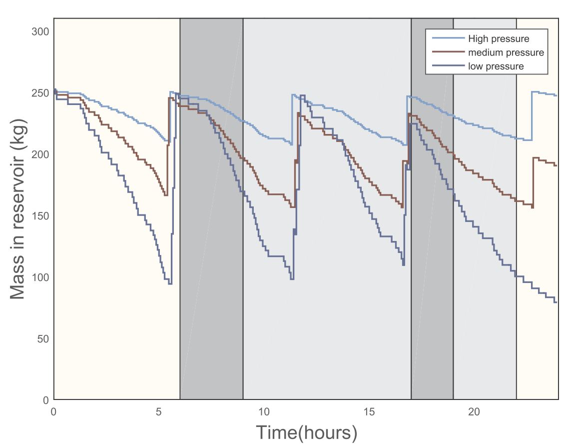
$$b_1 = \begin{bmatrix} m_{hp}(0) - m_{hp}^{min} - m_{ohp}(1) \\ m_{hp}(0) - m_{hp}^{min} - (m_{ohp}(1) + m_{ohp}(2)) \\ \vdots \\ m_{hp}(0) - m_{hp}^{min} - (m_{ohp}(1) + m_{ohp}(2) + \cdots + m_{ohp}(N)) \end{bmatrix}_{N\times 1} \tag{27}$$



(a) Compressor switch action



(b) Valve action



(c) Mass of gas in reservoir

Fig. 5. System behaviour in the high demand electricity pricing season using the non-linear objective function method of minimizing compressor switching.

$$b_2 = \begin{bmatrix} m_{hp}^{max} - m_{hp}(0) + m_{ohp}(1) \\ m_{hp}^{max} - m_{hp}(0) + (m_{ohp}(1) + m_{ohp}(2)) \\ \vdots \\ m_{hp}^{max} - m_{hp}(0) + (m_{ohp}(1) + m_{ohp}(2) + \cdots + m_{ohp}(N)) \end{bmatrix}_{N\times 1} \tag{28}$$

Table 2
Comparison of performance for the control strategies.

	Electricity cost (Rands)	Switching instances	Computing time (sec)
Optimal operation without switching minimization	148.90	16	0.5
Optimal operation with Pretoria switching minimization	148.90	4	71
Optimal operation with non-linear Objective function switching minimization	148.90	4	21

$$b_3 = \begin{bmatrix} m_{mp}(0) - m_{mp}^{min} - m_{omp}(1) \\ m_{mp}(0) - m_{mp}^{min} - (m_{omp}(1) + m_{omp}(2)) \\ \vdots \\ m_{mp}(0) - m_{mp}^{min} - (m_{omp}(1) + m_{omp}(2) + \cdots + m_{omp}(N)) \end{bmatrix}_{N\times 1} \tag{29}$$

$$b_4 = \begin{bmatrix} m_{mp}^{max} - m_{mp}(0) + m_{omp}(1) \\ m_{mp}^{max} - m_{mp}(0) + (m_{omp}(1) + m_{omp}(2)) \\ \vdots \\ m_{mp}^{max} - m_{mp}(0) + (m_{omp}(1) + m_{omp}(2) + \cdots + m_{omp}(N)) \end{bmatrix}_{N\times 1} \tag{30}$$

$$b_5 = \begin{bmatrix} m_{lp}(0) - m_{lp}^{min} - m_{olp}(1) \\ m_{lp}(0) - m_{lp}^{min} - (m_{olp}(1) + m_{olp}(2)) \\ \vdots \\ m_{lp}(0) - m_{lp}^{min} - (m_{olp}(1) + m_{olp}(2) + \cdots + m_{olp}(N)) \end{bmatrix}_{N\times 1} \tag{31}$$

$$b_6 = \begin{bmatrix} m_{lp}^{max} - m_{lp}(0) + m_{olp}(1) \\ m_{lp}^{max} - m_{lp}(0) + (m_{olp}(1) + m_{olp}(2)) \\ \vdots \\ m_{lp}^{max} - m_{lp}(0) + (m_{olp}(1) + m_{olp}(2) + \cdots + m_{olp}(N)) \end{bmatrix}_{N\times 1} \tag{32}$$

$$b_7 = [0]_{N\times 1} \tag{33}$$

then the linear inequality constraints become

$$A = \begin{bmatrix} A_1 \\ -A_1 \\ A_2 \\ -A_2 \\ A_3 \\ -A_3 \\ A_4 \end{bmatrix}_{7N\times 5N} \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \end{bmatrix}_{7N\times 1} \tag{34}$$

for the equality constraint (14) we denote

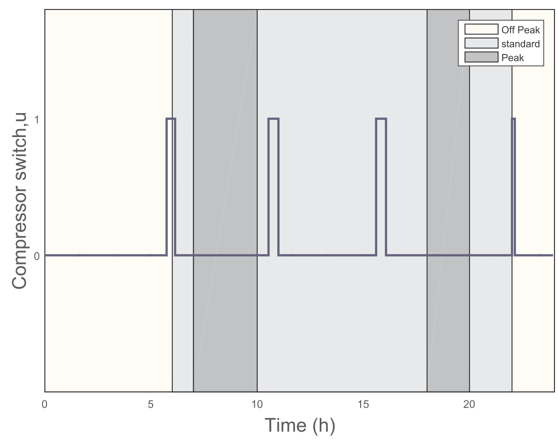
$$A_{eq} = [-I \quad I \quad I \quad I \quad 0]_{N\times 5N} \quad b_{eq} = [0]_{N\times 1} \tag{35}$$

This binary linear optimization problem is solved using the MATLAB Solving Constraint Integer Programs (SCIP) solver in the OPTI toolbox interface.

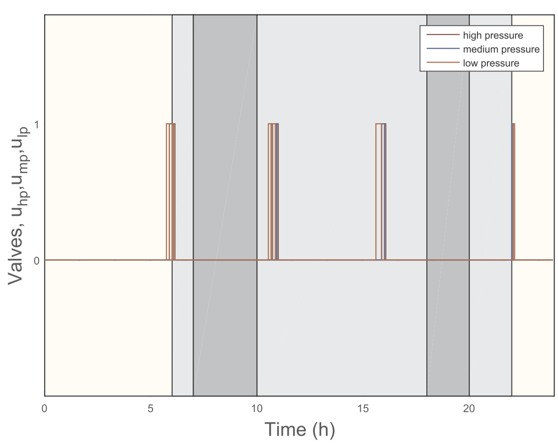
2.6. The non-linear objective function method to minimize compressor switching instances

In this method, we introduce the quadratic element $\sum(u(t+1) - u(t))^2$ to the objective function. The element minimizes the rate of change of status of the switch over the control horizon so as to achieve longer operating bands in both on and off states. The control variables are u,u_{hp},u_{mp} and u_{lp} and the objective function (36) using this method becomes

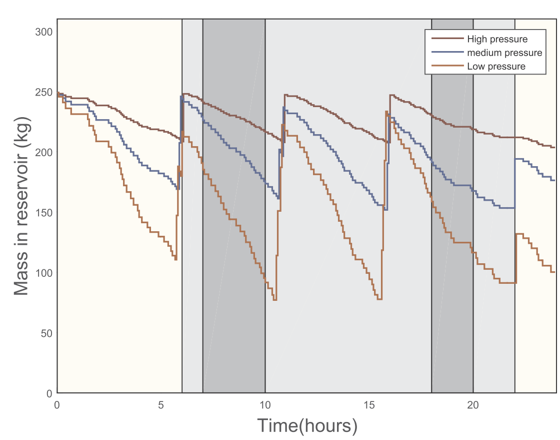
$$J = \psi \sum_{t=1}^{N} p_{co} p_e(t) u(t) t_s + (1-\psi) \sum_{t=1}^{N-1} (u(t+1) - u(t))^2 \tag{36}$$



(a) Compressor switch action



(b) Valve action



(c) Mass of gas in reservoir

Fig. 6. System behaviour in the low demand electricity pricing season using the non-linear objective function method.

where ψ is a weighting factor. The constraints in this approach remain as in Sections 2.2.1 and 2.2.2. The optimization problem can be written in the standard form

$$x = [u(1)\cdots u(N) \quad u_{hp}(1)\cdots u_{hp}(N) \quad u_{mp}(1)\cdots u_{mp}(N) \quad u_{lp}(1)\cdots u_{lp}(N)]^T_{4N\times 1} \tag{37}$$

and in the general OPTI toolbox solver algorithm the objective function is formulated as

$$min_x\ f^Tx\ subject\ to = \begin{cases} A\cdot x \leqslant b \\ A_{eq}\cdot x \leqslant b_{eq} \\ l_b \leqslant x \leqslant U_b \\ x \in \{0,1\} \end{cases} \tag{38}$$

from the linear inequality constraints (6)–(8), we can denote

$$A_1' = [0 \quad A_c \quad 0 \quad 0]_{N\times 5N} \tag{39}$$

$$A_2' = [0 \quad 0 \quad A_c \quad 0]_{N\times 5N} \tag{40}$$

$$A_3' = [0 \quad 0 \quad 0 \quad A_c]_{N\times 5N} \tag{41}$$

then the linear inequality constraints become

$$A' = \begin{bmatrix} A_1' \\ -A_1' \\ A_2' \\ -A_2' \\ A_3' \\ -A_3' \end{bmatrix}_{6N\times 4N} \qquad b' = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \end{bmatrix}_{6N\times 1} \tag{42}$$

for the equality constraint (14) we can denote

$$A_{eq}' = [-I;-I;-I;-I]_{N\times 4N} \quad b_{eq}' = [0]_{N\times 1} \tag{43}$$

This non-linear optimization problem is also solved using the MatLab SCIP solver in the OPTI toolbox interface.

2.7. Consideration for terminal constraints

It is desirable that the quantity of gas available at the end of the control horizon be similar to the initial quantity of gas in the cascade storage to ensure that the initial conditions are repeated for the next control period because open loop strategies do not guarantee proper operation in the subsequent control periods if initial conditions are not the same [74,77]. From Eqs. (2)–(4), the mass of gas in the three tanks of the cascade storage at the end of the control horizon can be set as;

$$m_{hp}(N) = m_{hp}(0) = m_{hp}(0) + \sum_{i=1}^{N} t_s\dot{m}_{cmp}u_{hp}(i) - \sum_{i=1}^{N} m_{ohp}(i) \tag{44}$$

$$m_{mp}(N) = m_{mp}(0) = m_{mp}(0) + \sum_{i=1}^{N} t_s\dot{m}_{cmp}u_{mp}(i) - \sum_{i=1}^{N} m_{omp}(i) \tag{45}$$

$$m_{lp}(N) = m_{lp}(0) = m_{lp}(0) + \sum_{i=1}^{N} t_s\dot{m}_{cmp}u_{lp}(i) - \sum_{i=1}^{N} m_{olp}(i) \tag{46}$$

It was deemed necessary to implement the terminal constraints as soft constraints, considering a fixed mass of gas flows into the cascade storage when the compressor is switched on in a sampling interval. This may not allow the solution to be automatically feasible with a hard constraint. The softened terminal constraint is a restriction of the final mass of gas in the cascade storage to within 90% of the initial mass in a control period. Therefore from Eqs. (44)–(46)

$$0.9\times m_{hp}(0) \leqslant m_{hp}(0) + \sum_{i=1}^{N} t_s\dot{m}_{cmp}u_{hp}(i) - \sum_{i=1}^{N} m_{ohp}(i) \leqslant m_{hp}(0) \tag{47}$$

$$0.9\times m_{mp}(0) \leqslant m_{mp}(0) + \sum_{i=1}^{N} t_s\dot{m}_{cmp}u_{mp}(i) - \sum_{i=1}^{N} m_{omp}(i) \leqslant m_{mp}(0) \tag{48}$$

$$0.9\times m_{lp}(0) \leqslant m_{lp}(0) + \sum_{i=1}^{N} t_s\dot{m}_{cmp}u_{lp}(i) - \sum_{i=1}^{N} m_{olp}(i) \leqslant m_{lp}(0) \tag{49}$$

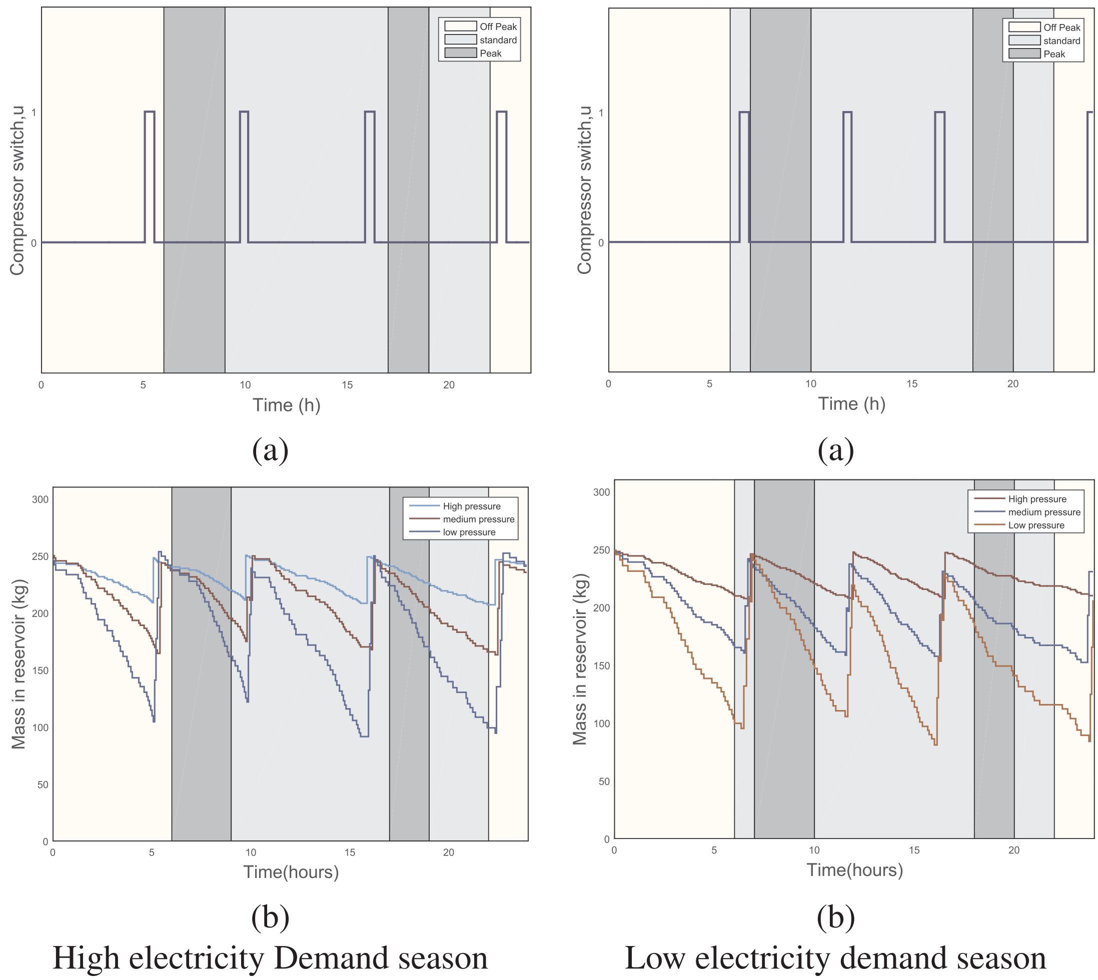


Fig. 7. (a) Compressor switching and (b) mass of gas in reservoir results with terminal constraints for high and low electricity demand seasons.

3. Case data

3.1. The CNG fuelling station

A CNG fuelling station in Johannesburg, South Africa is used as a case for the present study. The fuelling station has two dispensers with four refuelling nozzles and one reciprocating compressor. The fuelling station is supplied with gas from the utility company contracted by Johannesburg municipality to operate the municipality's gas pipeline. Table 1 shows the specifications of the CNG fuelling station unit. Under baseline operation, the low pressure limits of the cascade storage trigger the switching on of the compressor while the high pressure limits trigger the switching off.

3.2. Time-of-use electricity tariff

The time-of-use (TOU) electricity tariff is used in the electricity power industry so that retail electricity pricing is such that it reflects changes in the wholesale electricity market due to electricity demand [78]. It may vary by time of day, by day of the week and by seasons [79,36]. South Africa electric utility company, Eskom's tariff[2] for businesses named the TOU Miniflex is applicable to the CNG station in this study. The tariff is implemented at two levels; seasonal as well as time of day. Seasonal pricing is divided into high demand season in the winter months of June to August while the rest of the year is priced as low demand season, September to May. Further, the peak, standard and off-peak times in the day differ in the two seasons such that

$$p_{eHD}(t) = \begin{cases} p_{offpeak} = 0.5157 \text{ R/kW h} & \text{if } t \in [0,6] \cup [22,24] \\ p_{standard} = 0.9446 \text{ R/kW h} & \text{if } t \in [9,17] \cup [19,22] \\ p_{peak} = 3.1047 \text{ R/kW h} & \text{if } t \in [6,9] \cup [17,19] \end{cases} \tag{50}$$

$$p_{eLD}(t) = \begin{cases} p_{offpeak} = 0.4472 \text{ R/kW h} & \text{if} t \in [0,6] \cup [22,24] \\ p_{standard} = 0.7016 \text{ R/kW h} & \text{if} t \in [6,7] \cup [10,18] \cup [20,22] \\ p_{peak} = 1.0167 \text{ R/kW h} & \text{if} t \in [7,10] \cup [18,20] \end{cases} \tag{51}$$

where p_{eHD} and p_{eLD} is the price during high demand season and low demand season respectively, $p_{offpeak}$ is the off-peak price, $p_{standard}$ the price at standard time and p_{peak} the price at peak time, R is the South African Rand and t is the time of the day in hours. The tariff also has other charge components that are not considered in this study as they are constant [74].

3.3. Gas demand

The mass of gas flowing from the three CNG reservoir storage levels measured at the dispenser results in a gas demand profile shown in Fig. 2 for two days; one in the high demand electricity pricing season and the second in the low demand electricity pricing season. The dispenser uses mass flow measurements from each reservoir to make decision on valve sequencing when filling the vehicle tank and determine sale quantities with an operation log recording system performance

[2] http://eskom.co.za/tariffs.

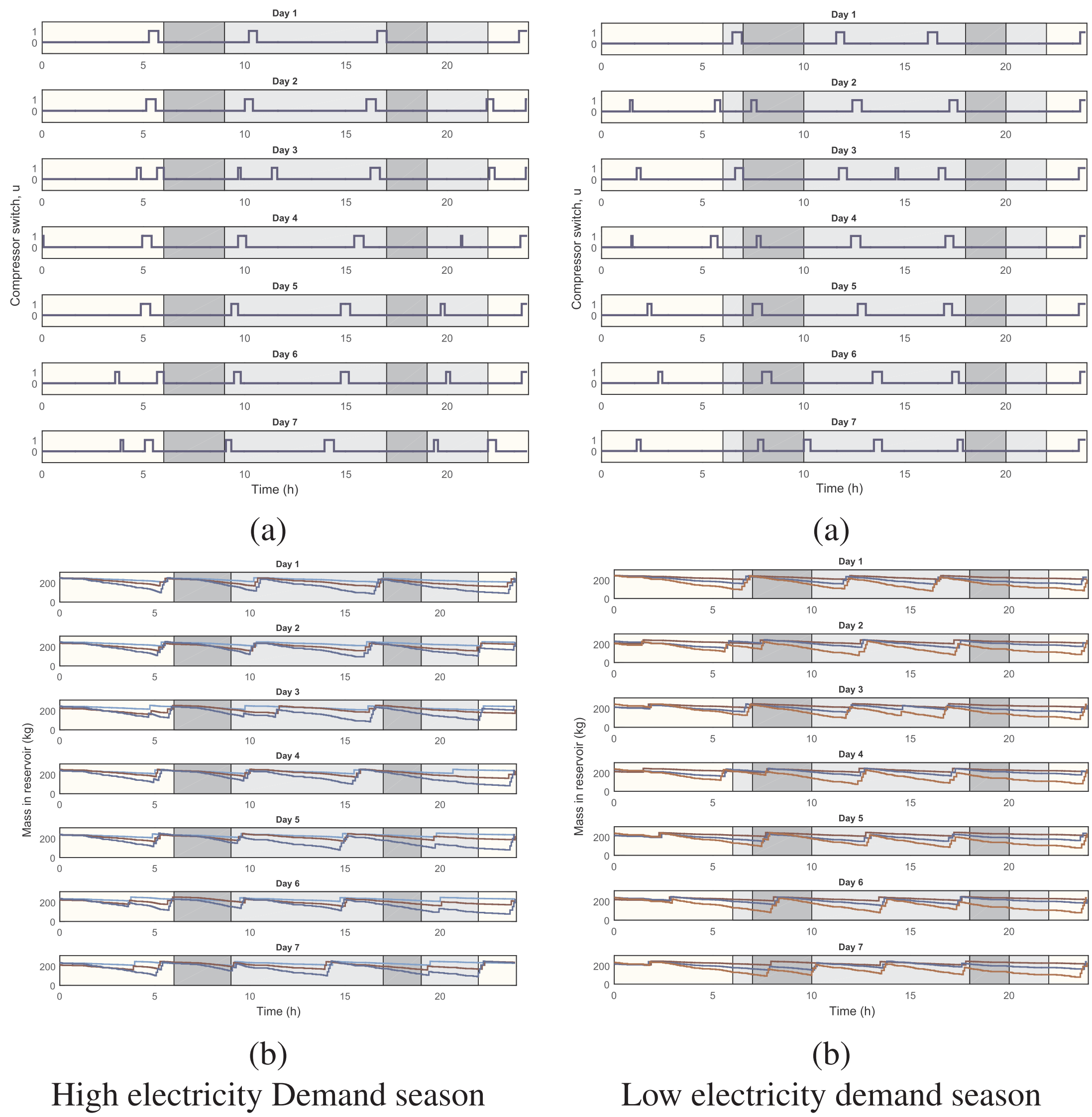


Fig. 8. (a) Compressor switching and (b) mass of gas in reservoir results with terminal constraints for one week in high and low electricity demand seasons.

[80]. The recorded mass of gas flowing from each reservoir for the two days in the two electricity pricing seasons bear similarities. From the profiles, there is an increase in gas demand in the early morning hours up to 10:00, due to motorists fuelling before beginning their journeys. There is also increased gas demand in the afternoon from 14:00 due to motorists refuelling in preparation for the evening rush hour. CNG powered vehicles are used mainly by courier fleet clients, security fleet clients and public service vehicles. Demand of gas tends to increase before and during people movement rush hours due to the public service transportation needs. Increased late afternoon and evening fuelling activity is also as a result of motorists who fuel prior to travelling the next day.

4. Results and discussion

4.1. High demand electricity pricing season

4.1.1. Optimization without consideration for compressor switching frequency

Fig. 3 shows the system behaviour when optimized without taking into consideration the switching frequency of the compressor in the objective function. Before the end of the off-peak electricity pricing at 06:00, the activity of the compressor increases in order to fill up the three reservoirs and therefore reduce compressor action in the peak electricity pricing band between 06:00 and 09:00. The method successfully avoids turning on the compressor in the peak electricity pricing time band thereby saving energy cost. Since the compressor stays off during the morning peak electricity pricing period, the compressor has to be switched on to satisfy gas demand for the standard electricity pricing period. The mass of gas in the reservoir is maintained at a high level before the beginning of the second electricity peak pricing period, to reduce the activity of the compressor in the undesirable time band between, 17:00 and 19:00. The approach also succeeds in preventing compressor activity in this second peak electricity pricing period, further reducing the cost of electricity consumed for the day. The compressor is turned on minimally in the following standard electricity pricing band between 19:00 and 22:00 to enable meet gas demand before the onset of the off-peak electricity pricing period when the compressor can supply gas to meet the rest of the day's demand. By operating the filling of each reservoir independently, the minimum electricity cost is achieved through ensuring the compressor runs only when there is predicted demand on an individual reservoir. From a baseline electricity cost of R432.59 for the day, a reduction of electricity cost to R148.90 is realised. Although this optimization regime minimizes the compressor action during the peak periods of electricity pricing thereby reducing electricity cost, the number of on/off switching of the compressor is too high at 14 and increases the probability of failure through fatigue to the compressor components which will results in an increase in maintenance cost [68,81].

4.1.2. Optimization while considering compressor switching frequency

Figs. 4 and 5 show the behaviour of the system when the switching frequency of the compressor is minimized using the Pretoria method with a weighting factor $\xi = 0.01$ and the non-linear objective function method with a weighting factor $\psi = 0.9$ respectively. These weighting factors were chosen to achieve the minimum number of switching instances without increasing the cost of electricity when compared to optimal operation without switching minimization. Both the Pretoria method and the non-linear objective function method are able to reduce the number of compressor turn-on instances to four when compared to the 14 instances in the optimized system without consideration for compressor switching frequency as can be seen in Figs. 4a and 5a. Compressor on-state is successfully avoided during the morning peak electricity pricing period between 06:00 and 09:00 by performing a reservoir refill just before the end of the off peak period at 06:00, prior to the onset of the peak electricity pricing period for both methods. Two compressor on-states occur in the standard electricity pricing time between 09:00 and 17:00 driven by the mid morning gas demand as well as the filling of the reservoirs just before the second peak electricity pricing period at 17:00. Both approaches are able to successfully prevent compressor activity in the second peak pricing period of the day and the subsequent standard electricity pricing period by predicting the demand in this time to ensure the compressor comes on only after the end of the standard electricity pricing period at 22:00.

The reduction in switching times for both the non-linear objective function method and the Pretoria method is achieved through synchronizing the utilization of the compressor to fill up each of the three reservoirs whenever an on-state occurs. Figs. 5b and 4b show how valve action is coordinated in the priority panel, to achieve gas levels in each of the reservoir that can sustain demand until the next synchronized need for a refill for the three reservoirs. The effects of the coordination on mass of gas in storage can be seen in the respective mass of gas in reservoir graphs in Figs. 4c and 5c. The reduction in switching occurrences represents a 71.4% reduction in the number of on/off actions that the compressor has to perform in comparison to optimization without consideration for the switching frequency. Fewer on/of instances mean a reduced probability of compressor component failure due to a high frequency of switching [82,83]. This also reduces the maintenance costs of the CNG fuelling station unit for which the compressor is an critical component [15,84].

A comparison of the three approaches to minimization of energy cost in the high demand electricity pricing season is shown in Table 2. From a baseline cost of power of R432.59, all the methods studied in the present work are able to reduce the cost to R148.90. The three strategies yield the same costs of electricity because they deliver the same effective compressor operation time in each electricity pricing period with variations only occurring in the exact time when the compressor-on state occurs and the length of time the compressor stays on under each strategy. Although the optimization strategy without switching control yields similar electricity cost savings as with the other two and with less computing time, the number of switching instances is too high and exposes the CNG fuelling station unit to higher probabilities of failure. The non-linear objective function method is demonstrably superior to the Pretoria method, given that it yields equal cost savings in a shorter computing time which corresponds to lower computing costs [85]. The optimization of the operation of a CNG fuelling station compressor and priority panel using the non-linear objective function method of compressor cost minimization is concluded as the superior approach to achieve the objective of electricity cost reduction in a TOU tariff electricity pricing regime as well as compressor care.

4.2. Low demand electricity pricing season

When applied to optimize operation of the CNG station considering the low electricity demand pricing season, the non-linear objective function method which has been determined to be the superior approach to the current problem in Section 4.1 results in the system behaviour shown in Fig. 6. To avoid compressor activity in the morning peak hours, the compressor is turned on some minutes before the morning standard electricity pricing period at 06:00 and stays on a few minutes into the standard electricity pricing period. This provides the cascade storage with sufficient gas to meet the demand without compressor activity past the morning peak electricity pricing time. Two compressor-on instances occur in the standard electricity pricing period between 10:00 and 18:00 which replenish the cascade storage sustaining the gas demand until after the night standard pricing period that ends at 22:00. A single compressor-on instance occurs in the subsequent off-peak period supplying gas for end of day demand. The profile accomplished by this approach reduces electricity cost for the CNG station in the low demand electricity pricing season from a baseline of R212.40 to R122.40 which is a 42.3% reduction in the day's electricity cost. This significant reduction in cost means that the energy cost reduction strategy is applicable throughout the year with significant savings in both electricity demand pricing seasons.

4.3. Solutions with terminal constraints

It is evident from Figs. 3c, 4c, 5c and 6c that the mass of gas in the reservoir at the end of the 24 h control horizon is different from the initial mass at the start of the control period under these strategies. This results in the initial conditions of the subsequent control period being different from those of the current one. When the terminal constraints are implemented with the non-linear objective function method of minimizing compressor switching frequency, the resulting profile of operation is shown in Fig. 7. In this regime, the strategy is able to keep the compressor operations outside of the high electricity pricing periods as well as raise the quantity of gas in storage close to the initial condition levels while keeping the number of compressor instances at four. However the cost of electricity incurred rises to R171.60 for the high electricity demand season and R139.84 for the low electricity demand season due to the operation for restoring the levels of gas to initial conditions compared to optimization without the terminal constraints. The effect of the terminal constraints can be observed when the optimization is repeated for seven consecutive days shown in Fig. 8. In general, the level of gas in storage after the end of each day remains similar to the initial conditions for that day. However the small variations have an effect on the performance of the strategy with a tendency to increase the number of compressor switching instances in some of the days. A maximum of seven switching instances for a single day occur during the high demand electricity pricing season while a maximum of six occur during the low electricity pricing season. These number of switching instances are still 50% lower than those observed under optimization without minimization of the switching frequency.

Over the course of the week in evaluation, the strategy is able to keep operation of the compressor outside the peak electricity pricing times for the high demand electricity pricing season. The resultant average cost of electricity is R176.13 per day, which represents 59.3% savings from the baseline. The highest cost at R179.91, happened on the fourth and seventh days while the lowest cost at R171.60 occurred on the first day. In the low demand electricity pricing season, the compressor is turned on during the morning peak electricity pricing time in the 4th, 5th, 6th and 7th days. However, the strategy still manages to keep the average cost of electricity per day at R158.54 with the highest cost observed on the fifth day at R172.54 and the lowest on the first day at R143.16. The average savings on electricity cost for the low demand electricity pricing period is 25%. The proposed strategy shows versatility in dealing with variations in initial conditions for consecutive days for the week under evaluation.

5. Conclusion

The use of compressed natural gas for the propulsion of motor vehicles can benefit greatly from optimization of the fuelling station operation by minimizing the energy cost. The present introductory work shows that 59.3% in electricity cost savings are achieved, while balancing the savings with a consideration for the life and reliability of the compressor. The subsequent savings translate to a reduction in compressor running cost by a 0.04 Rand cents margin per kilogram of gas sold in the low demand electricity pricing period and 0.23 Rand cents in the high demand electricity pricing period. These are significant margins that can allow compressed natural gas fuelling station operators to adjust the price of gas per unit of sale in order to attract more customers.

The control approach developed in this study can be used in setting operating schedules for compressors and priority panels in compressed natural gas stations operating in a time-based electricity tariff environment to save cost while prolonging the lifespan of the compressor by minimizing its frequent on/off switching. This can be done by changing the algorithm on the existing programmable logic controller that operates the compressor so that it is time-scheduled according to the results of the optimization, instead of the pressure limit cycled operation employed by the existing system. The main conditions arising from assumptions in this study that could affect the performance of the strategy in a real life scenario is a difference between the forecast and actual minimum and maximum temperatures over a control horizon which can result in unexpected changes in pressure. Safety interrupts for the maximum and minimum pressure can be implemented in the algorithm to deal with such cases where unpredictable events occur and expected optimized operation is violated. The safety interrupts can also deal with a variation in expected gas demand profile if it results in pressure limits in storage being reached.

The optimization model leads to improved economic efficiency of the compressed natural gas fuelling station and can reduce the contribution of the transportation industry to the emission of green house gases. This is because reduced costs have the potential to encourage consumer uptake of lower emission compressed natural gas powered vehicles over diesel and petrol vehicles. Further, the shifting of electric loads has the potential to reduce overall green house gas emission from the electricity generation infrastructure of the power utility. The outcome of the study justifies further optimization of the individual components of the fast-fill process to advance the goal of energy efficiency.

Acknowledgement

The authors thank the National Hub for Energy Efficiency and Demand Side Management at the University of Pretoria, South Africa for financial and other support.

References

[1] Bielaczyc P, Woodburn J, Szczotka A. An assessment of regulated emissions and CO_2 emissions from a European light-duty CNG-fueled vehicle in the context of Euro 6 emissions regulations. Appl Energy 2014;117:134–41.
[2] Movahed MM, Tabrizi HB, Mirsalim M. Experimental investigation of the concomitant injection of gasoline and CNG in a turbocharged spark ignition engine. Energy Convers Manage 2014;80:126–36.
[3] Yeh S. An empirical analysis on the adoption of alternative fuel vehicles: the case of natural gas vehicles. Energy Policy 2007;35(11):5865–75.
[4] Han Y, Kang J, Zhang G, Liu Z, Tian J, Chai J. Performance evaluation of free piston compressor coupling organic Rankine cycle under different operating conditions. Energy Convers Manage 2014;86:340–8.
[5] Akansu SO, Dulger Z, Kahraman N, Veziroglu TN. Internal combustion engines fueled by natural gas hydrogen mixtures. Int J Hydrogen Energy 2004;29(14):1527–39.
[6] Amirante R, Distaso E, Di Iorio S, Sementa P, Tamburrano P, Vaglieco B, et al. Effects of natural gas composition on performance and regulated, greenhouse gas and particulate emissions in spark-ignition engines. Energy Convers Manage 2017;143:338–47.
[7] Gilmore EA, Patwardhan A. Passenger vehicles that minimize the costs of ownership and environmental damages in the Indian market. Appl Energy 2016;184:863–72.
[8] Carle G. Erdgasfahrzeuge und ihr beitrag zu einer CO_2-reduktion im motorisierten personenverkehr der schweiz (Ph.D. thesis); 2006.
[9] Ally J, Pryor T. Life-cycle assessment of diesel, natural gas and hydrogen fuel cell bus transportation systems. J Power Sources 2007;170(2):401–11.
[10] Kirk JL, Bristow AL, Zanni AM. Exploring the market for compressed natural gas light commercial vehicles in the United Kingdom. Transp Res Part D: Transp Environ 2014;29:22–31.
[11] Ravindra K, Wauters E, Tyagi SK, Mor S, Van Grieken R. Assessment of air quality after the implementation of compressed natural gas (CNG) as fuel in public transport in Delhi, India. Environ Monit Assess 2006;115(1–3):405–17.
[12] Lin W, Zhang N, Gu A. LNG (liquefied natural gas): a necessary part in China's future energy infrastructure. Energy 2010;35(11):4383–91.
[13] D. of Minerals, E.R. of South Africa, Gas infra-structure plan, Tech. rep. Department of Minerals and Energy Republic of South Africa; 2005.
[14] Pretorius I, Piketh S, Burger R. The impact of the South African energy crisis on emissions. In: Longhurst JWS, Brebbia C, Brebbia CA, Barnes J, editors. Air pollution XXIII; 2015. p. 255–64.
[15] He HM, Lin WS, Gu AZ. Preliminary technical study on L/CNG stations. Nat Gas Ind 2007;27(4):126.
[16] Elgin RC, Hagen CL. Development and operation of a self-refueling compressed natural gas vehicle. Appl Energy 2015;155:242–52.
[17] Farzaneh-Gord M, Deymi-Dashtebayaz M, Rahbari HR. Studying effects of storage types on performance of CNG filling stations. J Nat Gas Sci Eng 2011;3(1):334–40.
[18] Ramoutar S, Riverol C. A thermodynamic analysis of refueling a natural gas vehicle cylinder from a cascade reservoir using chilled natural gas. J Nat Gas Sci Eng 2017;38:298–322.
[19] Kountz K, Blazek C, Liss W. A new natural gas dispenser control system. In: International gas research conference, vol. 4. Government Institutes Inc.; 1998. p. 135–45.
[20] Price BF. Apparatus and process for fast filling with natural gas. US Patent 5,370,159; Dec. 6 1994.
[21] Knowlton RE. Natural gas storage and retrieval system. US Patent 4,805,674; Feb. 21 1989.
[22] Barajas AM, Buckingham JC, Svedeman S. System for controlling the fill of compressed natural gas cylinders. US Patent 5,868,176; Feb. 9 1999.
[23] Han Y, You XD, Zhang X, Wang HY, Deng C. A novel design of the CNG dispenser electronic control system. In: Applied mechanics and materials, vol. 448. Trans Tech Publ.; 2014. p. 3119–22.
[24] Su C-L, Liu J-L, Qiu D-L, Cai J-Y, Li S-S. Intelligent compressed natural gas dispensers. China Measur Testing Technol 2008;6:041.
[25] Emmer CD, Gamble J, Zelasko C, Drube TK. Liquid and compressed natural gas dispensing system. US Patent 7,069,730; Jul. 4 2006.
[26] Gram A, Ursan M. Combined liquefied gas and compressed gas re-fueling station and method of operating same. US Patent 7,284,575; Oct. 23 2007.
[27] Tsuchiya M, Ogasa H, Matsuura H, Shimanuki H, Fujii I. Thermodynamic behavior of supply gas and influence on vehicle fuel fill line during CNG fast fill. Tech. rep. SAE Technical Paper; 1996.
[28] Thomas G, Goulding J, Munteam C. Measurement, approval and verification of CNG dispensers. NWML KT11 Report.
[29] Saadat-Targhi M, Khadem J, Farzaneh-Gord M. Thermodynamic analysis of a CNG refueling station considering the reciprocating compressor. J Nat Gas Sci Eng 2016;29:453–61.
[30] Mokhatab S, Poe WA. Handbook of natural gas transmission and processing. Gulf Professional Publishing; 2012.
[31] Xueqin J. Study of urban CNG station location arrangement and the size. Public Utilit 2007;1:005.
[32] Xia X, Zhang L. Industrial energy systems in view of energy efficiency and operation control. Ann Rev Control 2016;42:299–308.
[33] Xia X, Zhang J, Cass W. Energy management of commercial buildings-a case study from a POET perspective of energy efficiency. J Energy Southern Afr 2012;23(1):23–31.
[34] Xia X, Zhang J. Modeling and control of heavy-haul trains [applications of control]. IEEE Control Syst 2011;31(4):18–31.
[35] Xia X, Zhang J. Energy efficiency and control systems–from a POET perspective. IFAC Proc 2010;43(1):255–60.
[36] Zhuan X, Xia X. Optimal operation scheduling of a pumping station with multiple pumps. Appl Energy 2013;104:250–7.
[37] Xia X, Zhang J. Mathematical description for the measurement and verification of energy efficiency improvement. Appl Energy 2013;111:247–56.
[38] Zhuan X, Xia X. Development of efficient model predictive control strategy for cost-optimal operation of a water pumping station. IEEE Trans Control Syst Technol 2013;21(4):1449–54.
[39] Ye X, Xia X, Zhang J. Optimal sampling plan for clean development mechanism energy efficiency lighting projects. Appl Energy 2013;112:1006–15.
[40] Kountz KJ. Modeling the fast fill process in natural gas vehicle storage cylinders. Tech. rep. Chicago, IL (United States): Institute of Gas Technology; 1994.
[41] Farzaneh-Gord M. Compressed natural gas-single reservoir filling process. Int Gas Eng Manage 2008;48(6):16–8.
[42] Farzaneh-gord M, Hashemi S, Farzaneh-kord A. Thermodynamics analysis of cascade reservoirs filling process of natural gas vehicle cylinders 1.
[43] Olmos F, Manousiouthakis VI. Gas tank fill-up in globally minimum time: theory and application to hydrogen. Int J Hydrogen Energy 2014;39(23):12138–57.

[44] Deymi-Dashtebayaz M, Gord MF, Rahbari HR. Studying transmission of fuel storage bank to NGV cylinder in CNG fast filling station. J Braz Soc Mech Sci Eng 2012;34(4):429–35.
[45] Khadem J, Saadat-Targhi M, Farzaneh-Gord M. Mathematical modeling of fast filling process at CNG refueling stations considering connecting pipes. J Nat Gas Sci Eng 2015;26:176–84.
[46] Kountz KJ, Blazek CF, Christopher F. NGV fuelling station and dispenser control systems. Tech. rep., report GRI-97/0398. Chicago, Illinois: Gas Research Institute; November 1997.
[47] Stouffs P, Tazerout M, Wauters P. Thermodynamic analysis of reciprocating compressors. Int J Therm Sci 2001;40(1):52–66.
[48] Sun SY, Ren T-R. New method of thermodynamic computation for a reciprocating compressor: computer simulation of working process. Int J Mech Sci 1995;37(4):343–53.
[49] Elhaj M, Gu F, Ball A, Albarbar A, Al-Qattan M, Naid A. Numerical simulation and experimental study of a two-stage reciprocating compressor for condition monitoring. Mech Syst Signal Process 2008;22(2):374–89.
[50] Ndiaye D, Bernier M. Dynamic model of a Hermetic reciprocating compressor in on–off cycling operation (abbreviation: compressor dynamic model). Appl Therm Eng 2010;30(8):792–9.
[51] Frick M, Axhausen KW, Carle G, Wokaun A. Optimization of the distribution of compressed natural gas (CNG) refueling stations: Swiss case studies. Transp Res Part D: Transp Environ 2007;12(1):10–22.
[52] Malakoutirad M, Bradley TH, Hagen C. Design considerations for an engine-integral reciprocating natural gas compressor. Appl Energy 2015;156:129–37.
[53] Bang HJ, Stockar S, Muratori M, Rizzoni G. Modeling and analysis of a CNG residential refueling system. In: ASME 2014 dynamic systems and control conference, no. DSCC2014-6160. American Society of Mechanical Engineers; 2014.
[54] Jannati J, Nazarpour D. Optimal energy management of the smart parking lot under demand response program in the presence of the electrolyser and fuel cell as hydrogen storage system. Energy Convers Manage 2017;138:659–69.
[55] Albadi MH, El-Saadany E. A summary of demand response in electricity markets. Electr Power Syst Res 2008;78(11):1989–96.
[56] Zhang L, Xia X, Zhang J. Improving energy efficiency of cyclone circuits in coal beneficiation plants by pump-storage systems. Appl Energy 2014;119:306–13.
[57] Herter K, McAuliffe P, Rosenfeld A. An exploratory analysis of California residential customer response to critical peak pricing of electricity. Energy 2007;32(1):25–34.
[58] Setlhaolo D, Xia X, Zhang J. Optimal scheduling of household appliances for demand response. Electr Power Syst Res 2014;116:24–8.
[59] Ekpenyong UE, Zhang J, Xia X. Mathematical modelling for the social impact to energy efficiency savings. Energy Build 2014;84:344–51.
[60] Wu Z, Tazvinga H, Xia X. Demand side management of photovoltaic-battery hybrid system. Appl Energy 2015;148:294–304.
[61] Newhouse NL, Liss WE. Fast filling of NGV fuel containers. Tech. rep. SAE Technical Paper; 1999.
[62] Nguyen HH, Uraikul V, Chan CW, Tontiwachwuthikul P. A comparison of automation techniques for optimization of compressor scheduling. Adv Eng Softw 2008;39(3):178–88.
[63] Wanjiru EM, Xia X. Energy-water optimization model incorporating rooftop water harvesting for lawn irrigation. Appl Energy 2015;160:521–31.
[64] Heidaryan E, Moghadasi J, Rahimi M. New correlations to predict natural gas viscosity and compressibility factor. J Petrol Sci Eng 2010;73(1):67–72.
[65] Bahadori A, Mokhatab S, Towler BF. Rapidly estimating natural gas compressibility factor. J Nat Gas Chem 2007;16(4):349–53.
[66] Hopkins PF. Method and apparatus for dispensing compressed gas into a vehicle. US Patent 5,406,988; Apr. 18 1995.
[67] Nguyen HH, Chan CW. Applications of artificial intelligence for optimization of compressor scheduling. Eng Appl Artif Intell 2006;19(2):113–26.
[68] Bloch HP, Hoefner JJ. Reciprocating compressors:: operation and maintenance. Gulf Professional Publishing; 1996.
[69] Giacomelli E, Bernardini F, Pini A, Andree H. Application of gas turbines for large reciprocating compressor drive. In: ASME 1989 international gas turbine and aeroengine congress and exposition. American Society of Mechanical Engineers; 1989. p. V004T10A006–V004T10A006.
[70] Walsh P, Lamancusa J. A variable stiffness vibration absorber for minimization of transient vibrations. J Sound Vib 1992;158(2):195–211.
[71] Booysen W, Kleingeld M, Van Rensburg J. Optimising compressor control strategies for maximum energy savings. ICUE, Cape Town 2009:10–2.
[72] Sun CK, Uraikul V, Chan CW, Tontiwachwuthikul P. An integrated expert system/operations research approach for the optimization of natural gas pipeline operations. Eng Appl Artif Intell 2000;13(4):465–75.
[73] Mathaba T, Xia X, Zhang J. Analysing the economic benefit of electricity price forecast in industrial load scheduling. Electr Power Syst Res 2014;116:158–65.
[74] Wanjiru EM, Zhang L, Xia X. Model predictive control strategy of energy-water management in urban households. Appl Energy 2016;179:821–31.
[75] Bagirov AM, Barton A, Mala-Jetmarova H, Al Nuaimat A, Ahmed S, Sultanova N, et al. An algorithm for minimization of pumping costs in water distribution systems using a novel approach to pump scheduling. Math Comput Model 2013;57(3):873–86.
[76] Numbi BP, Zhang J, Xia X. Optimal energy management for a jaw crushing process in deep mines. Energy 2014;68:337–48.
[77] Faulwasser T, Korda M, Jones CN, Bonvin D. On turnpike and dissipativity properties of continuous-time optimal control problems. Automatica 2017;81:297–304.
[78] Zhang H, Zhao F, Fang K, Sutherland JW. Energy-conscious flow shop scheduling under time-of-use electricity tariffs. CIRP Ann-Manuf Technol 2014;63(1):37–40.
[79] Baladi SM, Herriges JA, Sweeney TJ. Residential response to voluntary time-of-use electricity rates. Resour Energy Econ 1998;20(3):225–44.
[80] Kountz KJ, Liss WE, Blazek CF. Automated process and system for dispensing compressed natural gas. US Patent 5,810,058; Sep. 22 1998.
[81] Bagirov AM, Barton A, Mala-Jetmarova H, Al Nuaimat A, Ahmed S, Sultanova N, et al. An algorithm for minimization of pumping costs in water distribution systems using a novel approach to pump scheduling. Math Comput Model 2013;57(3):873–86.
[82] Allan R, Billinton R, De Oliveira M. Reliability evaluation of electrical systems with switching actions. In: Proceedings of the institution of electrical engineers, vol. 123. IET; 1976. p. 325–30.
[83] Bosman M. Availability analysis of a natural gas compressor plant. Reliab Eng 1985;11(1):13–26.
[84] Wang HX, He M, Yang L, Li N, Wang Y-d. Suggestions on relevant design codes for CNG filling stations in China. Gas Heat 2009;12:027.
[85] He H, Daume III H, Eisner JM. Learning to search in branch and bound algorithms. In: Advances in neural information processing systems; 2014. p. 3293–301.

Solar Energy 170 (2018) 991–1000

Contents lists available at ScienceDirect

Solar Energy

journal homepage: www.elsevier.com/locate/solener

Tariff-driven demand side management of green ship

Zhou Wu[a,b,*], Xiaohua Xia[c]

[a] Key Laboratory of Complex System Safety and Control (Chongqing University), Ministry of Education, China
[b] School of Automation, Chongqing University, Chongqing, China
[c] Department of Electrical Electronic and Computer Engineering, University of Pretoria, Pretoria, South Africa

ARTICLE INFO

Keywords:
Green ship
Demand side management
Economic load dispatching
Receding horizon control

ABSTRACT

Green ships with hybrid renewable energy systems become important resources of demand side management, when ships in port have the grid connection. Variance of electricity tariff has influenced the optimal solutions to power management. Current power management methods for stand-alone green ships cannot be applied to this new situation. To enable tariff-driven power management, a unified model is proposed for a green ship under different time-of-use (TOU) tariffs. In the proposed model, diesel generation, solar energy, and battery storage could support auxiliary power demand, and the surplus of solar energy could be sold to grid when the ship is connected to grid. A power flow dispatching problem is then formulated as the optimization of operational cost. To cope with variance of tariff, solar energy, and on-board load demand, a receding horizon control approach is employed to ensure a closed-loop control mechanism. Experimental results indicate the tariff-driven model can effectively reduce the overall cost of green ships, and the receding horizon control can improve the performance in terms of fuel consumption and greenhouse gas emission.

1. Introduction

Over 90% of cargoes are transported by ships over the world, while greenhouse gas (GHG) emission and fossil fuel consumption are two critical problems in the shipping industry. In 2007, international shipping is responsible for approximately 3% global GHG emission, and 277 million tons of diesel/gasoline, in which the dry bulk shipping is the first contributor with about 52 million tons (Buhaug et al., 2009). To suppress the continuous increase of GHG emission and fossil fuel demand in the international shipping, international maritime organization (IMO) has issued strict regulations for shipping energy efficiency and GHG emission. Therefore, green ship technologies become urgent to improve shipping energy efficiency and reduce GHG emission. One of the most popular technology is to find clean energy to take the place of fossil fuel (Diab et al., 2016). Renewable energy (RE) resources have played increasingly significant roles to reduce fuel consumption and GHG emission in the green ship. Among available RE resources, solar energy is the most promising option of green ship, as solar is clean, safe, omnipresent, and freely available.

In general, photovoltaic (PV) panels have to be equipped together with storage components (battery, ultra-capacitor, and so on) for providing stable and sustainable power. Multiple renewable sources and storage components are usually combined in a hybrid renewable energy system (HRES). In the stand-alone application, e.g., remote communities, the HRES is able to supply electricity for off-grid customers (Tazvinga et al., 2013, 2015; Nema et al., 2009; Shaahid and El-Amin, 2009). In the grid-connected application, e.g., the berthing green ship, the HRES can also serve as distributed generation to sell the surplus of renewable energy on grid, which can bring financial profits on the electricity market (Palma-Behnke et al., 2013; Wu et al., 2015; Wu and Xia, 2015). Researchers have studied many theoretical and practical issues arisen in HRES applications, including optimal design (Arun et al., 2009), scheduling and control (Gabash and Li, 2013; Kanchev et al., 2011), maximum power point tracking (MPPT) (Soto et al., 2006), and economic analysis (Wies et al., 2005; Esen et al., 2007).

In recent years, the HRES has been applied to hybrid-electric ships and all-electric ships (Zahedi and Norum, 2013). On the one hand, new green ships are built with electric power systems, including PV, diesel generators (DGs), and battery (Lan et al., 2015; Wen et al., 2016; Banaei and Alizadeh, 2016). On the other hand, existing fossil fuel ships are undergoing energy efficient retrofit, and the HRES is installed to meet the axillary demand, such as loading, unloading, lighting, heating, cooling, and other on-board hotel services (Lee et al., 2013; Ovrum and Bergh, 2015). Compared with the fossil-fuel ships, the hybrid-electric ships are less dependent on fossil fuel, and have more integration of solar or wind energy. The use of renewable energy can improve energy efficiency of ship, enhance reliability and quality of power supply, and reduce shipping cost and GHG emission. The hybrid power system on

* Corresponding author.
E-mail addresses: wuzhsky@gmail.com (Z. Wu), xxia@up.ac.za (X. Xia).

https://doi.org/10.1016/j.solener.2018.06.033
Received 16 February 2018; Received in revised form 31 May 2018; Accepted 7 June 2018

Nomenclature

$P_1(t)$	power flow from diesel generator to internal bus (kW)
$P_2(t)$	power flow from internal bus to battery (kW)
$P_3(t)$	power flow from battery to internal bus (kW)
$P_4(t)$	bidirectional power flow between grid and internal bus (kW)
$P_{pv}(t)$	power output of PV panels (kW)
$P_{pl}(t)$	propulsion load of green ship (kW)
$P_{al}(t)$	auxiliary load of green ship (kW)
$P_D(t)$	power output of diesel generator (kW)
P_D^{max}	maximal power output of diesel generator (kW)
P_D^{max}	minimal power output of diesel generator (kW)
P_i^m	allowable maximal power on the ith line (kW)
v	status of switch on the grid connection
$\bar{v}$	inverse status of switch on the grid connection
$S(t)$	state of charge (SOC) of battery (%)
S^{max}	allowable maximum SOC (%)
S^{min}	allowable minimum SOC (%)
Q	capacity of battery (kWh)
η_C	charging efficiency of battery (%)
η_D	discharging efficiency of battery (%)
$\rho(t)$	price of electricity ($/kWh)

the green ship is usually regarded as a special case of mobile microgrid, which appears more complicated characteristics than the microgrid on land. System configurations are different when the ship is on voyage and berth, respectively. Environmental conditions are also extremely varying for the mobile microgrid. For the green ship, the mobile power system works on two modes, i.e., off-grid mode (stand-alone mode), and on-grid mode (grid-connected mode).

For the off-grid mode, many results have been reported in terms of optimal sizing (Lan et al., 2015; Wen et al., 2016; Yao et al., 2017), and power management (Banaei and Alizadeh, 2016; Ovrum and Bergh, 2015; Tsekouras and Kanellos, 2013). In Lan et al. (2015), an optimal sizing problem of stand-alone green ship has been formulated to minimize investment cost, fuel cost, and GHG emission, in which seasonal and geographical variation is considered for different routes. Interval optimization and clustering-based optimization methods have been proposed to determine the optimal size of energy storage system with uncertain PV power and load (Wen et al., 2016; Yao et al., 2017). To improve operational efficiency, power management has been studied for an electric ship with fuel cell, battery, PV panels, and diesel generators (Banaei and Alizadeh, 2016; Tsekouras and Kanellos, 2013). For crane ships, lithium-ion batteries have been employed to take part into power management, in which a hybrid control strategy is developed to reduce fuel cost and GHG emission (Ovrum and Bergh, 2015).

Other than the off-grid mode, green ships sometimes work on the grid-connected mode, when the shore-side grid power is available (Lee et al., 2013; Kanellos et al., 2017). As reported in Kökkülünk et al. (2016), average harboring time of bulk carrier ship is about 2 months per year. As the shore-side power is usually cleaner than the power generated on board, the use of shore-side power, called cold ironing, can effectively reduce annual fuel cost and GHG emission, when the green ship is on berth. With the help of HRES, solar energy can be used to supply the on-board demand instead of the shore-side electricity, and electricity cost can be significantly reduced. In Lee et al. (2013), a green cruise ship has been studied for delivering PV power to grid, and a rule-based strategy has been developed to satisfy auxiliary demand with batteries. In Kanellos et al. (2017) and Kanellos et al. (2014), a unit commitment problem has been studied to optimally allocate power output of each diesel generator, in which cold ironing is considered.

Considering bidirectional power flow between green ship and shore-side grid, electricity tariffs must influence electricity cost of cold ironing, and possible reward from selling renewable energy to grid. Thus, the change of electricity tariff will drive a different optimal solution to power management. To our best knowledge, very limited studies have evaluated tariff effects on power management of hybrid-electric ship. As a kind of demand side resources, on-grid green ships could take part into demand response programs, such as, time-of-use (TOU), and real time pricing tarrifs (Aalami et al., 2010). In this paper, the TOU tariff is studied as an instance of tariff-driven demand side management (DSM) of green ship. In the DSM, the HRES on a green ship can help owners to reduce electricity cost, and also can help utilities to enhance grid security and efficiency. Tariff-driven DSM of on-grid ship is more complicated than usual power management of off-grid ship, as demand-side management is required to consider the variance of electricity price and incentive reward, as well as the variance of renewable generation and load demand. One challenge of tariff-driven power management is to find an optimal control strategy for consuming grid power at the low-price period, and for selling renewable energy at the high-price period, while physical constraints have to be satisfied. Another challenge is to integrate the new capability of tariff-driven DSM into existing power management systems, which mainly focused on the off-grid management. The green ships often switch between on-grid and off-grid modes, especially for short-route ships, such as ferry and cruise. For this purpose, these challenging problems will be responded in the tariff-driven power management of green ship.

The contributions of this paper include three aspects. Firstly, tariff effects are studied for the power management of green ship with HRES, which is formulated as an optimal power dispatching problem to minimize the operational cost. Secondly, a unified tariff-driven power management system for off-grid and on-grid modes is proposed to optimally schedule the ship all the time. Thirdly, receding horizon control is proposed in the green ship application, so that system disturbances on solar energy and load demand can be detected and corrected. The resulted performance is promising with respect to energy efficiency and robustness. This paper is organized as follows. A HRES is introduced for the green ship in Section 2. Optimal power management problem of off-grid green ship is formulated in Section 3. A tariff-driven power management model is proposed in Section 4. Receding horizon control is proposed to control power flows for the minimization of operational cost in Section 5. Results and discussions are presented in Section 6, while the last section is the conclusion.

2. Hybrid renewable energy system of green ship

PV-DG-battery (PDB) hybrid systems are successfully applied to green ships (Banaei and Alizadeh, 2016; Tsekouras and Kanellos, 2013). The PDB system is made up of three main subsystems, i.e., PV panels, battery storage, and DG. The ship load includes propulsion load and auxiliary load. Auxiliary load consists of lights, water heating, air conditioners, plug-in devices, and other on-board hotel facilities. For the PDB hybrid system of green ship, the basic requirement is to keep the power balance, and to reduce operational cost and GHG emission.

Regarding to different volume and rated power, the hybrid electric ship can be categorized into two types. The first kind of ships, such as, bulk cargo vessels, which has large volume and rated power, only depends on DGs for the propulsion power. The solar energy is used to meet the hotel and auxiliary load, as shown in Fig. 1(a). The second kinds of ships, such as, cruises and ferries, usually have small volume and rated power. Both DG and solar energy are integrated to supply power for the propulsion load and auxiliary load, as shown in Fig. 1(b).

In this paper, we study the power management of a retrofitted green ship, which belongs to the first type, as shown in Fig. 1(a). The propulsion load is directly supplied by the DG. For the auxiliary load, the

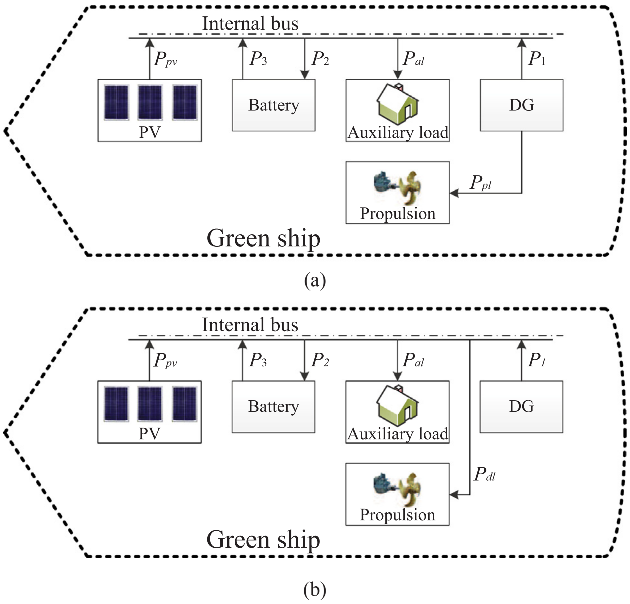


Fig. 1. Schematic of the off-grid PDB hybrid system on green ships: (a) DG propulsion and (b) electric propulsion.

PV power has the first priority of usage, and the battery takes part in the power supply when the PV output is not enough to meet the auxiliary load. Only when both PV and battery cannot meet the ship load, the DG eventually comes in due to its highest cost.

Note that there is an internal bus in the hybrid electric ship, as shown in Fig. 1. The shore-side grid can be connected with the ship internal bus for the cold ironing. The propulsion load is denoted as P_{pl}, and the auxiliary load is denoted as P_{al}. The power flows from the DG, battery and PV to the bus are denoted as P_1, P_3 and P_{pv}, respectively. P_2 represents the power flow from the bus to battery. The subsystems, i.e., PV, DG, and battery, are introduced as follows.

2.1. PV panel

A solar panel usually consists of several solar cells to convert solar irradiation into direct current power. In the application of green ship, the PV panels installed in different parts of ship can be categorized as different groups, e.g., the PV panels installed on the top deck, the lower deck, the vertical surface, and some discontinuous space. These groups may have different irradiance and shading characteristics during the long-term voyage. The power output of each PV panel can be simply formulated as:

$$P_{pv}(t) = \eta_{pv} I_{pv}(t) A_c, \tag{1}$$

where t is the time of day; P_{pv} is the power output from the PV panel; η_{pv} is the efficiency of solar generation; I_{pv} is the solar irradiation incident on the PV panel; A_c is the size of PV panel.

The hourly solar irradiation incident on the PV panel has complicated relations with time of a day, season of a year, tilt, location, global irradiation, and diffuse fraction. In this study, the simplified isotropic diffuse formula is used according to Tazvinga et al. (2014) and Collares-Pereira and Rabl (1979). The solar irradiation incident can be expressed as

$$I_{pv}(t) = [I_B(t) + I_D(t)]R_B(t) + I_D(t), \tag{2}$$

where I_B is the beam component of global irradiation, and I_D is diffuse irradiation. R_B is a geometric ratio of actual irradiation on the tilted plane to the standard irradiation on the horizontal plane.

The efficiency of solar generation can be expressed as a function of the irradiation I_{pv} and the ambient temperature T_A as

$$\eta_{pv} = \eta_R \left[1 - \frac{0.9\beta I_{pv}(T_{C0} - T_{A0})}{I_{pv0}} - \beta(T_A - T_R)\right], \tag{3}$$

where η_R is the PV generation efficiency that is measured at the referenced cell temperature T_R (25 °C); β is the temperature coefficient for cell efficiency (typically 0.004–0.005/°C); T_{C0} (typically 45 °C) and T_{A0} (typically 20 °C) are cell temperature and ambient temperature at the nominal operating cell temperature (NOCT) test, respectively; I_{pv0} is the average solar irradiation on the array at the NOCT test.

2.2. Diesel generator

Diesel generators are commonly used as engines in green ships. They are also incorporated in the PDB hybrid system to supply the auxiliary demand, when solar power and battery storage are insufficient. It is a common sense that the fuel consumption is determined by the power output. This relation is usually expressed as a quadratic model (Kanellos et al., 2017; Kanellos et al., 2014). The fuel consumption can be formulated as

$$\mu_{DG}(t) = d_1 P_D(t)^2 + d_2 P_D(t) + d_3, \tag{4}$$

where μ_{DG} is diesel consumption rate (the volume of diesel consumed per hour); $P_D(t)$ is the power output of DG; d_1, d_2, and d_3 are generation coefficients. When the power output is large, the DG efficiency is large (the fuel cost per kWh is small). According to Eq. (4), the hourly fuel cost can be calculated. DG's power output has to be restricted between the rated power and specified minimum value as

$$P_D^{min} \leqslant P_D(t) \leqslant P_D^{max}, \tag{5}$$

where P_D^{max} is the rated power and P_D^{min} is the minimum requirement of power output.

2.3. Battery bank

Many kinds of battery, such as Lead-acid, Nickel-based, and Lithium-ion cells, have been used in the PDB hybrid system. In general, the battery storage is closely related with maximum capacity and state of charge (SOC). Note that SOC is defined as the percent of remained storage.

The SOC could change dynamically due to possible charge or discharge. Let $S(t)$ denote the SOC of battery at time t, and S(0) denote the original SOC. The change of SOC can be formulated as

$$QS(t) - QS(0) = \eta_C \int_{\tau=0}^{t} P_2(\tau)d\tau - \frac{1}{\eta_D}\int_{\tau=0}^{t} P_3(\tau)d\tau, \tag{6}$$

where Q is the maximum capacity of battery; $P_2(t)$ is the power for charging the battery at time t;$P_3(t)$ is the power of discharge at time t. The first component at the right-hand side means the total energy stored to the battery, and the second component means the total energy consumed. $\eta_C \leqslant 1$ and $\eta_D \leqslant 1$ are charging efficiency and discharging efficiency (Wu and Xia, 2015; Wu et al., 2017). The charging/discharging loss comes from the heat loss of cells and converters.

By the differentiation at both sides of Eq. (6), the dynamics of SOC can be expressed as

$$\dot{S}(t) = \frac{\eta_C}{Q}P_2(t) - \frac{1}{Q\eta_D}P_3(t). \tag{7}$$

The battery has strict constraints on the upper and lower bounds of SOC. The upper bound is defined as S^{max}, and the lower bound is defined as S^{min} in this paper. The SOC must be bounded within the scale $[S^{min}, S^{max}]$.

3. Power management of off-grid mode

For the voyaging ship, how to minimize the fuel consumption for each day is a critical issue of power management, which is referred to power flow dispatching. Optimal dispatching will be studied to determine daily schedule of PDB hybrid system for minimizing the fuel cost. The daily fuel cost is formulated as

$$C_1 = p \sum_{k=0}^{N-1} [d_1 P_D^2(k) + d_2 P_D(k) + d_3], \tag{8}$$

where N denotes the evaluation period. The sampling period is an hour for instant, so $N = 24$ for a day. Note that the sampling period can be determined by users. C_1 is the fuel cost over the evaluation period; p is the fuel price. $P_D(k)$ is the diesel's power output over the period [k, k+ 1), which can be expressed as

$$P_D(k) = P_{pl}(k) + P_1(k), \tag{9}$$

where $P_{pl}(k)$ is the propulsion load over [k, k+ 1).

Furthermore, each component of PDB hybrid system suffers from continuous wearing over the rated lifetime (Tazvinga et al., 2015). According to (Wu et al., 2015; Yang and Xia, 2017), the daily wearing cost of system can be simplified as

$$C_2 = \tau_1 \sum_{k=0}^{N-1} [P_2(k) + P_3(k)] + N\tau_2, \tag{10}$$

where the first component is the wearing cost of battery, and the second component is the wearing cost of other subsystems, such as DG and solar panel. τ_1 is the coefficient of battery wearing, and τ_2 is the hourly wearing cost of other components. ($\tau_1 = 0.001$ and $\tau_2 = 0.002$ in the studied system.) Note that the first component can indicate the amount of charging/discharging cycle, as the battery usually works in full cycles due to SOC boundary. In the second component, we assume the constant wearing cost, as the wearing rate rarely changes for a given transportation task, e.g., the fixed propulsion load and the fixed frequency of start/stop.

Considering fuel cost and wearing cost, the objective of optimal power flow dispatching is to minimize off-grid operational cost J_f as

$$J_f = C_1 + C_2. \tag{11}$$

For the application of green ship, several physical and operational constraints have to be satisfied.

(1) Power balance constraint: The PV power, battery power, and possible DG power output must exactly match the auxiliary demand P_{al}. Power imbalance may harm all electric components in the PDB system. The power balance can be formulated as

$$P_1(k) + P_3(k) + P_{pv}(k) = P_2(k) + P_{al}(k). \tag{12}$$

(2) DG output constraint: The DG power output must be less than the rated power and larger than the specified minimum.

$$P_D^{min} \leqslant P_{pl}(k) + P_1(k) \leqslant P_D^{max}. \tag{13}$$

(3) Power flow constraint: For safety and other physical reasons, power flow on each line must be bounded by a maximum value as

$$0 \leqslant P_i(k) \leqslant P_i^m, \quad i = 1, 2, 3, \tag{14}$$

where P_i^m is the allowable maximum power delivered on the ith line.

(4) SOC boundary constraint: During charging or discharging, SOC has the upper and lower bound for ensuring state of health.

$$S^{min} \leqslant S(k) \leqslant S^{max}. \tag{15}$$

(5) SOC terminal state constraint: For the convenience of daily dispatching power, the terminal SOC of battery must be no less than the initial SOC as

$$S(0) \leqslant S(N). \tag{16}$$

For the off-grid mode, the power flow dispatching problem is modeled as a standard quadratic programming problem with equality and inequality constraints. In this optimization model, the objective function is (11), and the constraints include (12)–(16). The control variables are $P_1(k)$, $P_2(k)$, and $P_3(k)$ for each hour. Like other off-grid power flow dispatching models (Kanellos et al., 2017; Kanellos et al., 2014), PV output and auxiliary load are assumed known as priory knowledge in this study. Short-term deviations can be taken over by the ship real time control system through certain adjustment mechanisms (Kanellos et al., 2014). In a day, load demand and PV power are series of data indexed by time, which can be forecasted based on historical data. Time series analysis models, including autoregressive (AR) (Powell et al., 2014) and neural networks (Bacher et al., 2009; Mellit and Pavan, 2010; Suganthi and Samuel, 2012), have been studied for short-term and long-term forecast of future PV and load profiles. Note that electricity tariffs have no effects on power management for the off-grid mode.

4. Tariff-driven power management

When the green ship has stopped in the harbor for loading, unloading, or maintenance, the propulsion load is zero, and the green ship has the grid connection. Electricity tariffs at harbor have great effects on the solution to power management. For this on-grid mode, power management requires a tariff-driven power dispatching method, and the off-grid dispatching method is not able to suit the on-grid mode. The structure of on-grid green ship is given in Fig. 2. In this paper, the TOU tariff is considered as a typical incentive policy for studying the tariff-driven power management.

In the TOU tariff, electricity price changes over different periods according to the imbalance situation between power supply and demand. For example, a high price is paid for the peak load period; a medium price is paid for the standard period; and a low price is paid for the off-peak period. In this study, electricity price at the target harbor is

$$\rho(t) = \begin{cases} \rho_k, & t \in T_k, \\ \rho_o, & t \in T_o, \\ \rho_s, & t \in T_s, \end{cases} \tag{17}$$

where ρ_k is the price of peak load period T_k; ρ_o is the price of off-peak period T_k; ρ_s is the price of standard period T_s.

Let P_4 denote the bi-directional power flow between the grid and green ship. Define $P_4 > 0$ when the grid power flows to the ship, and $P_4 < 0$ when the ship supplies power to grid. It can be noticed that the role of ship, as load or distributed generation, determines the sign of P_4. For safety, the bi-direction power flow has to be bounded as

$$-P_4^m \leqslant P_4(k) \leqslant P_4^m, \tag{18}$$

where P_4^m is the allowable maximum of power flow on this connection.

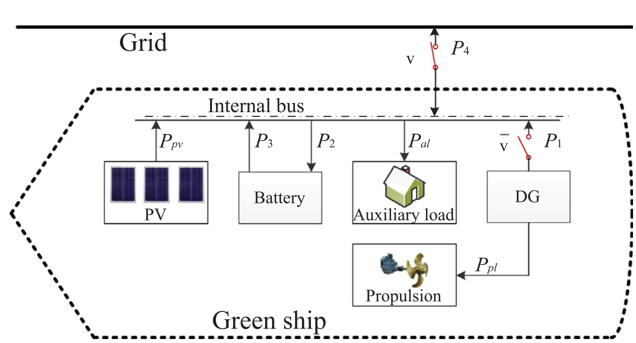


Fig. 2. A unified structure of green ship.

The daily cash flow, associated with buying and selling electricity, can be formulated as

$$C_3 = \sum_{k=0}^{N-1} \rho(k)P_4(k), \tag{19}$$

where C_3 represents the daily cash flow driven by the TOU tariff. Note that $C_3 > 0$ means cash-out, i.e., electricity cost. $C_3 < 0$ means cash-in, i.e., electricity reward. For the on-grid situation, the minimization of daily cash flow is expected in the tariff-driven power management.

The capability of tariff-driven dispatching will be integrated in the existing power management system. In this paper, a unified dispatching model will be studied to handle both off-grid and on-grid modes in an automatic manner. For the unified model, power flow dispatching has to consider the optimization of fuel cost, wearing cost, and possible electricity cost caused. The optimization problem is closely related with the status of switch. Based on the on/off status, the objective function can be expressed as

$$J_u = C_1 + C_2 + vC_3. \tag{20}$$

where J_u is the daily cost of the unified model. v is the status of switch. $v = 0$ means off-grid mode, and $v = 1$ means on-grid mode. C_1 is the fuel cost expressed as Eq. (8); C_2 is the wearing cost expressed as Eq. (10); C_3 is the electricity cost expressed as Eq. (19). If $v = 1$, grid connection is introduced, and tariff-driven power management is enabled.

Based on the status of switch, constraints, e.g., power balance and boundary, also have to be re-formulated as

$$\begin{cases} P_1(k) + P_3(k) + vP_4(k) + P_{pv}(k) = P_2(k) + P_{al}(k), \\ (1-v)P_D^{min} \leqslant (1-v)P_{pl}(k) + P_1(k) \leqslant (1-v)P_D^{max}, \\ 0 \leqslant P_1(k) \leqslant P_1^{max}, \\ 0 \leqslant P_2(k) \leqslant P_2^{max}, \\ 0 \leqslant P_3(k) \leqslant P_3^{max}, \\ -P_4^m \leqslant vP_4(k) \leqslant P_4^m, \\ S^{min} \leqslant S(k) \leqslant S^{max}, \\ S(0) \leqslant S(N), \end{cases} \tag{21}$$

For the unified model, three main characteristics are essential. Firstly, when the switch is off, the model must be equivalent with the proposed off-grid dispatching model. The optimal solution ensures minimal fuel cost and wearing cost during the voyage. Secondly, when the switch is on, the surplus of PV power can be sold to grid, and the hybrid system serves as a role of distributed generation. This could help release the peaking burden of grid, and earn possible incentive reward that depends on the policy of harbor. Thirdly, for taking advantage of incentive policies, battery can store the grid power at the off-peak time, and can be discharged during the peak time.

As a result, the unified model successfully covers off-grid and on-grid modes. In Fig. 2, two modes are changed by a switch V. The green ship is on-grid if the switch V is on (i.e., $v = 1$, $\bar{v} = 0$). If V is off (i.e., $v = 0$, $\bar{v} = 1$), the green ship is off-grid. For the off-grid situation, the system structure is the same as Fig. 1(a). In a unified model, the status of V is a variable detected in real time, and then the optimization of power flow is re-conducted periodically. If $v = 0$, it is obvious that the unified model is equivalent with the off-grid model studied in the previous section. If $v = 1$, the unified model can reflect all essential characteristics of the on-grid mode.

5. Receding horizon control

Based on current status, optimal schedule over next 24 h can be obtained via the optimization of the unified model. However, the status of switch and the SOC of battery could change over next 24 h due to uncertain solar generation and traveling time. This happens when the green ship is about to approach the harbor or to leave. Then, the original scheduling results cannot be used to control the PDB hybrid system. For this purpose, receding horizon control is proposed based on periodic optimization, as shown in Fig. 3.

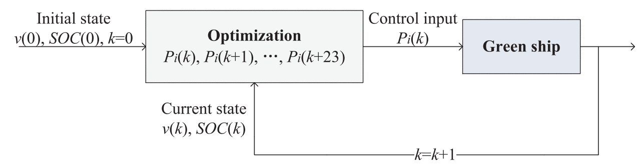


Fig. 3. Illustration of receding horizon control.

In the proposed receding horizon control, the optimization proceeds iteratively to utilize the real-time feedback information, i.e., the SOC and the status of switch. For each time, only the first component of optimal solution is employed to control the hybrid system. For example, a voyaging green ship will arrive at the harbor and connect to grid after 20 h. After detection of current state, the unified optimization model is an off-grid model, and the daily schedule of off-grid ship can be obtained by the optimization over the prediction horizon, i.e., 24 h. The first component of optimal solution is the power flow for the 1st hour, which is employed as the control input. The SOC may be changed due to possible charging or discharging. After 1 h, the switch is still off, and the same procedure is repeated, until the green ship is connected to grid. After 20 h, the status of switch is on. The unified optimization model is an on-grid model, and optimal dispatching can be obtained by the optimization of operational cost. The first component of optimal solution is the power flows for the 21th hour. This procedure is repeated till any stopping criterion is satisfied.

The procedure of receding horizon control for the green ship has been given in Algorithm 1. An optimization problem over the prediction horizon is repeatedly solved ($k = 0, 1, \ldots$). The optimization variable is the power flows over the following N intervals. At the kth sample, based on current states detected, an optimal solution denoted as $[\widetilde{P}_i(k|k), \widetilde{P}_i(k+1|k), \ldots, \widetilde{P}_i(k+N-1|k)]^T$ can be obtained. Only the first component of solution, i.e., $\widetilde{P}_i(k|k)$, is used as the control input over $[k, k+1)$. Note that the receding horizon control has the mechanisms of feedback and real-time control.

Algorithm 1. Pseudo-code of the receding horizon control approach

```
Set k = 0;
while the stopping criterion is not satisfied do
    Detect the SOC and the status of switch;
    Minimize the objective function (20) subject to
    constraint (21);
    For the optimal solution, apply P̃i(k|k) to the
    system at the period [k, k+1);
    k = k+1;
end
```

Receding horizon control is also called model predictive control (MPC) (Xia et al., 2011; Zhang and Xia, 2011). The key concept of receding horizon control is that control variables are calculated by using the optimization approach, but only the first component is taken as the control input at the current stage. As the optimization is conducted based on the current observation of state variables, state feedback is inherently incorporated in the receding horizon control. For the next interval, the prediction over the receding horizon is recalculated. As the close-loop control is implemented based on real-time updated information, the disturbance can be detected and corrected in the proposed approach.

In the receding horizon control, each optimization problem is a quadratic programming problem. Let $u(k) = [P_1(k), P_2(k), P_3(k), P_4(k)]$ denote the control inputs. Then the minimization of function (20) can be converted into a standard form of quadratic programming as

$$\min \frac{1}{2}(U^T * H * U + f * U), \tag{22}$$

where $U = [u(k), u(k+1), ..., u(k+N-1)]^T$. H and f are parameters that can be deduced according to (20).

For power flow dispatching, there are mainly two types of methods, i.e., ruled based and optimization-based methods. The proposed receding horizon control is an optimization-based method. For the purpose of comparison, a rule-based control is referred to fulfilling the satisfaction of constraints. In the rule-base control, the solar power has the highest priority of usage. The solar power is employed for satisfying the load demand or charging the battery. If the load demand cannot be satisfied by the solar power, the battery power is used. If the batter is over-discharged, the grid power or diesel is then integrated. For time t, the control input is decided as the following steps:

(1) If $P_{al}(t) \leqslant P_{pv}(t)$, $P_1(t) = 0$ and $P_3(t) = 0$. In this case, $P_2(t) = P_{pv}(t) - P_{al}(t)$ and $P_4(t) = 0$ if $S(t) < S^{max}$; otherwise $P_2(t) = 0$ and $P_4(t) = P_{pv}(t) - P_{al}(t)$.
(2) If $P_{al}(t) > P_{pv}(t)$ and $S(t) > 0.7S^{max}$, then $P_1(t) = 0$, $P_2(t) = 0$, $P_3(t) = P_{al} - P_{pv}(t)$, and $P_4(t) = 0$.
(3) If $P_{al}(t) > P_{pv}(t)$ and $S^{min} < S(t) \leqslant 0.7S^{max}$, then $P_2(t) = 0$, and $P_3(t) = 0$. In this case, $P_1(t) = 0$ and $P_4(t) = P_{al} - P_{pv}(t)$ if $v(t) = 1$, otherwise $P_1(t) = P_{al} - P_{pv}(t)$ and $P_4(t) = 0$.
(4) If $P_{al}(t) > P_{pv}(t)$ and $S(t) < S^{min}$, then $P_2(t) = S^{min} - S(t)$, and $P_3(t) = 0$. In this case, $P_1(t) = 0$ and $P_4(t) = P_{al} - P_{pv}(t) + P_2(t)$ if $v(t) = 1$, otherwise $P_1(t) = P_{al} - P_{pv}(t) + P_2(t)$ and $P_4(t) = 0$.

6. Results and discussions

6.1. Experimental results

A certain hybrid electric green ship with maximum power 500 kW is evaluated in this section. Note that the studied ship has been properly designed for matching its rated volume and power. Some advanced methods, such as optimal sizing and economic analysis (Arun et al., 2009; Lan et al., 2015; Yang et al., 2009), can be considered at the design stage of new green ships. As the scope of this paper is the power management for scheduling the operation of green ship, the issues on system design are excluded in this study.

Voyage tests at the ocean area of South Africa, are reported in the paper. Note the voyage schedule is calculated via other motion planning methods, while traveling constraints must be satisfied in the resulted voyage schedule. For the given route, the operational cost of ship will be evaluated under different seasons (summer vs. winter) and weather (sunny vs. cloudy). The structure of PDB hybrid system is the same as Fig. 2. In the application, the PV panels are installed in different parts of the ship, i.e., top deck, lower deck, vertical surface, and other discontinuous space. For the PDB hybrid system, configurations are mainly introduced here.

The storage bank consists of 272 Lithium-ion batteries. 4 batteries are serially connected as a set, and 68 sets, connected in parallel, form the bank. For each battery, the voltage is 12 V, and the capacity is 150 Ah. Therefore, the nominal capacity of storage bank is 489.6 kWh. The PV module consists of 240 PV panels, each of which has the capacity 250 W, so the rated PV output is 60 kW. The maximal power point tracking is integrated in each PV adapter. AC/DC and DC/AC inverters are also employed for each line. The parameters of this system are listed in Table 1. Note that charging and discharging efficiency are regarded as 85% and 95% in this paper for the target system. During the lifetime, energy efficiency may decrease due to system performance deterioration. This factor of efficiency decrease will be evaluated in the discussion part.

For regular cruising, the propulsion load is 100 kW. For berthing, the propulsion load is 0 kW. the daily profiles of auxiliary load and PV power are regarded as the average values over the past week before the test day (July 28, 2017, Cape Town), as shown in Fig. 4. Given known profiles of auxiliary load and PV output, optimal power dispatching can be obtained in the proposed control approach. Note that actual profiles of the test day could have small differences with the average profiles, differences will be corrected in the proposed receding horizon control, as evaluated in the discussion part.

Remark 1. Although solar irradiation mainly depends on time of day, it changes intensively under different environment, such as, location, season, orientation, and weather. For specific environmental conditions, daily profiles of overall PV generation on the ship show certain periodical characteristics. Advanced prediction methods can ensure promising accuracy, when environment change is trivial.

Remark 2. In some simple operating situations, ship load and PV output are fixed and known. For example, propulsion and axillary load is the same as historical days. The proposed model can deliver stable performance of minimal fuel cost and GHG emission. For complicated situations, the propulsion load is determined by the mass of ship and cargo, and auxiliary load is time-varying due to human behavior and external environment. Thus, system identification methods, such as model-based and data-driven methods, are required to determine propulsion and auxiliary load.

As the focus of this paper is system model and receding horizon control, the profiles of auxiliary load and PV power are regarded as the average historical values for simplicity. More advanced forecast methods (Kanellos et al., 2017; Kanellos et al., 2014; Powell et al., 2014; Li et al., 2018) can be utilized as preliminary steps of the proposed approach.

For the 24-h and 3-day tests, the TOU tariff (denoted as TOU-1) is

$$\rho_1(t) = \begin{cases} 0.157, & t \in [7, 10) \bigcup [18, 20), \\ 0.077, & t \in [0, 6) \bigcup [22, 24), \\ 0.113, & t \in [6, 7) \bigcup [10, 18) \bigcup [20, 22), \end{cases} \tag{23}$$

(1) *24-h test*
The 24-h test is conducted for both off-grid and on-grid modes. For each mode, optimal power flows obtained in the receding horizon control are plotted in Fig. 5. For the off-grid mode, the DG is the main power supplier. Battery is discharged at midnight, and is charged when the solar irradiation is sufficient at noon. For the on-grid mode, the grid is the main power supplier. However, battery is charged at midnight due to low electricity price, and discharged for selling electricity at the peak period. Although the DG power sometime decreases, i.e., the DG turns less efficient, energy efficiency of the whole system is improved. The reason is that the reduction of DG power is taken place by the cheap PV power or battery power.
The daily cost, including fuel cost and wearing cost, is evaluated.

Table 1
Parameters of PV-battery system.

Nominal battery capacity	489.6 kWh
Battery charging efficiency	85%
Battery discharging efficiency	95%
Initial SOC	60%
Minimum of SOC	40%
Maximum of SOC	100%
PV array's capacity	60 kW
fuel price	0.67 $/L
Rated power of diesel	500 kW
Regular cruising propulsion load	100 kW
Minimal output of diesel	5 kW
d_1	0.000036
d_2	0.1728
d_3	76.8

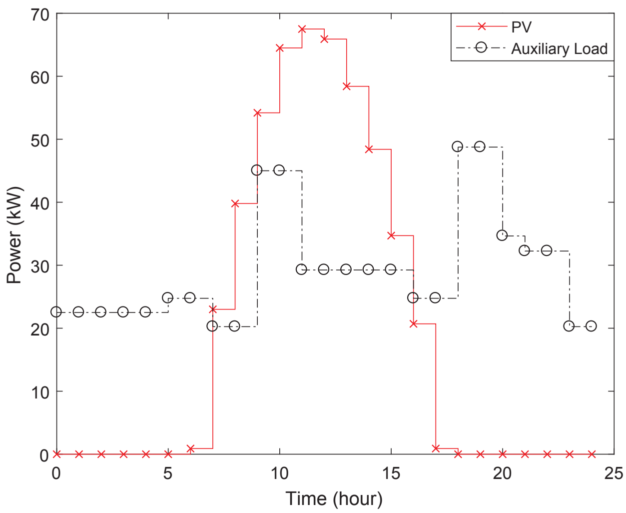


Fig. 4. Daily profiles of auxiliary load and solar energy in the test.

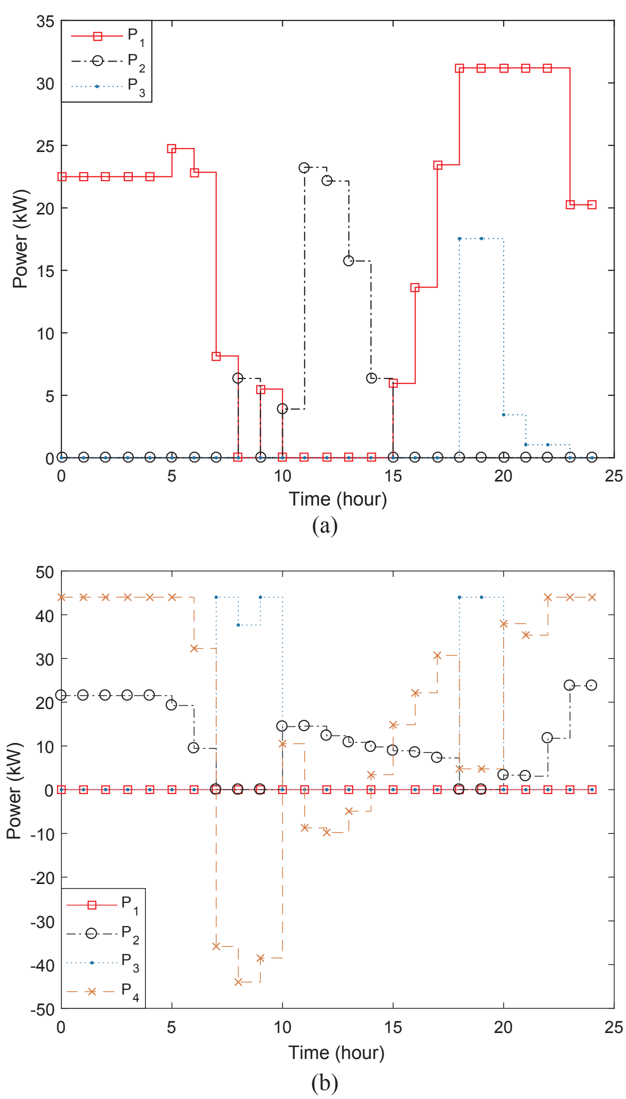


Fig. 5. Power flows of green ship in 24 h: (a) off-grid and (b) on-grid.

Without the integration of hybrid system, the daily cost is $1604.2. In the rule-based control, the daily cost is $1571.3. In the receding horizon control, the daily cost can be reduced to $1566.9. If the

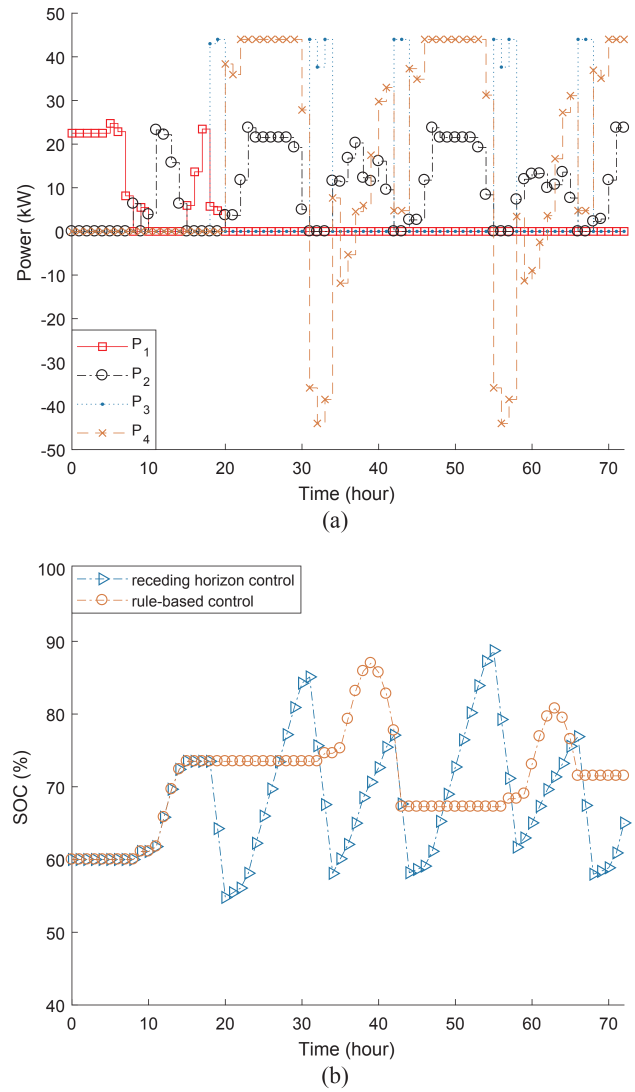


Fig. 6. Experimental results of green ship in 3 days: (a) power flows of receding horizon control and (b) SOC profiles.

Table 2
Daily cost under different environment.

	Cloudy winter	Sunny winter	Cloudy summer	Sunny summer
Off-grid cost ($)	1589.4	1566.9	1581.4	1557.9
On-grid cost ($)	55.63	29	47.37	13.05

Table 3
Daily cost under different charging efficiencies.

η_C	95%	85%	75%	65%	55%
Off-grid cost($)	1562.8	1563.7	1564.6	1565.5	1566.4
On-grid cost ($)	25.87	29.04	32.67	35.08	37.48

green ship is stopping in port with the grid connection, the daily cost includes electricity cost and wearing cost. Without the integration of hybrid system, the daily cost is $81.0. In the rule-based control, the daily cost is $47.2. In the receding horizon control, the daily cost can be reduced to $29.0. It can be noticed that the

Table 4
Daily cost under different discharging efficiencies.

η_D	95%	85%	75%	65%	55%
Off-grid cost ($)	1563.7	1564.5	1565.3	1566.1	1566.9
On-grid cost ($)	29.04	32.42	34.57	36.72	38.15

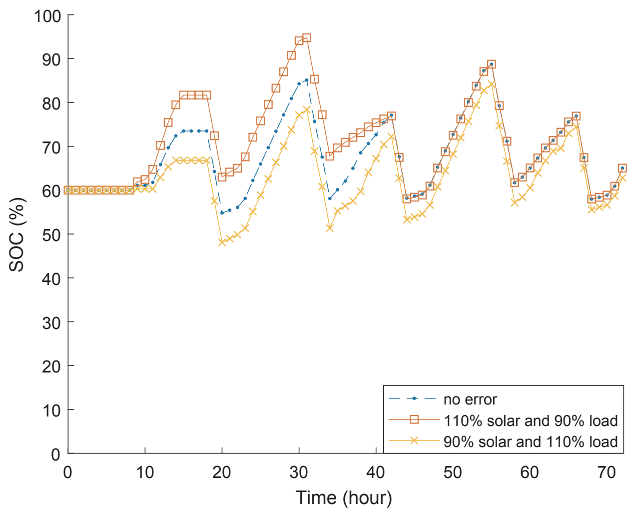


Fig. 7. SOC's sensitivity on forecast errors.

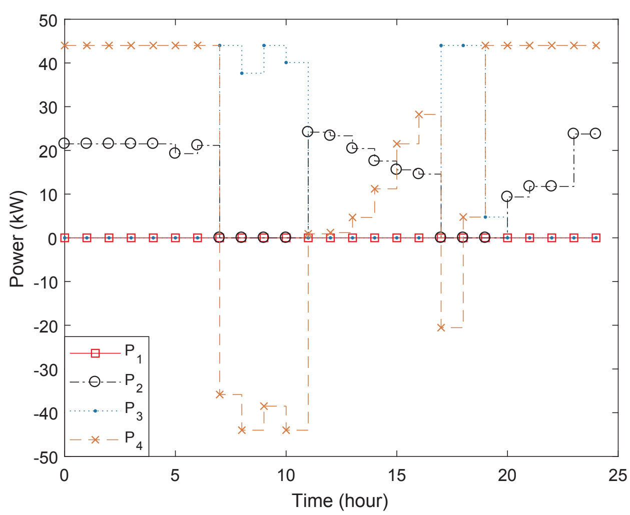


Fig. 8. Optimal power dispatching under the TOU-2 tariff.

integration of hybrid system can effectively reduce the expense of ship, and the receding horizon control can achieve the minimal cost.

(2) *3-day test*

The 3-day test is conducted to verify the power management for complicated situations. The change of off-grid and on-grid modes will be evaluated in a 3-day route. The green ship is off-grid at 0am of the first day, and gets the grid connection since 8 pm of the first day.

Based on the unified model, results of power flow and SOC are plotted in Fig. 6. Before the arriving time, the results of receding horizon control are similar with those of the 24 h off-grid experiment. The main power supplier is the DG, and the battery is charged at noon. In contrast, the results after arrival are similar with the 24 h on-grid experiment. The main power supplier turns to be the grid, and the battery is charged at midnight.

The overall cost over 3 days will be evaluated. Without the integration of hybrid system, the overall cost is $1519.7. With the help of hybrid system, the daily cost is $1392.5, and the SOC is 71.5% for the rule-based control. For the receding horizon control, the overall cost is $1368.4, and the SOC is 65%. The difference of residual energy in the bank, worth about $3, can be negligible. It is obvious that receding horizon control can result in an optimal strategy with the minimal cost.

It can be concluded that the unified model can effectively handle two different modes, and that the overall cost can be minimized by the receding horizon control regardless to the change of mode. If the harboring period is 2 months per year, the operational cost of green ship can be reduced by about $14300 per year. Compared with fuel ships without PV generation, fuel consumption and GHG emission of green ship can be reduced by about 3% for each year.

6.2. Discussions

The aforementioned results are reported based on tests during sunny winter days. However, the environmental change must influence the solar energy on the green ship, and the operational cost as well. Firstly, environmental effects on the green ship will be discussed in this part. Secondly, charging and discharging efficiency must change month by month due to system deterioration. Effects of varying parameters on the green ship will also be discussed. Thirdly, effects of forecast error are discussed, as it has influenced the control performance. At last, effects of different TOU tariffs are evaluated in the tariff-driven approach.

(1) *Effects of environmental conditions*

The green ship is tested on 4 kinds of environment, i.e., a sunny winter day (July 28, 2017, Cape Town), a cloudy winter day (August 3, 2017, Cape town), a sunny summer day (January 29, 2017, Cape Town), and a cloudy summer day (February 18, 2017, Cape Town). Different environmental conditions mainly influence daily profiles of PV power. Other parameters are assumed the same as listed in Table 1.

The daily cost under different environment is listed in Table 2. For a sunny summer day, solar energy generation is the largest, so the green ship has the smallest operational cost for each mode. For a cloudy winter day, solar energy generation decreases the most, so the operational cost is the largest. In the same season, solar generation on a sunny day is larger than a cloudy day, so the operational cost on a sunny day is smaller than a cloudy day. In the comparison of summer and winter, daily solar generation in summer is larger than winter, so the daily cost in summer is usually smaller than winter for the sunny and cloudy weather, respectively.

(2) *Effects of charging and discharging efficiency*

To test effects of charging efficiency, the charging efficiency is set as 95%, 85%, 75%, 65%, and 55%, respectively. The other settings are kept the same as listed in Table 1. Note that the initial SOC is 60% and the discharging efficiency is 95%.

For the receding horizon control, the daily cost under different charging efficiency is listed in Table 3. It can be observed that high charging efficiency is preferred to reduce the daily cost. When the battery gets old with low charging efficiency, the daily cost will increase. Especially for the on-grid ship, more reward can be earned when the charging efficiency is larger. It is suggested to retrofit a new battery when the charging efficiency is lower than 70%.

To test the discharging efficiency, the discharging efficiency is set as 95%, 85%, 75%, 65% and 55% respectively. The other settings are kept the same as listed in Table 1. Note that the initial SOC is 60% and the charging efficiency is 85%.

The daily cost under different discharging efficiency is given in Table 4. Effects of discharging efficiency are similar with those of charging efficiency. The cost increases, as the battery has relatively low discharging efficiency. A new battery is suggested for retrofitting when the discharging efficiency is lower than 80%.

(3) *Effects of forecast error*

To test effects of forecast error, the actual load demand is assumed as 90% of the forecast of load, and the actual solar power is assumed as 110% of the forecast of solar power at the first day. There is no forecast error in the next two days in this test. The SOC sensitivity on uncertain forecast errors is analyzed as shown in Fig. 7. When no forecast error exists, the SOC profile is a baseline for the sensitivity analysis. It can be observed that forecast errors of load and solar cause variance of SOC. More power is stored in the battery when the system has less load and more solar power than the predicted values. The SOC profiles keep close and converge in finite time, which can indicate the proposed receding horizon control has good robustness when the forecast errors exist.

In comparison, the actual load demand is also assumed as 110% of the load forecast, and the actual solar power is assumed as 110% of the forecast. When the load demand is 90% and the PV power is 110%, the electricity cost decreases to \$1356.1, because more solar power is stored in the battery and less grid power is consumed. When the load demand is 110% and the PV power is 90%, the electricity cost increases to \$1379.3, because the actual load demand is larger than the forecast value. Note that the electricity cost is \$1368.4 if forecast error is zero.

(4) *Effects of tariff change* To test effects of different tariff, another TOU tariff (denoted as TOU-2) is considered as

$$\rho_2(t) = \begin{cases} 0.132, & t \in [11, 17), \\ 0.065, & t \in [19, 24) \bigcup [0, 7), \\ 0.095, & t \in [7, 11) \bigcup [17, 19), \end{cases} \tag{24}$$

For the 24-h test of on-grid mode, the daily cost under TOU-1 tariff is \$29, but the daily cost under TOU-2 tariff is \$-125.4. In other words, the hybrid electric system can earn \$125.4 under TOU-2 tariff. Fig. 8 shows the optimal solution to power dispatching. Comparing Figs. 5(b) and 8, it can be observed that the optimal solutions under different TOU tariffs are also different, as peak/off-peak period and electricity price changes.

7. Conclusion

Considering effects of different tariff, power management of green ship, with the PDB hybrid system, is studied in the receding horizon control approach. Both the stand-alone and grid-connected modes are considered in a unified power flow dispatching model. The receding horizon control is proposed to iteratively optimize operational cost, including possible fuel cost, wearing cost, and electricity cost. Regardless of variant environmental conditions, optimal dispatching strategies of green ship can be obtained to reduce fuel consumption and GHG emission by about 3% per year.

Experimental results have indicated several conclusive points. Firstly, the green ship is an effective resource to join in demand side management. Under TOU tariffs, optimal power management of green ship can contribute energy efficiency improvement on shipping industry and electricity market. Secondly, the capability of tariff-driven dispatching is successful integrated in the unified model of power management. Two working modes, i.e., off-grid management and on-grid management, can be handled in an automatic way. Thirdly, receding horizon control is a robust approach to power management of green ship. With the feedback mechanism, forecast errors and other disturbance have been detected and corrected in the control, and the performance of energy efficiency and cost saving is lasting.

The green ship studied in this paper is a retrofitted ship with hybrid electrification. The proposed model can be extended to all-electric green ships as future work. Multiple generation resources, such as different kinds of distributed energy and energy storage systems, will be studied in the green ship in future.

Conflicts of interest

None.

Acknowledgement

This work is partially supported by the Fundamental Research Funds for the Central Universities, 106112016CDJRC000014, and 106112017CDJXY170003.

References

Aalami, H., Moghaddam, M.P., Yousefi, G., 2010. Demand response modeling considering interruptible/curtailable loads and capacity market programs. Appl. Energy 87 (1), 243–250.

Arun, P., Banerjee, R., Bandyopadhyay, S., 2009. Optimum sizing of photovoltaic battery systems incorporating uncertainty through design space approach. Solar Energy 83 (7), 1013–1025.

Bacher, P., Madsen, H., Nielsen, H.A., 2009. Online short-term solar power forecasting. Solar Energy 83 (10), 1772–1783.

Banaei, M.R., Alizadeh, R., 2016. Simulation-based modeling and power management of all-electric ships based on renewable energy generation using model predictive control strategy. IEEE Intel. Transport. Syst. Magaz. 8 (2), 90–103.

Buhaug, Ø, Corbett, J., Endresen, Ø, Eyring, V., Faber, J., Hanayama, S., et al., 2009. Second IMO CGC Study 2009. International Maritime Organization (IMO), London, UK.

Collares-Pereira, M., Rabl, A., 1979. The average distribution of solar radiation correlation between diffuse and hemispherical and between daily and hourly insolation values. Solar Energy 22, 155–164.

Diab, F., Lan, H., Ali, S., 2016. Novel comparison study between the hybrid renewable energy systems on land and on ship. Renew. Sustain. Energy Rev. 63, 452–463.

Esen, H., Inalli, M., Esen, M., 2007. A techno-economic comparison of ground-coupled and air-coupled heat pump system for space cooling. Build. Environ. 42 (5), 1955–1965.

Gabash, A., Li, P., 2013. Flexible optimal operation of battery storage systems for energy supply networks. IEEE Trans. Power Syst. 28 (3), 2788–2797.

Kanchev, H., Lu, D., Colas, F., Lazarov, V., Francois, B., 2011. Energy management and operational planning of a microgrid with a PV-based active generator for smart grid applications. IEEE Trans. Indust. Electron. 58 (10), 4583–4592.

Kanellos, F.D., Tsekouras, G.J., Hatziargyriou, N.D., 2014. Optimal demand-side management and power generation scheduling in an all-electric ship. IEEE Trans. Sustain. Energy 5 (4), 1166–1175.

Kanellos, F.D., Anvari-Moghaddam, A., Guerrero, J.M., 2017. A cost-effective and emission-aware power management system for ships with integrated full electric propulsion. Electr. Power Syst. Res. 150, 63–75.

Kökkülünk, K.Y.G., Parlak, A., Karakaş, A., 2016. Energy cost assessment of shoreside power supply considering the smart grid concept: a case study for a bulk carrier ship. Marit. Pol. Manage. 43 (4), 469–482.

Lan, H., Wen, S., Hong, Y.Y., Yu, D.C., Zhang, L., 2015. Optimal sizing of hybrid PV/diesel/battery in ship power system. Appl. Energy 158, 26–34.

Lee, K.J., Shin, D., Yoo, D.W., Choi, H.K., Kim, H.J., 2013. Hybrid photovoltaic/diesel green ship operating in standalone and grid-connected mode – experimental investigation. Energy 49, 475–483.

Li, Q., Wu, Z., Xia, X., 2018. Estimate and characterize pv power at demand-side hybrid system. Appl. Energy 218, 66–77.

Mellit, A., Pavan, A.M., 2010. A 24-h forecast of solar irradiance using artificial neural network: application for performance prediction of a grid-connected PV plant at Trieste, Italy. Solar Energy 84 (5), 807–821.

Nema, P., Nema, R., Rangnekar, S., 2009. A current and future state of art development of hybrid energy system using wind and PV-solar: a review. Renew. Sustain. Energy Rev. 13 (8), 2096–2103.

Ovrum, E., Bergh, T., 2015. Modelling lithium-ion battery hybrid ship crane operation. Appl. Energy 152, 162–172.

Palma-Behnke, R., Benavides, C., Lana, F., Severino, B., Reyes, L., Llanos, J., et al., 2013. A microgrid energy management system based on the rolling horizon strategy. IEEE Trans. Smart Grid 4 (2), 996–1006.

Powell, K.M., Sriprasad, A., Cole, W.J., Edgar, T.F., 2014. Heating, cooling, and electrical load forecasting for a large-scale district energy system. Energy 74, 877–885.

Shaahid, S., El-Amin, I., 2009. Techno-economic evaluation of off-grid hybrid photovoltaic-diesel-battery power systems for rural electrification in Saudi Arabia – a way forward for sustainable development. Renew. Sustain. Energy Rev. 13 (3), 625–633.

Soto, W.D., Klein, S., Beckman, W., 2006. Improvement and validation of a model for photovoltaic array performance. Solar Energy 80 (1), 78–88.

Suganthi, L., Samuel, A.A., 2012. Energy models for demand forecasting–a review. Renew. Sustain. Energy Rev. 16 (2), 1223–1240.

Tazvinga, H., Xia, X., Zhang, J., 2013. Minimum cost solution to photovoltaic-diesel-battery hybrid power systems for remote consumers. Solar Energy 96, 292–299.

Tazvinga, H., Zhu, B., Xia, X., 2014. Energy dispatch strategy for a photovoltaic-wind-diesel-battery hybrid power system. Solar Energy 108, 412–420.

Tazvinga, H., Zhu, B., Xia, X., 2015. Optimal power flow management for distributed energy resources with batteries. Energy Convers. Manage. 102, 104–110.

Tsekouras, G., Kanellos, F., 2013. Optimal operation of ship electrical power system with

energy storage system and photovoltaics: analysis and application. WSEAS Trans. Power Syst. 8 (4), 145–155.
Wen, S., Lan, H., Hong, Y.Y., Yu, D.C., Zhang, L., Cheng, P., 2016. Allocation of ESS by interval optimization method considering impact of ship swinging on hybrid PV/diesel ship power system. Appl. Energy 175, 158–167.
Wies, R., Johnson, R., Agrawal, A., Chubb, T., 2005. Simulink model for economic analysis and environmental impacts of a PV with diesel-battery system for remote villages. IEEE Trans. Power Syst. 20 (2), 692–700.
Wu, Z., Xia, X., 2015. Optimal switching renewable energy system for demand side management. Solar Energy 114 (0), 278–288.
Wu, Z., Tazvinga, H., Xia, X., 2015. Demand side management of photovoltaic-battery hybrid system. Appl. Energy 148, 294–304.
Wu, Z., Ling, R., Tang, R., 2017. Dynamic battery equalization with energy and time efficiency for electric vehicles. Energy 41, 937–948.
Xia, X., Zhang, J., Elaiw, A., 2011. An application of model predictive control to the dynamic economic dispatch of power generation. Control Eng. Practice 19 (6), 638–648.
Yang, F., Xia, X., 2017. Techno-economic and environmental optimization of a household photovoltaic-battery hybrid power system within demand side management. Renew. Energy 108, 132–143.
Yang, H., Wei, Z., Lou, C., 2009. Optimal design and techno-economic analysis of a hybrid solar–wind power generation system. Appl. Energy 86 (2), 163–169.
Yao, C., Chen, M., Hong, Y.Y., 2017. Novel adaptive multi-clustering algorithm-based optimal ESS sizing in ship power system considering uncertainty. IEEE Trans. Power Syst. (99), 1.
Zahedi, B., Norum, L.E., 2013. Modeling and simulation of all-electric ships with low-voltage dc hybrid power systems. IEEE Trans. Power Electron. 28 (10), 4525–4537.
Zhang, J., Xia, X., 2011. A model predictive control approach to the periodic implementation of the solutions of the optimal dynamic resource allocation problem. Automatica 47 (2), 358–362.

Applied Energy 227 (2018) 180–189

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

A power dispatch model for a ferrochrome plant heat recovery cogeneration system

Lijun Zhang[a,*], Michael Chennells[a,b], Xiaohua Xia[a]

[a] Department of Electrical, Electronic and Computer Engineering, University of Pretoria, 0002, South Africa
[b] Rustenburg Smelter a Glencore Merafe Venture Operation, South Africa

HIGHLIGHTS

- A power dispatching model is developed for a heat recovery cogeneration system.
- The model maximizes plant owner benefits considering power export to the grid.
- The cogeneration system designed generates both electrical and cooling power.
- Operation of furnaces is modeled to determine the waste heat available for recovery.
- Energy and cost savings obtained are used to evaluate feasibility of the system.

ARTICLE INFO

Keywords:
Waste heat recovery
Organic Rankine Cycle
Economic power dispatching
Optimal power flow

ABSTRACT

A Organic Rankine Cycle waste heat recovery cogeneration system for heat recovery and power generation to relieve grid pressure and save energy cost for a ferrochrome smelting plant is investigated. Through the recovery and utilization of previously wasted heat from the facility's internal smelting process off-gases, the cogeneration system is introduced to generate electrical power to supply the on-site electricity demand and feed electricity back to the utility grid when it is necessary and beneficial to do so. In addition, the cogeneration system generates cooling power through a lithium bromide-water solution absorption refrigeration cycle to meet the cooling requirements of the plant. The heat recovery process for power generation is modeled and the optimal power dispatching between the on-site loads and the utility grid is formulated as an economic power dispatching (EPD) problem, which aims to maximize the plant's economic benefits by means of minimizing the cost of purchasing electricity from the utility and maximizing revenue from selling the generated electricity to the grid. Application of the developed model to a ferrochrome smelting plant in South Africa is presented as a case study. It is found that, for the studied case, more than $1,290,000 annual savings can be obtained as a result of the proposed heat recovery power generation system and the associated EPD model. In addition to this, more than $920,000 annual savings is obtained as a result of the generated cooling power via the proposed absorption refrigeration system. The combined cogeneration system is able to generate up to 4.4 MW electrical power and 11.3 MW cooling power from the recovered thermal energy that was previously wasted.

1. Introduction

The world is in the midst of an energy crisis where a limited energy generation capacity is struggling to keep up with a continuously increasing demand for energy. This is particularly the case in South Africa. It has therefore never been more crucial to look towards and embrace renewable energy resources and new energy technologies to aid in the alleviation of this energy crisis. In conjunction with technology development, the recovery and utilization of waste energy have shown significant potential in the management of this crisis by introducing considerable energy savings [1,2]. One such energy saving opportunity exists in the mining and smelting industry, for example in the ferrochrome (FeCr) industry, in the form of furnace off-gas thermal energy recovery.

It was estimated that around 80% of the world's chromium deposits can be found in the Bushveld Complex in South Africa, which spans an estimated cumulative diameter of almost 300 km [3,4]. Because of the sheer size of the area and the overwhelming deposits of precious metals, such as chromium, in the Complex rock, the mining and smelting of these metals form a vital and influential sector of South Africa's economy [4].

* Corresponding author.
E-mail address: lijun.zhang@up.ac.za (L. Zhang).

http://dx.doi.org/10.1016/j.apenergy.2017.08.019
Received 15 January 2017; Received in revised form 1 August 2017; Accepted 7 August 2017
Available online 12 August 2017
0306-2619/

Nomenclature

AD	maximum installed access demand for consumption in MVA
CAC_r	consumption administration charge rate in \$/day
COP	coefficient of performance of the cooling system
CRC	consumption reliability charge in \$
CRC_r	consumption reliability charge rate in \$/kWh
CSC_r	consumption service charge rate in \$/day
C_{ph}	specific heat of hot material in kJ/kg·K
DLF	distribution loss factor
E	electrical power generated by the cogeneration system in MW
ERC_r	electricity and rural subsidy charge rate in \$/MWh
GAC_r	the generation administration charge rate in \$/day
GRC	generation reliability charge in \$
GRC_r	generation reliability charge rate in \$/kWh
GSC_r	the generation service charge rate in \$/day
NAC_r	consumption network access charge rate in \$/MVA
NDC_r	network demand charge rate in \$/MVA
$P_{i,j}^{load}$	active power consumption of the plant, including consumptions of furnaces and induced draft fans, in MW
Q_h^k	heat transfer of the k-th furnace in kW
$Q_{cool,cold}$	cooling power generated by the cooling system in MW
$Q_{cool,low}$	available low temperature power in MW
$Q_{h,total}$	total extracted heat in MW
$S_{i,j}^{load}$	apparent power consumption of the plant, including consumptions of furnaces and induced draft fans, in MVA
TLF	transmission loss factor
TNC_r	transmission network charge rate in \$/MVA
T_{cold}	temperature of hot material outlet from heat exchanger in °C
T_{hot}	temperature of hot material inlet to heat exchanger in °C
$ULVSC_r$	the urban low voltage subsidy charge rate in \$/MVA
$\%_{Cr_2O_3,s}$	the mass percentages of the Cr_2O_3 in a dry sample of the ore
$\%_{FeO,s}$	the mass percentages of the FeO in a dry sample of the ore in kg/s
$\%_{H_2O}$	the required moisture percentage in the feed ore to a furnace
η_{ne}	net efficiency of the ORC electricity generation system
gAD	maximum installed access demand for generation in MVA
$gNAC_r$	the generation network access charge rate in \$/MVA
m_h^k	hot material mass flow rate of the k-th furnace in kg/s
m_{CO_2}	the mass flow rate of CO_2 extracted from the off-gas of a furnace in kg/s
$m_{Cr_2O_3}$	the mass flow rate of Cr_2O_3 to a furnace in kg/s
m_{FeO}	the mass flow rate of FeO to a furnace in kg/s
m_{N_2}	the mass flow rate of N_2 in the extracted hot material from a furnace in kg/s
m_{O_2}	the mass flow rate of O_2 in the extracted hot material from a furnace in kg/s
m_{ore}	the mass flow rate of the raw material ore to a furnace in kg/s
n	number of furnaces
p_p, p_s, p_o	the price for energy consumed in \$/MWh during peak, standard and off-peak periods, respectively
p_p^g, p_s^g, p_o^g	the price for energy sold in \$/MWh during peak, standard and off-peak periods, respectively

The smelting of chrome is an energy-intensive production process requiring approximately 3.3–3.8 MWh of electrical energy per ton of FeCr produced [5]. Of the country's 40 GW supply capacity, Ferro-Alloy smelting industries account for almost 5%, a staggering 2 GW of required power.[1] FeCr industries in South Africa have become severely constrained nowadays because of their high energy intensity and the increasing electricity price in the country. As a result, these industries need to seek solutions for more efficient utilization of the limited energy supply, which involves improving operational technologies and processes, and the potential recovery and re-use of wasted energy. Through such improvements, the efficiency of energy utilization can be improved and an overall improvement in the country's economy can be achieved by allowing the FeCr industries to be competitive on a global scale once again.

Various methods and techniques for increasing energy efficiency in the chrome smelting industry have been reported [6–8]. An important topic, the utilization of waste thermal energy for the generation of useful energy, has recently come under scrutiny.

The smelting processes of chrome involve the separation and fusion of materials according to process-specific chemical reactions inside a molten material bath in order to produce FeCr. The chemical processes and reactions require a carbonaceous reductant and extremely high temperatures for the extraction of iron (Fe) and chromium (Cr) metals from the raw feed material, which ultimately fuse to form FeCr [9,5,10].

The two most important furnace internal chemical reactions are therefore the reduction of iron and chromium oxides in the raw material, FeO and Cr_2O_3 respectively, to produce the Fe and Cr. A byproduct of the smelting process and the chemical reactions, along with heat, is carbon monoxide (CO) gas. Because of the open nature of the furnaces and the extremely high temperatures, the CO gas exiting the top of the furnace auto-ignites, using oxygen in the surrounding air to produce carbon dioxide (CO_2). The heat, CO_2 gas and dust particles thrown up from the raw material feed process are extracted from the furnaces and treated at the bagplant section of the facility. Currently, these off-gases are extracted by induced draft fans (ID fans) and passed through trombone coolers, which utilize vast surface areas and ambient temperature to cool the hot material. The cooled off-gasses then flow to the bagplant where they are combined with water and pumped to slimes dams for treatment.

Significant waste of energy occurs in the current cooling process because the thermal energy of the extracted hot material is simply dissipated into the atmosphere. The implementation of a cogeneration system instead of the trombone coolers will allow for the recovery and utilization of the wasted thermal energy for the generation of electricity. In the literature, many applications of waste heat recovery technologies to industrial processes have been published. For example, application of a waste heat recovery system to a company manufacturing large ship and offshore oil-platform chains was reported in [11], with the focus on determining the size of the main cogeneration equipment. A similar study on the recovery of multiple waste heat streams in a refinery was done by [12], in which the procedures for designing the heat recovery network were presented in detail. Only preliminary studies on the application cogeneration systems utilizing furnace off-gasses in FeCr smelting plants have been reported [13]. According to the literature, a waste heat recovery system is most suitable for implementation in a FeCr smelting industry that rejects heat from the furnaces at medium to high temperatures via the off-gas extraction system [14–16]. In addition to electricity generation, an absorption refrigeration cycle can be used to generate cooling power by utilizing the byproduct of the electricity generating system, low-grade thermal energy, which is traditionally directed to the power generation cycle cooling system [17,18]. Therefore, a combined cogeneration

[1] Rodney Jones. Electric Smelting in Southern Africa. http://www.mintek.co.za/Pyromet/Files/2013Jones-ElectricSmelting.pdf.

system is proposed for the recovery and utilization of thermal energy rejected from the smelting process for the generation of additional electrical and cooling energy in this study.

The facilities required for the cogeneration system are widely available today [19,20,17,21]; a waste heat recovery system using Organic Rankine Cycle (ORC) [22–25] as working fluid is identified as the most suitable waste heat recovery system for the specific application. A waste heat recovery cogeneration system from Turboden s.r.l. was considered for electricity generation and a lithium bromide-water solution absorption refrigeration system recommended by Voltas Technologies was adopted as the core equipment for cooling power generation. The interest of this study is, in particular, the optimal operation of this system when applied to FeCr smelting plants. Existing studies on heat recovery cogeneration systems either do not consider the power management of such facilities or only study stand-alone operations of such systems. No study on the operation optimization of such systems in a grid-tied environment has been reported so far. Lack of such operating strategies leads to poor performance of the system in terms of both operating efficiency and financial benefits to the plant owner. This is evidenced by many studies concluding that the proper operation and planning of the equipment and facilities are some of the key factors affecting the effectiveness of systems in both the industrial and residential sectors [26–31]. For instance, energy and associated cost savings were achieved by optimal operation of mining facilities, such as conveyor belts, crushers, coal washing plants and so on [32–40]. Moreover, existing studies on the application of heat recovery cogeneration systems to mineral processing plants, such as [41], are centered around the detailed modeling of the heat recovery efficiency instead of looking into the availability of the thermal energy for recovery and optimal operation of the cogeneration systems.

The heat recovery cogeneration system studied in this paper produces on-site electricity and cooling power supply for the plant and provides support and assistance to the utility grid by feeding the generated energy back into the grid during severely high demand periods. This not only helps the national grid but also enables the mine to obtain savings either through the substitution of electrical energy consumed directly from the utility grid or through the export of the generated electrical energy to the utility grid. Optimal operation of the proposed combined cogeneration system in such a grid-connected environment is a challenging task that depends heavily on the operating status of both the on-site smelting processes and the utility grid. The main function of the operation strategy is similar to the traditional economic power dispatching (EPD) problem [42,36,43–46] and the power flow management problem of hybrid renewable energy systems [47–51]. However, the EPD for the studied cogeneration system is even more challenging. Firstly, unlike the traditional power dispatching for power plants, the power generated by the cogeneration system is not controllable because it is directly affected by the process generating the waste heat. Secondly, the amount of waste heat available from the FeCr smelting plant is difficult to determine because of the chemical reactions involved. Therefore, this study focuses on the development of an EPD algorithm that optimizes the operation of an on-site cogeneration system tied to the national grid in order to maximize the benefits of the plant and help to relieve grid strain by feeding electricity back into the grid during peak demand periods. This will make an already implemented cogeneration system more efficient and help to reach the potential benefits of introducing a new cogeneration system to the plant from the plant owner's perspective. From the utility's point of view, the cogeneration system, together with its optimal operation strategy, helps to deal with peak demand and reducing the its generating costs.

The rest of this study is organized as follows: Section 2 presents the modeling process of the waste thermal energy carried in the furnace off-gas together with the efficiencies of the cogeneration system to determine its potential electricity and cooling power generation capacity. The EPD problem is formulated in Section 3. Section 4 provides a case study based on a real world mine in South Africa, followed by some further discussions on the results achieved in Section 5. Finally, conclusions are given in Section 6.

2. Modeling of the cogeneration system

In order to determine the electrical and cooling power generation capacity of the cogeneration system, the available heat from the internal material smelting process and the efficiencies of the cogeneration system must be determined first. A brief description the FeCr smelting process, the determination of the available waste heat for cogeneration, and modeling of the electrical and cooling power generation systems are given in the following subsections. In particular, the furnace process is modeled from first principles to determine the available heat for recovery. After that, the efficiencies of the electrical and cooling power generation facilities are estimated based on manufacture supplied information. Finally, an EPD model is developed to optimally control the power flows between the cogeneration system, the on-site load, and the utility grid in pursuit of maximizing the plant owners benefits.

2.1. Description of the FeCr smelting process

The FeCr smelting plant studied utilizes three-phase AC submerged-electrode arc furnaces for the smelting of raw materials to form a molten bath. This bath is tapped off from the furnaces at regular intervals throughout the day and separated into waste slag and molten FeCr using a density separator. After the cooling, crushing and treatment processes, the FeCr is stored in a range of rock sizes for dispatch. The product FeCr is used in the manufacture of stainless steel.

The furnaces operate at a temperature around 1500 °C. Hot dust and gas are extracted from the furnaces via extraction vents and stacks, at trend-based temperatures from 250 °C to 600 °C, which are determined by the operating conditions of the furnaces. The hot material is then transferred via the extraction ducts to a bagplant, where it is compressed into a fine powder and mixed with water to produce sludge. This sludge is pumped to slimes dams around the facility for treatment. In the existing system, an intermediate cooling process using trombone coolers is implemented between the furnaces and the bagplant to bring the temperature of the hot materials below the maximum temperature rating of the bags to ensure safety. It is proposed to implement a waste heat recovery cogeneration system, consisting of a heat exchanger for thermal energy recovery and a turbine generator for electricity generation, instead of the trombone coolers to generate electricity from the waste heat, while still performing the required cooling of the extracted hot material.

2.2. Calculation of available recovered heat

The calculation of the total available heat from the furnace off-gases requires the temperatures of the hot materials before and after the proposed heat recovery cogeneration system. The temperatures before and after the trombone coolers are used to determine the available heat per furnace and then combined into total available heat for the cogeneration system. The available heat from each furnace is calculated by:

$$Q_h^k = m_h^k C_{ph}(T_{hot} - T_{cold}), \tag{1}$$

where m_h^k is the sum of the flow rates of CO_2, N_2 and O_2 gasses determined according to the furnace feed receipt and the relevant chemical reactions.

The calculations of m_h^k begin with the calculation of the actual mass flow rates of FeO and Cr_2O_3 in the raw feed material by

$$m_{\mathrm{FeO}} = \%_{\mathrm{FeO},s} m_{ore}(1 - \%_{\mathrm{H_2O}}), \tag{2}$$

$$m_{\mathrm{Cr_2O_3}} = \%_{\mathrm{Cr_2O_3},s} m_{ore}(1 - \%_{\mathrm{H_2O}}). \tag{3}$$

Thereafter, the constitution of the off-gas is then calculated according to the following chemical reactions:

$$FeO + C = Fe + CO, \tag{4}$$

$$Cr_2O_3 + 3\ C = 2Cr + 3CO, \tag{5}$$

$$2CO + O_2 = 2CO_2. \tag{6}$$

Using Eqs. (4)–(6), the mass flow rate of CO_2, m_{CO_2}, in the hot material can be obtained. The total mass flow rate of hot materials m_h^k is then obtained by

$$m_h^k = m_{CO_2} + m_{N_2} + m_{O_2}$$

considering the excess air flows, consisting of N_2 and O_2 gases, caused by the operation of ID fans.

To account for the operational status of the system, such as the furnaces being off at certain time intervals for maintenance, potential faults in the temperature sensors and the minimum temperature required for the recovery of heat via the ORC cogeneration system, two vital assumptions are made to facilitate the estimation of the overall available heat. These assumptions, which apply to each furnace individually, are:

- If the furnace outlet extracted off-gas temperature is below 200 °C at a time instant, the actual furnace itself is assumed to be off, and the overall heat recovery cogeneration system will not consider this specific furnace during this time interval.
- If the measured bagplant inlet off-gas temperature is higher than the measured furnace outlet off-gas temperature, the heat recovery process cannot occur, and the overall heat recovery cogeneration system will once again not consider this specific furnace during this time interval.

In particular, the above assumptions are realized by means of the flow control of ID fans used to extract the off-gasses for each furnace, *i.e.* if a furnace is off, the flow rate of the corresponding ID fan will be set to zero. With the aforementioned assumptions, the overall extracted heat available for cogeneration can be calculated using (7).

$$Q_{h,total} = 0.001 \sum_{k=1}^{n} Q_h^k. \tag{7}$$

2.3. Systems for electrical and cooling power generation

For the specific application, Turboden s.r.l., an Italian leading company in the production and development of ORC heat recovery and turbo generator solutions,[2] was consulted and a specialised heat recovery ORC power generation system was recommended based on the information on the plant shown in Table 1.

The most appropriate working fluid selected by Turboden s.r.l. was Hexamethyldisiloxane. The proposed indirect exchange ORC heat recovery cogeneration system is shown in Fig. 1, in which the thermal energy is transferred from the furnace off-gases to the power generation ORC working fluid via the intermediate heat exchanger which utilizes thermal oil as the heat transfer medium. The ORC working fluid absorbs the heat transferred and is vaporized. The fluid vapor then expands through the turbine which drives an electric generator. The ORC working fluid in the vaporous phase that leaves the turbine passes through the regenerator component, where it is condensed utilizing the condenser and water cooling subsystems. Finally, the working fluid, pre-heated by an internal heat exchanger, cycles back at the required pressure by means of the flow control pressurizing pump and is passed back to the main heat exchanger where the cycle begins again.

Therefore, the power generation cycle in Fig. 1 produces electricity and low temperature heat through the closed thermodynamic cycle that

Table 1
Customer supplied and assumed system data descriptions.

Data description	Source	Data value	Unit
Thermal energy source	Customer	Smelting off-gases	–
Number of furnaces	Customer	4	–
Total exhaust gas flow rate	Customer	73.5	kg/s
Average exhaust gas temperature	Customer	413	°C
Minimum exhaust gas temperature	Customer	200	°C
Average air temperature (dry bulb)	Assumed	23	°C
Average cooling water temperature (tower water)	Assumed	30	°C
Grid voltage connection for unit	Assumed	Medium voltage	–

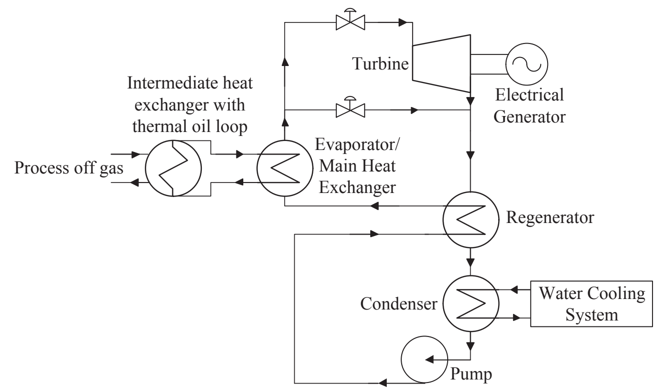


Fig. 1. Turboden s.r.l. indirect exchange ORC heat recovery cogeneration system.

enforces the working fluid to change as defined by the working fluid's characteristic ORC. From the evaluation performed by Turboden s.r.l. utilising the data and descriptions shown in Table 1, the proposed system and performance calculations are shown in Table 2. The net electricity produced by the Rurboden TD40 ORC unit is obtained by

$$E = \eta_{ne} Q_{h,total}, \tag{8}$$

where η_{ne} = net electrical output power/net available thermal power × 100% = 23% for the specific unit.

Traditionally, the low-temperature heat from the ORC cogeneration system is dissipated into the atmosphere via the cooling subsystem. The utilization of the absorption refrigeration cycle allows for the use of this low-grade heat for cooling and refrigeration applications, thereby further improving the overall system energy utilization efficiency. The amount of low-temperature thermal power available, 12.78 MW in the system design case (shown in Table 2), is calculated using

$$Q_{cool,low} = (100\% - \eta_{ne} - 2\%) Q_{h,total}. \tag{9}$$

The low-temperature thermal power calculated in (9) is in the form of hot water. This is because it is the water in the cooling subsystem that picks up the low-grade heat from the power generation cycle via the condenser component. For the suitable operation of an absorption refrigeration cycle, the cooling system fuel or supply heat must be in the form of hot water around 92 °C. Although the proposed power generation system usually operates with condenser design inlet and outlet cooling water temperatures of 23 °C and 30 °C respectively, these design values can be altered with a relatively small reduction in the net efficiency of the electricity generating cogeneration system (about a 2%) in order to obtain a condenser cooling water outlet at approximately 90 °C. Therefore it is assumed that the cooling water exiting the condenser of the power generation system, hot water at approximately 92 °C, will be an acceptable fuel source for the absorption refrigeration system. The utilization of this low-grade thermal energy results in a net efficiency decrease of 2% for the electrical power-generating unit.

A lithium bromide-water solution absorption refrigeration cycle

[2] Turboden s.r.l.: http://www.turboden.eu/en/home.index.php.

Table 2
Turboden s.r.l. system and calculated performance characteristics.

Data description	Source	Data value	Unit
Heat source calculations			
Output temperature from exchanger	Turboden	200	°C
Exhaust gas average specific heat capacity	Turboden	1.1	kJ/kg.K
Heat losses from heat exchanger	Turboden	2	%
Net available thermal power	Calculated	17,060	kW
ORC power generation unit			
ORC unit type	Turboden	TD40	–
Heat exchange configuration	Turboden	Indirect exchange	–
ORC gross power output at generator terminals	Calculated	4130	kW
ORC captive power consumption	Calculated	195	kW
ORC net output power	Calculated	3935	kW
Thermal power to cooling source	Calculated	12,700	kW
Electrical generator			
Generator type	Turboden	Asynchronous	–
Generator frequency	Turboden	50	Hz
Generator voltage	Turboden	Medium voltage	–
Cooling subsystem (if required)			
Cooling type (ORC condenser)	Turboden	Dry WCC	–
Cooling system internal consumption	Calculated	180	kW

recommended by Voltas Technologies is chosen for the cooling power generation. Fig. 2 shows the diagram of this cooling power generating unit. The coefficient of performance (COP) of this cooling system is 0.7. Therefore Eq. (10) is used to calculate the cooling power output.

$$Q_{cool,cold} = COP \times Q_{cool,low} = 0.7 Q_{cool,low}. \tag{10}$$

The equations from (1)–(10) will be used to determine the available heat from the smelting process, the generated cooling power, and the generated electrical power, which is then used by the EPD algorithm to develop the optimal power dispatch schedule in the following section.

3. EPD model development

The cogeneration system and the energy flow diagram is shown in Fig. 3. As mentioned earlier, the EPD model determines the optimal dispatching of the generated electricity, E, between the on-site loads and the utility grid so that the maximum possible overall savings can be obtained.

The decision variable of the EPD problem is thus the amount of generated electricity, denoted by $C_{i,j}$ in MW, that is dispatched back to the furnace loads. In $C_{i,j}, i = 1,2,...,m$ is the index of days in a month and $j = 1,2,...,48$ is the index of hours in a day. To account for the maximum demand charge cost of the plant, which is determined by the recorded maximum power drawn by the plant in a month, the EPD problem is formulated over the period of a month. The sampling period of the EPD problem is taken as half an hour, which is the integrating period of the utility for energy and demand charges, to ease calculations of the overall cost.

The system's overall cost to be minimized is the sum of the system-related energy costs/incentives, use-of-service (UoS) charges and costs, and costs associated with the generation of energy via the cogeneration system. The overall energy cost represents all costs associated with the consumption of energy from the utility grid less the financial incentives obtained through the sell-back of cogeneration generated power to the utility grid. The UoS charges and costs account for all costs and rebates associated with the power supply from the utility grid, including administration and network reliability costs. Lastly, the power generation costs include all costs that are incurred through the process of generating the additional useful electricity via the cogeneration system. To be specific, the cost of power generation includes two parts. The first part is the captive power consumed by the cogeneration system, which is accounted for by (8). The second part is the energy cost of the operation of the ID fans, which is part of the on-site loads.

The overall objective function of the EPD model is therefore summarised in the following equation:

$$\begin{aligned}\text{Overall Cost} &= \text{cost of on-site energy consumption}\\ &\quad -\text{total incentive from sold energy} + \text{UoS charges}.\end{aligned} \tag{11}$$

According to the consumption and generation tariff structures set by the local utility, Eskom, overall energy consumption and generation costs are to be determined according to the time-of-use (TOU) tariff [52,53]:

$$p_j = \begin{cases} p_p, & \text{if } j \in \{15,16,...,20\} \cup \{37,38,...,40\};\\ p_s, & \text{if } j \in \{13,14\} \cup \{21,22,...,36\} \cup \{41,42,43,44\};\\ p_o, & \text{if } j \in \{1,2,...,12\} \cup \{45,46,47,48\}. \end{cases} \tag{12}$$

Therefore the cost function must account for these TOU periods using peak, standard and off-peak TOU period flag variables. The peak, standard and off-peak flag variables take values of either one or zero and are defined by $P_{i,j} = 1$ if $D(i) = 1, j \in \{15,...,20\} \cup \{37,...,40\}; S_{i,j} = 1$ if $D(i) = 1, j \in \{13,14\} \cup \{21,...,36\} \cup \{41,...,44\}$ or $D(i) = 2, j \in \cup \{15...,24\} \cup \{37...,40\}$; and $O_{i,j} = 1$ if $D(i) = 1, j \in \{1,12\} \cup \{45,...,48\}$ or $D(i) = 2, j \in \cup \{1...,14\} \cup \{25...,36\} \cup \{41,...,48\}$ or $D(i) = 3, j \in \cup \{1...,48\}$ with $D(i)$ defined by

$$D(i) = \begin{cases} 1, & \text{if day } i \text{ is a Weekday;}\\ 2, & \text{if day } i \text{ is a Saturday;}\\ 3, & \text{if day } i \text{ is a Sunday.} \end{cases} \tag{13}$$

With the optimization variables and TOU periods defined, the overall system costs for the plant can be determined. The relevant costs under the MEGAFLEX tariff structure [52,53] are discussed in the following subsections.

3.1. Energy consumption-related costs

3.1.1. Network demand charge

A consumption network demand charge (NDC) for the plant's maximum demand is shown in (14).

$$\begin{aligned} NDC = \max\left[|S_{i,j}^{load}| \times \left(1 - \frac{C_{i,j}}{P_{i,j}^{load}}\right) \times (P_{i,j} + S_{i,j}) \right] \\ \times NDC_r, \text{ for all } i \text{ and } j \end{aligned} \tag{14}$$

3.1.2. Active energy consumption charges

The peak, standard and off-peak active energy consumption charges, PEC, SEC and OEC, respectively, for the total amount of energy consumed, are shown in (15)–(17).

$$PEC = 0.5 p_p \sum_{i=1}^{m} \sum_{j=1}^{48} [(P_{i,j}^{load} - C_{i,j}) \times P_{i,j}] \tag{15}$$

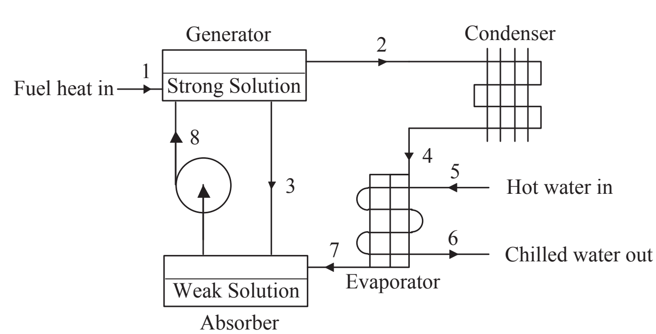


Fig. 2. System incorporated lithium bromide-water solution absorption refrigeration cycle.

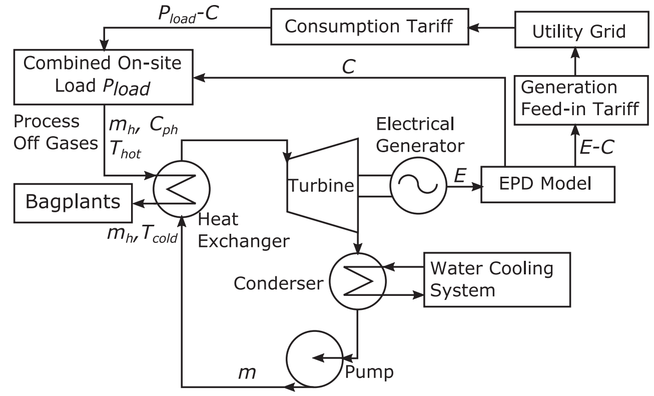


Fig. 3. Waste heat recovery cogeneration system and EPD model energy flow diagram.

$$SEC = 0.5p_s \sum_{i=1}^{m} \sum_{j=1}^{48} [(P_{i,j}^{load} - C_{i,j}) \times S_{i,j}] \tag{16}$$

$$OEC = 0.5p_o \sum_{i=1}^{m} \sum_{j=1}^{48} [(P_{i,j}^{load} - C_{i,j}) \times O_{i,j}] \tag{17}$$

3.1.3. Consumption network related charges

A consumption network access charge (NAC) based on the voltage of power supply and the annual utilised capacity is shown in (18).

$$NAC = AD \times NAC_r \tag{18}$$

A transmission network charge (TNC) is shown in (19).

$$TNC = AD \times TNC_r \tag{19}$$

An urban low voltage subsidy charge (ULVSC) is determined by (20).

$$ULVSC = AD \times ULVSC_r \tag{20}$$

3.1.4. Electrification and rural subsidy charge

An electrification and rural subsidy charge (ERC), applied to the total amount of active energy consumed, is shown in (21).

$$ERC = 0.5ERC_r \sum_{i=1}^{m} \sum_{j=1}^{48} [(P_{i,j}^{load} - C_{i,j}) \times (P_{i,j} + S_{i,j} + O_{i,j})] \tag{21}$$

3.1.5. Reactive energy charge

A reactive energy charge (REC) based on the total amount of excess reactive energy required by the plant is shown in (22).

$$REC = \begin{cases} 0.5REC_r \sum_{i=1}^{n} \sum_{j=1}^{48} [Q_{i,j}^{exc} \times (P_{i,j} + S_{i,j})], & \text{if } Q_{i,j}^{exc} > 0; \\ 0, & \text{otherwise;} \end{cases} \tag{22}$$

where $Q_{i,j}^{exc} = (P_{i,j}^{load} - C_{i,j}) \times \left(\sqrt{\left(\frac{|S_{i,j}^{load}|}{P_{i,j}^{load}} \right)^2 - 1} - 0.3 \right)$.

3.1.6. Consumption service and administration charges

A consumption administration charge (CAC) and a consumption service charge (CSC) for the utilisation of the utility grid are shown in (23) and (24).

$$CAC = m \times CAC_r \tag{23}$$

$$CSC = m \times CSC_r \tag{24}$$

3.2. Energy generation related costs

3.2.1. Generation network access charge

A generation network access charge (gNAC) for the cogeneration system to sell electricity back to the grid is shown in (25).

$$gNAC = gAD \times gNAC_r \tag{25}$$

3.2.2. Active energy generation charges and total rebate

Peak, standard and off-peak active energy generation incentives, PEI, SEI and OEI respectively, for the total amount of active energy sold to or wheeled through the utility grid to third party customers, and a rebate to be subtracted from the *gNAC*, shown in (26)–(29).

$$PEI = 0.5p_p^g \sum_{i=1}^{n} \sum_{j=1}^{48} [(E_{i,j} - C_{i,j}) \times P_{i,j}] \tag{26}$$

$$SEI = 0.5p_s^g \sum_{i=1}^{n} \sum_{j=1}^{48} [(E_{i,j} - C_{i,j}) \times S_{i,j}] \tag{27}$$

$$OEI = 0.5p_o^g \sum_{i=1}^{n} \sum_{j=1}^{48} [(E_{i,j} - C_{i,j}) \times O_{i,j}] \tag{28}$$

$$\text{Rebate} = (PEI + SEI + OEI) \times (DLF \times TLF - 1) \tag{29}$$

The distribution and transmission loss factors, DLF and TLF, in Eq. (29) are given in [52,53]. The rebate is to be subtracted from *gNAC* only if this charge is applicable. If the gNAC is not applicable, the rebate will be 0.

3.2.3. Generation service and administration charges

A generation administration charge (GAC) and a generation service charge (GSC) for the utilisation of the utility grid are shown in (30) and (31).

$$GAC = m \times GAC_r \tag{30}$$

$$GSC = m \times GSC_r \tag{31}$$

3.3. System network reliability service charge

A combined reliability service charge (RSC) based on the supply voltage of the utility grid for both energy consumption and generation is shown in (32)–(34).

$$CRC = 0.5CRC_r \sum_{i=1}^{n} \sum_{j=1}^{48} [(P_{i,j}^{load} - C_{i,j}) \times (P_{i,j} + S_{i,j} + O_{i,j})] \tag{32}$$

$$GRC = 0.5GRC_r \sum_{i=1}^{n} \sum_{j=1}^{48} [E_{i,j} - C_{i,j}) \times (P_{i,j} + S_{i,j} + O_{i,j})] \tag{33}$$

$$RSC = \max(CRC, GRC). \tag{34}$$

3.4. The final cost function

Considering all costs discussed in Sections 3.1–3.3, the final cost function of the EPD problem is re-written as:

$$\begin{aligned} Cost &= NDC + (PEC + SEC + OEC) + ERC + REC \\ &+ CAC + CSC + TNC + \max(NAC: (gNAC \\ &+ \text{Rebate})) + ULVSC - (PEI + SEI + OEI) \\ &+ GAC + GSC + RSC. \end{aligned} \tag{35}$$

The EPD optimization problem is eventually formulated as minimizing system associated cost (35) subject to the available heat for cogeneration (7), the efficiencies of the electrical and cooling power generating units detailed in Sections 2.2 and 2.3, and the energy consumption of the plant.

4. Case study and result analysis

4.1. EPD model data requirements

A case study of a chrome smelting plant in South Africa that utilizes four AC submerged electrode arc furnaces is presented. The EPD algorithm utilizes the raw facility and process-related data in order to calculate the energy generation capacity of the proposed cogeneration system. In particular, the raw data required include the real and apparent load powers of all four furnaces and the associated induced draft fans, in MW and MVA respectively, and the extracted off-gas temperatures before and after the cogeneration system. All data were obtained with corresponding date-time stamps (yyyy/mm/dd hh:mm) with a sampling period of half an hour.

The EPD algorithm utilizes a monthly-based cost function and therefore requires an entire month's data for the optimal power dispatch schedule development. However, the EPD model is expected to operate in real-time, generating a current optimal power dispatch schedule for any time interval in a given month. A forecast model based on historical data is developed to facilitate the real-time operation of the EPD model. The forecast model developed predicts the average daily load profile data array using historical data acquired. This forecast data array is then used by the EPD model and updated every 30 min by the most recent measurement.

The main objective of the EPD model is to provide the facility with energy and cost savings. Therefore, the primary and most important result is the overall system associated energy and cost savings from the heat recovery cogeneration system and the associated EPD model. In addition, this study also serves to investigate the overall capacity of the previously wasted thermal energy from the off-gasses and the cooling power capacity that can be additionally generated using the byproduct, low-grade thermal energy, of the waste heat recovery system.

Process and facility-related raw data were obtained for the time period from 2014/08/01 00:00 to 2014/11/12 12:30. The EPD model was tested using a winter and a summer month, August 2014 and October 2014 respectively, to investigate the effect of seasonal variations.

4.2. Potential of the cogeneration system

Fig. 4 shows the combined on-site loads of the plant studied and the corresponding available thermal powers that can be used by the cogeneration system for a winter month (August) and a summer month (October).

Making use of the raw data and the characteristics of the combined cogeneration system, Fig. 5 depicts the average daily profiles of generated electricity and cooling power from the combined cogeneration system for the studied winter and summer months. In addition, Table 3 shows statistics on the profiles in Figs. 4 and 5 and the overall waste thermal energy recovery efficiency of the cogeneration system.

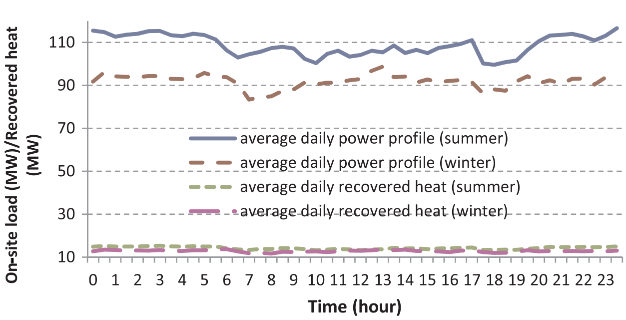


Fig. 4. Average on-site load and recovered heat.

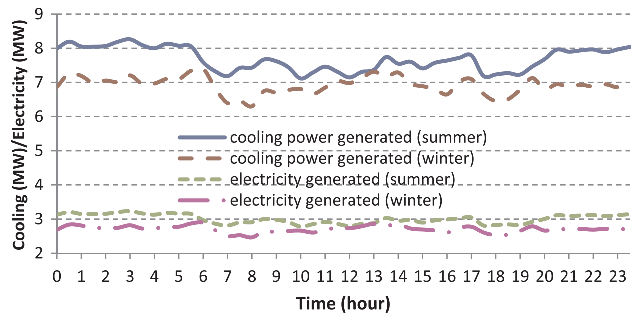


Fig. 5. Average daily generated electricity and cooling power.

4.3. EPD developed optimal power flow schedule

The EPD optimization problem is solved by the *sqp* algorithm built into Matlab. The optimal power dispatch schedule is generated for each half hour interval in the given month. In real-time operation, as time goes by and the following time interval is reached, a new optimal power dispatch schedule is calculated according to the most recent measurement. This process continues until the end of the given month and starts again at the beginning of the following month.

Because of the relatively recent nature of feeding electrical power back to the utility grid in South Africa, no specific feed-in energy tariff structure has been implemented. Currently, a facility feeding electrical power back into the grid obtains a financial incentive from a third-party customer, or Eskom itself, which buys this electrical power. The rates that are implemented for this transaction are the base or wholesale electricity pricing system (WEPS) energy rates. It is noticed that the WEPS energy rates have a significant impact on the EPD model schedule and the overall system associated savings. The WEPS energy rates set by the local utility [52] were found to be too low to allow for or encourage the feed of electrical power back into the utility grid. Consequently, the EPD schedule that was developed dispatched all generated electricity to the on-site loads. For the purpose of this research, the WEPS energy rates are adjusted in order to fully investigate the benefits of the cogeneration system under competitive feed-in tariffs. Table 4 shows the adjusted WEPS energy rates.

The schedules generated by the EPD model for a winter day, 05 August 2014, and a summer day, 10 October 2014, under the adjusted WEPS rates are presented in Figs. 6 and 7, in which different color backgrounds represent the peak (red[3]), standard (yellow) and offpeak (green) periods of the TOU tariff. From these figures, it can be seen that if the WEPS rates encourage this, the EPD would feed the cogeneration generated power back to the grid to support the utility during critical periods.

4.4. System and facility associated savings

To investigate the financial benefits and viability of the proposed cogeneration system, the final system associated cost savings are calculated by subtracting the energy costs of the plant after implementation of the cogeneration system and the associated EPD algorithm from that of the existing system. This is in line with standard protocols on measurement and verification of energy savings (theory and case studies of measurement and verification can be found in [54,55]).

For the chosen winter and summer months the final calculated cost savings obtained are shown in Table 5. Assuming these cost savings are similarly achieved for each winter and summer month respectively, an approximate annual energy cost savings figure of $1351282.80 can be obtained for the plant.

[3] For interpretation of color in Figs. 6 and 7, the reader is referred to the web version of this article.

Table 3
Statistics on the combined averaged daily profiles.

Results	Winter: August 2014			Summer: October 2014		
	Minimum	Average	Maximum	Minimum	Average	Maximum
Average daily load: P_{load} (MW)	84	92.81	100	100	109.5	118
Average daily recovered heat: $Q_{h,total}$ (MW)	11	12.85	14	13	14.26	16
Average daily generated electricity: E (MW)	2.4	2.71	2.9	2.7	3	3.3
Average daily generated cooling power: $Q_{cool,cold}$ (MW)	6.2	6.92	7.4	7.1	7.68	8.3
Average overall waste energy utilization (cooling + electrical) efficiency (%)		74.94			74.89	

4.5. Additional savings from generated cooling power

There are many requirements for cooling throughout the plant, the most prominent being the cooling plant that provides cooling water for the furnace shells. Additional cooling requirements include cooling fans throughout the plant in numerous plant rooms such as the transformer rooms and the hydraulic rooms, office buildings, and so on. The studied plant requires a total facility-wide cooling power of 1.81 MW. The total combined average available cooling power generated via the combined cogeneration system is significantly more than the required cooling capacity of the facility as shown in Section 4.2.

Additional cost savings can be calculated by applying the MEGAFLEX active energy charges to the cooling demand of the plant that is now supplied by the combined cogeneration system. Also, the furnace load can be further reduced because of the reduction in the cooling power required by the furnaces. Utilizing the chosen winter and summer months to calculate additional savings, the overall additional savings are obtained:

- $46742.37 and $30007.80 savings from reduction of the furnace load for August and October 2014, respectively.
- $43658.72 and $45502.95 savings from substituted cooling power for August and October 2014, respectively.

Again assuming these savings are similarly achieved for each summer and winter month, an approximate annual cost savings of $2302082.93 can be obtained for the plant, an increase of $950800.13 due to the utilization of the available cooling power from the proposed absorption refrigeration system.

5. Further discussion of the EPD model results

5.1. Projected payback period

The cogeneration and power generation technologies are still relatively expensive and the success of such a system or project is often determined by its payback period, which is the time it takes for the savings to pay back the capital or project start-up costs. A summary of the potential costs required for the implementation of the proposed combined system is shown in Table 6. Taking into account additional costs for installation and labor, the payback period for the entire system is found to be from 4 to 5 years, which is acceptable for industrial projects.

Table 4
Initial and adjusted WEPS energy rates.

	Peak (c/kWh)	Standard (c/kWh)	Off-Peak (c/kWh)
Adjusted rates for high demand season			
Initial WEPS rates	17.16	5.20	2.82
Adjusted WEPS rates	19.45	6.27	3.61
% Increase	13.39	20.61	28.05
Adjusted rates for low demand season			
Initial WEPS rates	5.60	3.85	2.44
Adjusted WEPS rates	6.69	4.75	3.19
% Increase	19.50	23.43	30.69

5.2. Seasonal variations

The results obtained for winter and summer months are compared with each other to identify the seasonal performance variations. It is found that, although the results do not differ significantly, the comfort cooling required throughout the facility is considerably less in winter. In fact, heating is required rather than cooling in winter. Therefore, an improvement in the system could be made to provide the facility with comfort heating during winter months. In general, it is noticed that slightly more, about 0.18%, cost savings are obtained during winter months. This is because of the much higher active energy consumption rates during winter months, especially during peak and standard TOU periods. Being able to significantly reduce the amount of active energy consumed from the utility grid by the cogeneration system allowed for the slight increase in the overall system associated cost savings during winter months.

5.3. Potential improvements

A number of improvements are identified with regards to this research. The most significant improvements identified are:

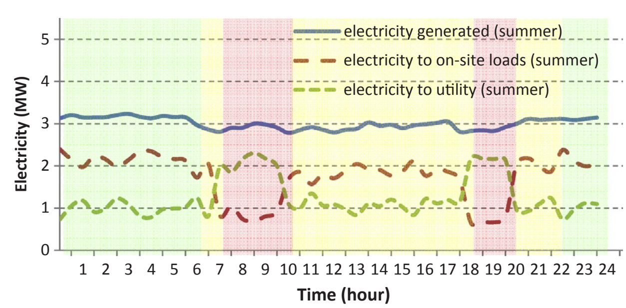


Fig. 6. EPD result of a summer month.

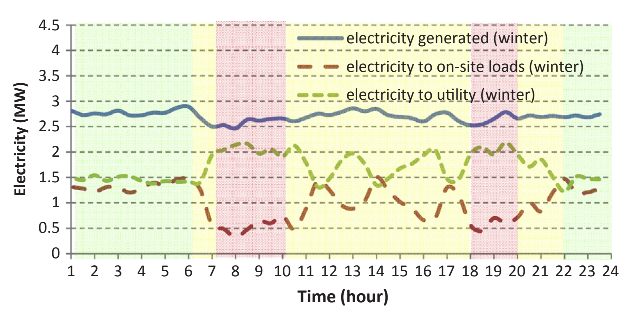


Fig. 7. EPD result of a winter month.

Table 5
Total energy cost and savings.

Parameter	Winter: August 2014	Summer: October 2014
Total initial cost	$5356482.74	$4209939.62
Total final cost	$5215508.85	$4106788.38
Total savings	$140973.88	$103151.24
% Savings	2.63%	2.45%

Table 6
Capital costs associated with the combined cogeneration system.

System component	Cost ($)
Turboden TD40 ORC Unit	4935785.83
Intermediate gas/oil thermal exchanger	3286310.67
2 MW single stage hot water chiller	276578.92
Total cost	8498675.41

- The thermal energy recovery efficiency can be improved by allowing for a lower bagplant inlet temperature.
- The thermal energy recovery efficiency can be improved by convert the open furnaces to closed ones. This will allow for the direct burning of the CO gas in a gas combustion engine, which is much more efficient than the proposed power generation system.
- The excess cooling power, more than 5 MW, can be sold and dispatched to outlying or adjacent facilities to meet additional cooling requirements and to allow for additional financial incentives.

6. Conclusion

Obtaining an optimal operation strategy for cogeneration systems in a grid-connected environment is a challenging task, which has not been well studied in the literature. In this study, the optimal operation of a grid-tied cogeneration system aiming at maximizing the benefits of a ferrochrome smelting plant and aiding the utility grid during peak demand periods is formulated into an optimization problem. The raw material smelting processes and the characteristics of the cogeneration system are modeled first to determine the available process waste heat and the corresponding electrical and cooling power generation capacities of the cogeneration system. The optimization model is designed to make use of the process models and the consumption and feed-in tariffs determined by the utility to dispatch the generated electricity between the on-site loads and the utility grid optimally. The effectiveness of the model is demonstrated by a case study. Further, the optimization model developed can be adapted for similar grid-connected cogeneration systems to improve their efficiency. It can also be used to evaluate the financial viability of new cogeneration projects.

Acknowledgement

The authors would like to thank the reviewers for their insightful comments and suggestions. The quality of this work has been improved significantly because of their inputs.

References

[1] Oluleye G, Jobson M, Smith R, Perry SJ. Evaluating the potential of process sites for waste heat recovery. Appl Energy 2016;161:627–46.
[2] Zhang C, Zhou L, Chhabra P, Garud SS, Aditya K, Romagnoli A, et al. A novel methodology for the design of waste heat recovery network in eco-industrial park using techno-economic analysis and multi-objective optimization. Appl Energy 2016;184:88–102.
[3] Cameron EN. Chromite in the central sector of the eastern Bushveld Complex, South Africa. Am Miner 1977;62:1082–96.
[4] Cawthorn RG. The platinum group element deposits of the Bushveld Complex in South Africa. Platin Met Rev 2010;54(4):205–15.
[5] Pan X. Effect of South African chrome ores on ferrochrome production. In: International conference on mining, mineral processing and metallurgical engineering, Johannesburg, South Africa; 2013.
[6] Roos J, Hearn A. Optimising the effective use of energy in the ferroalloy industry through innovative technology. In: Proceedings of the tenth international ferro alloys congress, Cape Town, South Africa; 2004. p. 712–25.
[7] Ishitobi T, Ichihara K, Homma T. Operational improvements of a submerged arc furnace in kashima works (KF-1) relined in 2006. In: The 12th international ferroalloys congress: sustainable future, Helsinki, Finland; 2010. p. 509–15.
[8] Daavittila J, Honkaniemi M, Jokinen P. The transformation of ferrochromium smelting technologies during the last decades. J South Afr Inst Min Metall 2004;104(9):541–9.
[9] Niemelä P, Kauppi M. Production, characteristics and use of ferrochromium slags. In: 11th International congress on ferroalloys, New Delhi, India; 2007.
[10] Hayes P. Aspects of SAF smelting of ferrochrome. In: 10th International ferroalloys congress, Cape Town, South Africa; 2004.
[11] Sala A, Flores I, Sala J, Millán J, Gómez I, López L. Cogeneration technology for the metal-processing sector. Appl Energy 2008;85(6):516–27.
[12] Yu H, Feng X, Wang Y, Biegler LT, Eason J. A systematic method to customize an efficient organic Rankine cycle (ORC) to recover waste heat in refineries. Appl Energy 2016;179:302–15.
[13] Murray R, de Kock JA. Potential in utilising furnace off-gas at a ferrochrome smelter with gas engines. In: 2015 International conference on the industrial and commercial use of energy, Cape Town, South Africa; 2015. p. 171–7.
[14] Beaty H, Fink D. Standard handbook for electrical engineers. 16th ed. McGraw-Hill Education; 2012.
[15] Dong H, Li J, Li Z, Zhao Y, Cai J. Cogeneration system utilizing waste heat from sinter-cooling process. In: The 2nd international symposium on power electronics for distributed generation systems, Hefei, China; 2010. p. 674–7.
[16] Hung T, Shai T, Wang S. A review of organic Rankine cycles (ORCs) for the recovery of low-grade waste heat. Energy 1997;22(7):661–7.
[17] Sun L, Han W, Jing X, Zheng D, Jin H. A power and cooling cogeneration system using mid/low-temperature heat source. Appl Energy 2013;112:886–97.
[18] Wei D, Lu X, Lu Z, Gu J. Performance analysis and optimization of organic Rankine cycle (ORC) for waste heat recovery. Energy Convers Manage 2007;48(4):1113–9.
[19] Ziviani D, Beyene A, Venturini M. Advances and challenges in ORC systems modeling for low grade thermal energy recovery. Appl Energy 2014;121:79–95.
[20] Liew PY, Lim JS, Alwi SRW, Manan ZA, Varbanov PS, Klemeš JJ. A retrofit framework for total site heat recovery systems. Appl Energy 2014;135:778–90.
[21] Moussawi HA, Fardoun F, Louahlia H. Selection based on differences between cogeneration and trigeneration in various prime mover technologies. Renew Sustain Energy Rev 2017;74:491–511.
[22] Saleh B, Koglbauer G, Wendland M, Fischer J. Working fluids for low-temperature organic Rankine cycles. Energy 2007;32(7):1210–21.
[23] Chen H, Goswami DY, Stefanakos EK. A review of thermodynamic cycles and working fluids for the conversion of low-grade heat. Renew Sustain Energy Rev 2010;14(9):3059–67.
[24] Lecompte S, Huisseune H, van den Broek M, Vanslambrouck B, Paepe MD. Review of organic rankine cycle (ORC) architectures for waste heat recovery. Renew Sustain Energy Rev 2015;47:448–61.
[25] Chen H, Goswami DY, Rahman MM, Stefanakos EK. A supercritical rankine cycle using zeotropic mixture working fluids for the conversion of low-grade heat into power. Energy 2011;36(1):549–55.
[26] Xia X, Zhang J. Energy efficiency and control systems – from a POET perspective. IFAC Proc Vol 2010;43(1):255–60.
[27] Zhuan X, Xia X. Optimal operation scheduling of a pumping station with multiple pumps. Appl Energy 2013;104:250–7.
[28] Wang B, Xia X, Zhang J. A multi-objective optimization model for the life-cycle cost analysis and retrofitting planning of buildings. Energy Build 2014;77:227–35.
[29] Setlhaolo D, Xia X. Optimal scheduling of household appliances with a battery storage system and coordination. Energy Build 2015;94:61–70.
[30] Wu Z, Xia X, Wang B. Improving building energy efficiency by multiobjective neighborhood field optimization. Energy Build 2015;87:45–56.
[31] Ekpenyong UE, Zhang J, Xia X. An improved robust model for generator maintenance scheduling. Electr Power Syst Res 2012;92:29–36.
[32] Zhang S, Xia X. A new energy calculation model of belt conveyor. In: AFRICON, Nairobi, Kenya; 2009.
[33] Xia X, Zhang L. Industrial energy systems in view of energy efficiency and operation control. Annu Rev Control 2016;42:299–308.
[34] Zhang L, Xia X. Control of industrial systems: mining industry as a case study. Control Eng China 2016;23(12).
[35] Xia X, Zhang J. Operation efficiency optimisation modelling and application of model predictive control. IEEE/CAA J Autom Sinica 2015;2(2):166–72.
[36] Elaiw AM, Xia X, Shehata AM. Hybrid DE-SQP and hybrid PSO-SQP methods for solving dynamic economic emission dispatch problem with valve-point effects. Electr Power Syst Res 2013;103:192–200.
[37] Zhang S, Xia X. Optimal control of operation efficiency of belt conveyor systems. Appl Energy 2010;87(6):1929–37.
[38] Zhang L, Xia X, Zhang J. Improving energy efficiency of cyclone circuits in coal beneficiation plants by pump-storage systems. Appl Energy 2014;119(0):306–13.
[39] Badenhorst W, Zhang J, Xia X. Optimal hoist scheduling of a deep level mine twin rock winder system for demand side management. Electr Power Syst Res 2011;81(5):1088–95.
[40] Mathaba T, Xia X. A parametric energy model for energy management of long belt conveyors. Energies 2015;8(12):12375.
[41] Maruoka N, Akiyama T. Exergy recovery from steelmaking off-gas by latent heat storage for methanol production. Energy 2006;31(10-11):1632–42.

[42] Nwulu N, Xia X. A combined dynamic economic emission dispatch and time of use demand response mathematical modelling framework. J Renew Sustain Energy 2015;7(4).
[43] Xia X, Zhang J, Elaiw A. A model predictive control approach to dynamic economic dispatch problem. In: 2009 IEEE Bucharest PowerTech, Bucharest, Romania; 2009. p. 1–7.
[44] Talaq J, El-Hawary F, El-Hawary M. A summary of environmental/economic dispatch algorithms. IEEE Trans Power Syst 1994;9(3):1508–16.
[45] Xia X, Zhang J, Elaiw A. An application of model predictive control to the dynamic economic dispatch of power generation. Control Eng Pract 2011;19(6):638–48.
[46] Xia X, Elaiw A. Optimal dynamic economic dispatch of generation: a review. Electr Power Syst Res 2010;80(8):975–86.
[47] Wu Z, Tazvinga H, Xia X. Demand side management of photovoltaic-battery hybrid system. Appl Energy 2015:294–304.
[48] Tazvinga H, Zhu B, Xia X. Optimal power flow management for distributed energy resources with batteries. Energy Convers Manage 2015;102:104–10.
[49] Sichilalu S, Tazvinga H, Xia X. Optimal control of a fuel cell/wind/pv/grid hybrid system with thermal heat pump load. Sol Energy 2016;135:59–69.
[50] Zhu B, Tazvinga H, Xia X. Switched model predictive control for energy dispatching of a photovoltaic-diesel-battery hybrid power system. IEEE Trans Control Syst Technol 2015;23(3):1229–36.
[51] Tazvinga H, Zhu B, Xia X. Energy dispatch strategy for a photovoltaic-wind-diesel-battery hybrid power system. Sol Energy 2014;108:412–20.
[52] Eskom. Tariff & charges booklet 2014/15; 2014.
[53] Eskom. generator tariff charges 2014/15; 2014.
[54] Xia X, Zhang J. Mathematical description for the measurement and verification of energy efficiency improvement. Appl Energy 2013;111:247–56.
[55] Xia X, Zhang J, editors. Energy efficiency measurement & verification practices: demystifying M & V through south african case studies. Media in Africa, Pretoria, South Africa; 2012.

Applied Energy 228 (2018) 2140–2152

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Building retrofit optimization models using notch test data considering energy performance certificate compliance

Yuling Fan[a,b,*], Xiaohua Xia[b]

[a] College of Informatics, Huazhong Agricultural University, Wuhan, Hubei 430072, China
[b] Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa

HIGHLIGHTS

- Two simplified optimization models for whole-building retrofit planning are proposed.
- The models reduce the complexity of systematic building retrofit planning problems.
- The models eliminate a costly detailed bottom-up energy audit process.
- The models consider green building policy based on EPC and the tax incentive.
- The models can be of great help for retrofit plans for a building portfolio.

ARTICLE INFO

Keywords:
Whole-building retrofit
Energy performance certificate
Grouping
Notch test data
Green building

ABSTRACT

Determining a systematic whole-building retrofit plan for envelope components and indoor appliances to achieve targets such as cost savings and policy compliance is a challenging task. To be specific, the systematic whole-building retrofit problem, when solved by an optimization approach, is highly complicated. It is sometimes even impossible to find a solution with given computation resources and algorithms. In addition, a costly comprehensive bottom-up audit is required to establish the parameters of the problem. This study presents two models to reduce the complexity of the systematic whole-building retrofit optimization problem. Firstly, the proposed models use the grouping concept to optimize the retrofit of subsystems in a building instead of individual components/appliances, which reduces the dimension of the problem effectively. Secondly, the models use so-called 'notch test' data, which are sampled and verified savings of an intervention, to eliminate the need for bottom-up energy audits. This further simplifies the retrofit optimization problem and reduces the retrofit cost. The models are based on our previous work and aim at energy savings maximization and payback period minimization, considering the green building policy and tax incentive initiatives. A case study shows that about 2530 MWh energy savings and an A rating from the energy performance certificate standard can be obtained with a payback period of 59 months, which verifies the feasibility and effectiveness of the models proposed.

1. Introduction

Globally, the building sector accounts for around 30–40% of total energy consumption [1]. Statistics show that this number is even higher in the European Union [2]. This high energy usage by the building sector is mainly attributed to existing buildings [3] and still keep increasing, because of the low construction rate of new buildings and the fact that new buildings are more energy-efficient owing to tighter energy regulations introduced [4]. In view of this, retrofitting existing buildings with energy-efficient technologies to bring down their energy intensities is an effective and common approach to facilitate the transition to a green building sector, which is proved by the investigation on building retrofit potential in [5], the studies on office building retrofit in [6] and sustainable building retrofit decisions in [7]. For instance, energy-efficient lightings are useful to reduce the energy usage of artificial lighting [8], good window technologies promote better energy-saving ventilation [9], and heating, ventilation and air conditioning (HVAC) help to reduce the energy consumed by cooling [10], heating and ventilation [11] and to promote a healthy indoor environment [12].

Many policies around the world are implemented to promote a green building sector that utilizes energy efficiently, such as the

* Corresponding author at: College of Informatics, Huazhong Agricultural University, Wuhan, Hubei 430072, China.
E-mail address: ylfan.up@gmail.com (Y. Fan).

https://doi.org/10.1016/j.apenergy.2018.07.043
Received 23 April 2018; Received in revised form 6 July 2018; Accepted 12 July 2018

Leadership in Energy and Environmental Design (LEED) certification program [13], Building for Environment and Economic Sustainability [14], the Green Star rating system [15], the Canadian green building tool [16], the Italian regulation [17], etc. The South African government has also released a green building rating policy based on the energy intensity of buildings, namely the energy performance certificate (EPC) for buildings [18]. The purpose is to compel building owners or managers to reduce the energy demand of their buildings by implementing energy-efficient interventions. The green building policy is proposed to be applied to public buildings first and to all kinds of buildings at a later stage. Seven energy intensity ratings, ranging from grade A (most energy-efficient) to grade G (most energy-inefficient), are available from the EPC system. The effectiveness of the EPC standard depends on two aspects. One is the effective implementation and monitoring of the EPC. South African government has published a policy that prohibits the use of buildings not complying with the required EPC level to promote implementation and monitoring of buildings' energy usage. The other is the accuracy of energy performance evaluation [19]. The evaluation is addressed by a scientific measurement and verification (M&V) approach [20]. The uncertainty of the energy performance evaluation depends on many factors, such as measurement uncertainty, modeling uncertainty etc. [21]. In this study, real-world 'notch test' data are used to improve accuracy and hence uncertainty of the EPC certification process. Given the aforementioned background, it is essential to develop methods to retrofit buildings in a cost-effective manner, not only to achieve energy and cost savings, but also to adhere to the green building policy introduced.

In the literature, studies on the economic perspective of green building rating and building energy performance contracting projects were reported recently. For instance, Qian and Guo [22] built a revenue-sharing bargaining model for energy performance contracting projects. Castro-Lacouture et al. [23] developed an optimal design model for buildings for the purpose of maximizing the credits under the LEED rating system and [24] proposed an optimization model aiming at maximizing the economic benefits, energy and water savings as well as LEED points. However, no study that can be used to support decision makers technically, considering building retrofit investment for the purpose of EPC compliance, can be found in the literature. In particular, the EPC standard requires the energy intensity of the whole building to be reduced, which calls for a whole-building retrofit approach, considering both indoor appliances and the envelope components and interactions between them.

Existing studies relevant to building retrofit from the literature can be categorized into two general types, namely studies on the building envelope system and indoor appliances. With respect to the envelope system, Asadi et al. [25] proposed a multi-objective optimization method to help decision makers to determine intervention measures for the purpose of minimizing building energy consumption in a cost-effective manner. Güçyeter and Günaydın [26] evaluated and optimally determined retrofit strategies for a building envelope system by a calibrated simulation method based on energy audit and monitoring. Edeisy and Cecere [27] investigated envelope retrofit as a tool to increase comfort levels and decrease cooling loads in hot, arid climates. Fan and Xia [28] developed an optimization method for building envelope retrofit planning, considering a roof-top photovoltaic (PV) system, for energy efficiency improvement. Regarding indoor appliances, Kang and Liu [29] proposed a multi-objective optimization model on a heat exchanger network retrofit with a heat pump for simultaneously minimizing the retrofit cost and maximizing the CO_2 emission reduction. Cartens et al. [30] developed a model for reducing the cost and energy consumption in clean development mechanism lighting retrofit projects. Wang and Xia [31] introduced a control system framework to tackle the retrofit planning problems for indoor appliances to reduce energy intensity.

Very few studies on determining systematic retrofit plans for a whole building have been reported. To this end, our previous work [32] presented an approach to identifying systematic whole-building retrofit plans for existing buildings considering both the envelope components and the indoor appliances with the purpose of maximizing energy savings and green building policy compliance. However, determining such a systematic whole-building retrofit plan with the approach proposed in [32] is quite complex. Firstly, the large numbers of items to be retrofitted and those of the available alternatives for retrofit result in a high-dimensional optimization problem. This makes the problem difficult to solve when coupled with the mixed integer decision variables involved. Secondly, the conflicting objectives, such as maximizing energy savings and minimizing the payback period, and the nonlinear characteristics of the optimization problem make it even more challenging to find the optimal solution. This situation is further worsened especially when the building to be retrofitted has a large number of floors (or similar functional areas) that cause a linear increase in the dimension of the decision variables. The same problem is experienced by managers investigating retrofit options for a building portfolio consisting of multiple buildings. Thirdly, there are a large number of parameters to be obtained for the systematic whole-building retrofit problems. This usually requires a detailed energy audit of the buildings to be retrofitted, which is an expensive bottom-up modeling exercise.

Therefore, this study puts forward two methods to reduce the complexity of the systematic whole-building retrofit optimization problem and to eliminate the need for a bottom-up energy audit. These methods are based on the concept of grouping and measured energy savings data from sample retrofits.

The grouping method is used to categorize items to be retrofitted into several homogeneous groups [33]. Items are considered to be homogeneous and assigned to a group if they have a similar energy performance, inherent properties, working environment and operating schedules. This is motivated by the fact that energy-consuming systems in a building can be classified into lighting systems, HVAC systems, envelope systems, etc., and each of these systems usually consists of items that have the same characteristics. On a larger scale, each of these systems in a big building or building group can be treated as a virtual 'item' because of their similar functionality and characteristics.

Given the large number of items in a building for possible retrofit, it is very difficult to evaluate all the possibilities of retrofitting each energy-consuming item. In contrast, the dimension of the decision variables can be reduced significantly [34] by making use of the grouping method, because the solution will only determine whether a group of items should be retrofitted or not and which retrofit option should be chosen for the group instead of determine this for each single item. This is also in good agreement with the expectations of the decision makers because they will usually retrofit the whole group of similar units to facilitate easy maintenance and retrofit labor cost, etc.

In this study, items with the same energy performance and cost implications are grouped together. In addition, it must be pointed out that this study considers buildings with a large number of similarly designed and operated floors or functional areas. All homogeneous items within the boundary of a floor or a functional area comprise a subset of the overall homogeneous group of items for the whole building and will be termed an 'item' of the overall group in the rest of this study. For example, all light bulbs in a building belong to the same group and lambs on one floor constitute a virtual 'item' of the lighting group. After dividing the retrofitted items of the building into several homogeneous groups, the overall retrofit performance of the building, such as energy savings and cost, can be evaluated by investigating the performance of retrofitting an individual member and the number of retrofitted members of each homogeneous group.

The whole-building retrofit problem is further simplified by making use of measured energy savings achieved by retrofitting items of a homogeneous group. This is supported by the large number of energy conservation initiatives implemented across the world. In South Africa, for example, many building retrofit projects have been implemented and the energy savings of these projects have been quantified by the M&

V approach [35]. The verified energy savings of retrofitting different systems in a general building, including an envelope system, lighting system, HVAC system, etc., are the so-called 'notch test' data, which can be used to simplify the optimization problem. To be specific, knowing the potential energy savings and corresponding cost of retrofitting each subsystem on one floor of the building with a certain alternative, one can determine the best combination of subsystems and alternatives that could be used for the whole building retrofit so that the given objectives of the optimization problem are achieved.

In summary, the two models proposed categorize the items of target buildings into several homogeneous groups. Knowing the available energy savings of retrofitting an item of each group from existing retrofits, the models can work out systematic optimal retrofit plans for buildings by optimizing the numbers of virtual 'items' of each group and the retrofit options for them. The difference between the two methods is that the first one limits the retrofit options for the 'items' in the same group to be the same, while the second one does not.

The main contributions of this paper are listed as follows:

- Two simplified optimization models are proposed to reduce the complexity of systematic whole-building retrofit planning problems.
- The simplification is based on grouping method and 'notch test' data, which decreases the dimension of the optimal building retrofit planning problem and eliminates the need for a costly detailed bottom-up energy audit process.
- The models developed can help decision makers to determine energy-efficient and cost-effective whole-building retrofit plans in a computationally less expensive manner.
- The models take into account the South African green building policy based on EPC and the tax incentive initiative program for energy saving projects such that all possible benefits of the building retrofit project are explored and all constraints are considered in the planning phase.
- The proposed models can be of great help to decision makers to investigate retrofit plans for a building portfolio consisting of multiple buildings.
- The simplified models developed can be applied to similar building retrofit optimization projects that aim at reducing complexity and eliminating a comprehensive energy audit.

It is also noted that although the models presented are developed with particularly the South African environment in mind, they are applicable to general green building retrofit projects where energy intensity reduction and cost-effectiveness are the main concerns.

Because the systematic whole-building retrofit problem is a nonlinear mixed integer programming problem, modern optimization methods must be employed to solve this problem. Given the wide variety of modern optimization approaches, the literature has been investigated and it was found that the genetic algorithm is proved to be a better method to solve this type of problem [36]. In [37], a real coded genetic algorithm is proposed for solving integer and mixed integer optimization problems. Juan et al. [38] also chose a genetic algorithm to solve office building renovation problems considering energy performance improvement. The genetic algorithm is a method for solving optimization problems based on natural selection and evolutionary biology. It reflects the process of natural selection, which is that the fittest individuals are chosen for reproduction, aiming at producing offspring of the next generation. Genetic algorithms are capable of solving a variety of optimization problems, which cannot be dealt with by standard optimization algorithms, such as discontinuous, non-differentiable, mixed integer or highly nonlinear issues [39].

The remaining part of this paper is organized as follows. Two models for the simplification of the systematic whole-building retrofit problems are presented in Section 2. Section 3 provides a case study and results analysis. Conclusions are drawn in Section 4.

2. Optimization models

In this section, the aforesaid two simplified optimization models for systematic whole-building retrofit planning considering both the envelope components and the indoor appliances are developed. The purposes of the two optimization models are the same as those given in [32], which is to maximize the energy savings, minimize the payback period of building retrofit projects and make sure the buildings can obtain a good energy rating from the EPC standard for green building policy compliance.

The two simplified optimization models are built under the premise given below:

- The building to be retrofitted has the same structure for each floor.
- The intention is to retrofit energy consumption subsystems, such as lighting envelope, etc. on each floor of the building rather than single items. For instance, all the luminaries rather than part of them on one floor will be replaced with new ones if the lighting system on that floor is to be retrofitted.
- Proper maintenance for the items retrofitted during the project period is implemented so that the energy savings of the retrofit project are persistent.

In this study, the energy consumption in a building is divided into lighting systems, envelope systems (window and wall), HVAC systems (chiller and heat pump) and the roof system for upgrading with energy-efficient interventions. In addition, a PV power supply system is considered to be installed to reduce the building's energy demand from the grid [40] and ensure better life quality for occupants [41] owing to the rich solar resource in South Africa. Because the structure of each floor of the building is the same, the energy performance, inherent properties, working environment and operating schedules of the lighting subsystems and envelope subsystems of each floor are considered to be the same. According to grouping, all the lights within the building can be grouped into a homogeneous group, with all the lights installed on each floor as a virtual item of this group. The same is done for the envelope systems. The roof only has one item because for each building, there is only one roof structure. The HVAC systems in this study are of a centralized type. With this grouping and notch test data for retrofitting an item in these homogeneous groups, one can determine the impact of retrofitting a homogeneous group of items (subsystems) on the whole building.

Assume that there are I alternatives of windows and J alternatives of wall insulation materials for retrofitting the envelope systems, K alternatives of roof insulation materials for retrofitting the roof, C alternatives of chillers and H alternatives of heat pumps for retrofitting the HVAC systems, and P alternatives of solar panels for the PV system installation. For the lighting systems, assume that there are m types of existing lighting to be retrofitted and there are L_m alternatives for retrofitting the m-th type. It follows that there are $(I+1)(J+1)$ retrofit options for the envelope systems, $(C+1)(H+1)$ retrofit options for the HVAC systems, $(L_1+1)(L_2+1)$, ...,(L_m+1) retrofit options for the lighting systems, $(K+1)$ retrofit options for the roof, and $(P+1)$ options for the PV system installation. Let e, v and u denote that the e-th option for the envelope systems, the v-th option for the HVAC systems and u-th option for the lighting systems are chosen to replace the corresponding existing components, respectively. The values of e, v and u take integer values defined in (1)–(3).

$$e \in \{1, 2, ..., (I+1)(J+1)\}, \tag{1}$$

$$v \in \{1, 2, ..., (C+1)(H+1)\}, \tag{2}$$

$$u \in \{1, 2, ..., (L_1+1)(L_2+1)...(L_m+1)\}. \tag{3}$$

There is strong coupling between the envelope and the HVAC systems in their energy performance because the thermal performance of the envelope systems affects the load of the HVAC systems. As a consequence, these two subsystems are considered together to achieve energy savings. In this case, there are $(I+1)(J+1)(C+1)(H+1)$ retrofit options for the combined system. Let r, defined in (4), denote the r-th option for the combined system, i.e., the e-th option for the envelope systems and the v-th option for the HVAC systems, are chosen for the retrofit. The selection of the envelope, HVAC and lighting systems can thus be represented by the values of r and u.

$$r \in \{1, 2, \ldots, (I+1)(J+1)(C+1)(H+1)\}. \tag{4}$$

With the above information, the detailed formulations of the two models considering the retrofit of a building with F floors over the project period of T years are given in the following subsections.

2.1. Optimization model I

Optimization model I solves the whole building retrofit problem by assuming that the optimal retrofit options for each floor of the building are the same to simplify the problem further. For instance, if the e-th option for the envelope system and the u-th option for the lighting system are chosen by the optimization model, each floor of the building will use these options for its retrofit. Because the structure and functions of all the floors are the same, the optimization determines the optimal retrofit options r, u and the number of floors to retrofit their subsystems with these optimally selected options. In addition, the optimization will, at the same time, optimally determine the option of the PV system, the number of PV panels to be installed, and the optimal solution for the roof retrofit.

2.1.1. Decision variables of optimization model I

The decision variable of the systematic building retrofit optimization problem following optimization model I is given by:

$$X_1 = [r, u, N_{env,f}, N_{lig,f}, k, p, N_{pv}],$$

where $N_{env,f}$ denotes the number of floors to retrofit the envelope systems, $N_{lig,f}$ denotes the number of floors to retrofit the lighting systems, N_{pv} is the number of solar panels to be installed; $k \in \{1, 2, \ldots, (K+1)\}$ and $p \in \{1, 2, \ldots, (P+1)\}$ mean that the k-th roof alternative is chosen and the p-th solar panel alternative is installed, respectively.

2.1.2. Objectives of optimization model I

The objectives of the building retrofit project include energy savings and the payback period, which are important indicators to evaluate the profitability of an investment [42].

The energy savings of the building retrofit project in year t, $ES_1(t)$, can be calculated by

$$ES_1(t) = N_{lig,f} ES_{lig}(u) + ES_{rof}(k, v) + ES_{pv}(p) N_{pv} + N_{env,f} ES_{mix}(r) + (F - N_{env,f}) ES_{mix}(r - e + 1), \tag{5}$$

where $ES_{mix}(r)$ is the energy savings on one floor after retrofitting the floor's envelope and the building's HVAC system with the r-th option measured in Wh, $ES_{lig}(u)$ is the energy savings of retrofitting one floor's lighting system with the u-th option measured in Wh, $ES_{rof}(k, v)$ is the energy savings of retrofitting the roof of the building with its k-th option when the HVAC systems are retrofitted with the v-th option, measured in Wh and $ES_{pv}(p)$ is the energy production of one solar panel of the p-th option measured in Wh. The second term in (5) represents the energy savings achieved by retrofitting the centralized HVAC systems on the floors whose envelope systems are not retrofitted.

Taking into account the discount rate and the tax incentive program, the payback period of the building retrofit project T_{p1} is calculated by the following equations:

$$T_{p1} = t + \frac{|\overline{C_f}(t)|}{C_f(t+1)}, \tag{6}$$

$$C_f(t) = \frac{p(t) ES_1(t) + R(t)}{(1+d)^t} - C_{r1}, \tag{7}$$

$$R(t) = \begin{cases} (E_{pre} - E_{post}) \zeta_a \zeta_t, & t = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

In Eqs. (6)–(8), t is an integer and is the last period with a negative cumulative discounted cash flow, $\overline{C_f}(t)$ is the absolute value of cumulative cash flow at the end of period t measured in Dollar ($), $C_f(t+1)$ is the discounted cash flow in the period after t measured in $, $p(t)$ is the electricity price in year t measured in $/Wh, d is the discount rate, $R(t)$ is the tax incentive measured in $, E_{pre} and E_{post} are the total energy consumption of the building before and after the retrofit, respectively, measured in Wh/year, ζ_a is the allowance rate and ζ_t is the tax rate for general businesses in South Africa. C_{r1} is the retrofit cost making use of optimization model I measured in $ and can be calculated by

$$C_{r1} = N_{env,f}(C_{mix}(r) - C_{hva}(v)) + N_{lig,f} C_{lig}(u) + C_{rof}(k) + C_{pv}(p) N_{pv} + C_{hva}(v), \tag{9}$$

where $C_{mix}(r)$ is the cost of retrofitting one floor's envelope systems and the building's HVAC systems with the r-th option measured in $, $C_{lig}(u)$ is the cost of retrofitting one floor's lighting system with the u-th option measured in $, $C_{hva}(v)$ is the cost of retrofitting the HVAC systems of the building with the v-th option measured in $, $C_{rof}(k)$ is the cost of retrofitting the roof of the building with the k-th option measured in $, $C_{pv}(p)$ is the cost of one solar panel of the p-th option measured in $.

In the literature, the weighted sum method was widely used to solve multiple objective optimization problems [43]. For instance, Kim and de Weck [44] investigated adaptive weighted sum method for multi-objective issues and [45] employed this approach for energy-efficient investment decision problems. Therefore, the weighted sum method is chosen to solve the optimization problem formulated, resulting the following objective function:

$$J = -w_1 \sum_{t=1}^{T} ES_1(t) + w_2 T_{p1}. \tag{10}$$

2.1.3. Constraints of optimization model I

The constrains of the optimal retrofit problem include three parts, which are the EPC limit, budget limit and physical limits.

The EPC rating system assigns a rating to a building based on its energy intensity compared to a reference value set by the South African national standard [46]. The requirements of getting a certain rating from the EPC are detailed in Table 1. The item E_r is the reference energy intensity, which depends on the occupancy class and location of the building.

Based on the requirements of different ratings in Table 1, the EPC limit used to ensure that the building obtains the desired rating from the EPC standard for the purpose of green building policy compliance, can be described by the following general formulas [18]:

$$E_p < \delta E_r, \tag{11}$$

Table 1
Energy performance scale.

Grade	Requirement
A	Energy intensity $< 0.3E_r$
B	$0.3E_r \leqslant$ Energy intensity $< 0.6E_r$
C	$0.6E_r \leqslant$ Energy intensity $< 0.9E_r$
D	$0.9E_r \leqslant$ Energy intensity $< 1.1E_r$
E	$1.1E_r \leqslant$ Energy intensity $< 1.4E_r$
F	$1.4E_r \leqslant$ Energy intensity $< 1.7E_r$
G	Energy intensity $\geqslant 1.7E_r$

$$E_p = \frac{E_{post}}{A_g}, \tag{12}$$

where E_p denotes the energy intensity of the building measured in kWh/m^2, A_g is the gross area of the building measured in m^2, δ is a coefficient, taking the values from Table 1. For instance, $\delta = 1.1$ means that at least a D rating must be obtained for the building.

The budget limit for the retrofit can be described with the following formula:

$$C_{r1} \leqslant \beta, \tag{13}$$

where β is the retrofit budget measured in $.

The physical limits include the available roof area for the PV system installation, given as follows:

$$A_{pv}(p)N_{pv} \leqslant A_{eff}, \tag{14}$$

where $A_{pv}(p)$ is the area of one solar panel of the p-th option measured in m^2 and A_{eff} is the usable area of the roof for PV system installation measured in m^2.

All the decision variables must satisfy the following integer constraints:

$$\begin{aligned} &N_{env,f} \in \{0, 1, ..., F\},\\ &N_{lig,f} \in \{0, 1, ..., F\},\\ &r \in \{1, 2, ..., (I+1)(J+1)(C+1)(H+1)\},\\ &u \in \{1, 2, ..., (L_1+1)(L_2+1)...(L_m+1)\},\\ &k \in \{1, 2, ..., (K+1)\},\\ &p \in \{1, 2, ..., (P+1)\}. \end{aligned} \tag{15}$$

2.2. Optimization model II

Based on optimization model I and on [32], which allows each item to flexibly choose desired alternatives for retrofit, one can naturally think of a second simplified method, which might find better solutions compared with model I. The differences between the two methods are detailed as follows.

- Method II makes it possible for each floor to have different retrofit options, *i.e.* the same subsystem on all floors can be retrofitted with different options.
- It might be capable of making more complete use of investment and finding better retrofit plans owing to the flexibility of retrofit options.
- It might make a relatively small compromise in the complexity reduction of the retrofit optimization problem.

Since each floor of the building can determine whether its subsystems are to be retrofitted or not, the aim of second optimization model is to prepare an optimal retrofit plan for the whole-building retrofit with a given budget by determining the retrofit states and retrofit options for the energy-consuming subsystems of each floor and the roof and HVAC systems of the building, the installation option for the PV system and the number of solar panels to be installed.

2.2.1. Decision variables of optimization model II

The decision variable of the building retrofit optimization following model II is described by:

$$X_2 = [v, e_1, ..., e_f, ..., e_F, u_1, ..., u_f, ..., u_F, k, p, N_{pv}],$$

where e_f and u_f denote that the e_f-th option for the envelope system and the u_f-th option for the lighting systems are chosen for retrofitting the f-th floor.

2.2.2. Objectives of optimization model II

The same objectives, including energy savings and the payback period, are considered.

The energy savings of the building retrofit project in year t, $ES_2(t)$, can be calculated by the following equation:

$$ES_2(t) = \sum_{f=1}^{F} (ES_{mix}(v, e_f) + ES_{lig}(u_f)) + ES_{rof}(k, v) + ES_{pv}(p)N_{pv}, \tag{16}$$

where $ES_{mix}(v, e_f)$ is the energy savings of the f-th floor after its envelope systems retrofitted with the e_f-th option and the building's HVAC systems have been retrofitted with the v-th option, measured in Wh, $ES_{lig}(u_f)$ is the energy savings of the f-th floor after its lighting systems have been retrofitted with the u_f-th option, measured in Wh.

The resulting retrofit cost of the second model, C_{r2}, can be calculated by

$$C_{r2} = \sum_{f=1}^{F} [C_{mix}(v, e_f) - C_{hva}(v) + C_{lig}(u_f)] + C_{rof}(k) + C_{pv}(p)N_{pv} + C_{hva}(v), \tag{17}$$

where $C_{mix}(v, e_f)$ is the cost of retrofitting the building's HVAC systems with the v-th option and the envelope systems of the f-th floor with the e_f-th option, measured in $, $C_{lig}(u_f)$ is the cost of retrofitting the lighting systems of the f-th floor with its u_f-th option, measured in $.

The payback period of the building retrofit project T_{p2} can be calculated following Eqs. (6)–(8).

Taking advantage of the Eqs (16) and (17), the objective function of this model is given by

$$J = -w_1 \sum_{t=1}^{T} ES_2(t) + w_2 T_{p2}. \tag{18}$$

2.2.3. Constraints of optimization model II

The budget limit can be described with the following in equation:

$$C_{r2} \leqslant \beta. \tag{19}$$

The EPC rating limit can be described with formulas (11) and (12). The PV installation area limit is described by formula (14). The limits on the design variables are:

$$\begin{aligned} &k \in \{1, 2, ..., (K+1)\},\\ &p \in \{1, 2, ..., (P+1)\},\\ &v \in \{1, 2, ..., (C+1)(H+1)\},\\ &e_f \in \{1, 2, ..., (I+1)(J+1)\},\\ &\forall f \in \{1, 2, ..., F\},\\ &u_f \in \{1, 2, ..., (L_1+1)(L_2+1)...(L_m+1)\},\\ &\forall f \in \{1, 2, ..., F\}. \end{aligned} \tag{20}$$

3. Case study

3.1. Case information

In this section, an existing office building is used as a case study to verify the viability of the two optimization models. The building studied comprises six floors with the same structure, shown in Fig. 1. The area of each floor is 266 m^2. Before the retrofit, the EPC rating of the building under study is grade E. Therefore, this building has to improve its energy efficiency to achieve a D rating at least to comply with the green building policy. The information on the alternatives for retrofitting the envelope, lighting, HVAC and roof systems and installing the PV system are detailed in Tables 2–8, (The data in this paper are obtained from manufactures product datasheets, data obtained from hundreds of M&V projects, published technical reports, and the South African national standards, etc.). For example, Table 8 gives the information of the alternative lighting technologies used to retrofit the corresponding existing lighting technologies. The economic parameters

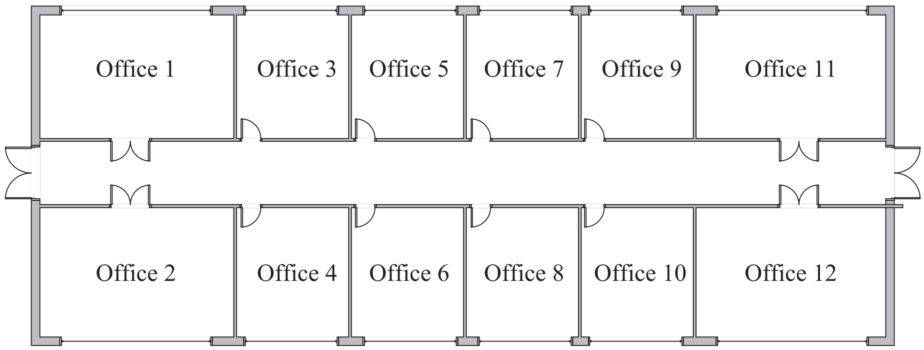


Fig. 1. Floor design of the office building under study.

Table 2
Window alternatives.

i	Alternatives	U_i (W/m °C)	$C_{win,i}$ (\$/m^2)
1	Double glazing, tinted uncoated air-filled metallic frame	0.49	50.00
2	Double glazing, tinted coated air-filled metallic frame	0.38	80.00
3	Double glazing, low-e window, air-filled metallic frame	0.32	97.00

Table 3
Wall insulation material alternatives.

j	Alternatives	d_j (m)	λ_j (W/m °C)	$C_{wal,j}$ (\$/m^2)
1	Glass wool	0.05	0.038	16.32
2	EPS	0.08	0.033	21.10
3	Cork	0.30	0.040	69.38

Table 4
Roof insulation material alternatives.

k	Alternatives	d_k (m)	λ_k (W/m °C)	$C_{rof,k}$ (\$/m^2)
1	SPF	0.020	0.042	8.23
2	EPS	0.060	0.033	10.49
3	Stone wool	0.105	0.037	44.84

Table 5
Chiller alternatives.

c	Alternatives	SEER	$C_{chi,c}$ (\$)
1	Trane chiller type 1	17.0	8580
2	Trane chiller type 2	15.0	7590

Table 6
Heat pump alternatives.

h	Alternatives	HSPF	$C_{pum,h}$(\$)
1	Trane heat pump type 1	9.5	7920
2	Trane heat pump type 2	8.6	7425

Table 7
Solar panel alternatives.

p	Alternatives	$C_{pv,p}$ (\$)	η_l (%)	$A_{pv,p}$ (m^2)
1	YL190P-23B	592.62	14.7	1.297
2	CS6X-300P	870.33	15.6	1.919
3	SW 275 MONO	1042.50	16.4	1.593

Table 8
Lighting technology alternatives.

l_m	Existing lighting	N_{l_m}	Alternatives	C_{lig_m,l_m} (\$)
l_1	2-lamp 4′ T8 fixture 70 W	80		
			2-lamp 4′ T5 14 W	19.0
			2-lamp 4′ T5 18 W	20.5
			2-lamp 4′ T5 36 W	10.0
l_2	PAR 38–65 W	48		
			CFL lamp 7 W	35.4
			CFL lamp 14 W	37.1
			CFL lamp 20 W	27.6
l_3	Incandescent 100 W	32		
			LED bulb 12 W	79.5
			LED bulb 17 W	53.0
			LED bulb 20 W	42.4

involved in the optimization models include the discount rate and the increased rate of the electricity price, which are determined as 6% and 12.69%, respectively, according to South Africa's economic statistics and the largest utility, Eskom, in South Africa.

3.2. Data collection

According to Section 2, there are 144 retrofit options for the combined envelope and HVAC systems, 64 retrofit options for the lighting systems, 36 retrofit options for the roof considering the HVAC systems and four options for the PV system installation. The notch test data on retrofitting the envelope, lighting, HVAC and roof of the building and installing a roof-top PV system on the building obtained following the M&V method are detailed in Tables 11–14, which are presented in the appendix. For instance, the resulting energy savings and the corresponding cost of retrofitting the envelope of one floor and the HVAC systems of the building with different combined options are detailed in Table 11. The numbers in the row corresponding to $r = 103$ of Table 11 detail the 103-rd retrofit option for this floor's envelope and the building's HVAC systems. Specifically, the data mean that the heat pump in the HVAC is not retrofitted, the chiller is retrofitted with its second alternative listed in Table 6, the windows are replaced with the first alternative listed in Table 2 and the walls are fitted with the second insulation alternative given in Table 3. Retrofitting one floor with this option results in 5349 kWh energy savings and costs \$14335.

In this study, the building retrofit optimization problem is solved by a genetic algorithm. To investigate the impact of investments on the optimal retrofit plans, the results of applying the optimal plans obtained by the two optimization models proposed in Section 2 with different budgets are presented in the following sections. Optimization results with different budgets set to \$10,000, \$25,000, \$45,000 and \$200,000 are analyzed.

The two optimization models are both solved using the weighted sum method to give decision makers a convenient way to obtain a desired retrofit plan according to their preferences on different objectives

Table 9
Results of applying optimization model I with different budgets.

Description	Budget1	Budget2	Budget3	Budget4
β ($)	10000	25000	45000	200000
C_{r1} ($)	9263	24586	44683	196490
r	1	1	49	85
$N_{env,f}$	0	0	0	2
u	52	24	23	22
$N_{lig,f}$	4	6	6	6
(v, k)	1	1	13	21
p	2	2	2	3
N_{pv}	2	0	16	163
T_{p1} (month)	22	22	27	59
ES_1 (kWh)	561286	1501978	1873954	2530403
E_p	0.927	0.617	0.494	0.278
RSD of T_{p1}	2.67%	2.65%	0.83%	3.55%
RSD of ES_1	3.40%	4.46%	0.14%	0.16%
RSD of E_p	0.68%	3.29%	0.18%	0.48%

by tuning the weighting factors. In order to verify this, the effectiveness of tuning the weighting factors is studied. The impact of the tax incentive program on the optimal retrofit plan is also analyzed.

3.3. Results analysis

3.3.1. Results analysis of optimization model I

To verify the feasibility of the first optimization model for systematic whole-building retrofit planning, the optimal solutions with different budgets based on the method are presented in Table 9.

In Table 9, the detailed optimal retrofit plans for the building with different investments are indicated by the contents from the fourth (starting with C_{r1} ($)) to the tenth row. r represents the retrofit options for the envelope systems of each floor and the HVAC systems of the building. u represents the retrofit option for the lighting systems of each floor. $N_{env,f}$ and $N_{lig,f}$ indicate the numbers of floors to retrofit their envelope systems and lighting systems, respectively. (v, k) represents the retrofit options listed in Table 13 for the roof system of the building considering the HVAC systems. p and N_{pv} indicate the option and number of installed PV panels shown in Table 14. For instance, the number '85' for r means that the 85-th option for the envelope systems and the HVAC systems is chosen for retrofit with a budget of $200000. The number '49' for r means the 49-th option is chosen, which indicates that the envelope systems of the building are not retrofitted. Only the HVAC systems of the building are retrofitted with the budget of $45000. '2' for $N_{env,f}$ means that the envelope systems of two floors of the building are retrofitted. The number '23' for u and '6' for $N_{lig,f}$ in the fourth column represent that the lighting systems of all six floors are retrofitted with the 23-rd option with a budget of $45000. The numbers '13' and '21' for (v, k) both represent that the roof system of the building is not retrofitted, referring to Table 13. The number '2' for p and '16' for N_{pv} in the fourth column mean that the second option in Table 14 is chosen for setting up the PV system and 16 of the selected solar panels are installed.

When investigating the optimal retrofitting solutions in Table 9, it is observed that the proposed methods do not simply choose the cheapest options or the most energy-efficient ones. For instance, the optimal retrofit plan with a budget of $10000 selects the 52-th option from Table 8 for retrofitting the lighting systems, which is not the cheapest or the most energy-efficient option among the alternatives in Table 8.

The items ES_1 and T_{p1} represent the resulting energy savings and payback period of the building retrofit project making use of optimization model I. It can be seen that the energy savings and payback period keep increasing with growing budgets. The reason for this phenomenon is that more investments allow more systems to be retrofitted, thereafter resulting in more energy savings and longer payback periods. One also finds that the growth rate of the payback period increases with growing budgets. This is because more and more systems with long payback are retrofitted when the budget increases. For instance, only the lighting systems are retrofitted with the budget of $25,000. However, 16 solar panels are installed with the budget of $45,000. When the budget increases to $200,000, more solar panels are installed and the envelope systems of some floors are also retrofitted.

An interesting phenomenon is that the payback period with a budget of $10,000 is the same as that with a budget of $25,000. This can be explained by the cost-effectiveness of retrofitting different subsystems with different options. Retrofitting the lighting systems is the most cost-effective method to save energy, followed by retrofitting the HVAC systems. Installing a PV system and retrofitting the envelope systems require long payback periods. With the budget of $25000, all the investment is used to retrofit the lighting systems, while part of the investment is used to install a PV system with the budget of $10,000. In addition, the 24-th option chosen for retrofitting the lighting systems with the budget of $25,000 is more energy-efficient compared with the 52-nd one and results in a relatively shorter payback period. This explains the nearly identical payback periods of the two investments.

When investigating the optimal retrofit actions with different budgets, one finds that the lighting systems of four floors of the building are retrofitted with the 52-nd option and two solar panels of its second option in Table 14 are installed with the budget of $10,000. However, the lighting systems of all the floors are retrofitted while no PV panel is installed when the budget increases to $25,000. When the investment grows to $45,000, all the lighting systems of the building are retrofitted with a more energy-efficient option and the HVAC systems are also retrofitted. In addition, 16 solar panels of the second option are installed. With an even higher budget, $200,000, available, the envelope systems of some floors are retrofitted, better options are selected for retrofitting other subsystems, and more solar panels are installed. In view of the above, a conclusion can be drawn that the investment gives priority to the subsystems of the building in the order of the lighting, HVAC, PV, envelope and the roof. This is because retrofitting the lighting systems is the most cost-effective choice to save energy, followed by the HVAC systems. Retrofitting the envelope and roof systems and installing a PV power supply system take a long time to pay back the cost in spite of their large energy saving potentials.

One of the purposes of this study is to improve the energy efficiency of the building to achieve a good EPC rating for green building policy compliance. In Table 9, E_p represents the energy performance of the building after applying the optimal retrofit plan obtained from optimization model I. Compared with the reference value in Table 1, one finds that the four optimal retrofit plans obtained with budgets of $10000, $25000, $45000 and $200000 can help the building to get a D, C, B and A rating from EPC, respectively.

Because the problem is solved by a genetic algorithm, which is essentially a metaheuristic method, the variance of the solutions must be investigated. In Table 9, the RSD values [47] of T_{p1}, ES_1 and E_p represent the relative standard deviations of the payback period, energy savings of the building retrofit project and the energy performance of the building achieved by the retrofit calculated from 20 runs of the genetic algorithm, respectively. It can be seen that the RSD values of these items are less than 5%, which means the results obtained with optimization model I are stable.

To demonstrate the effectiveness of weighting parameter tuning, the optimization problem is solved with two more sets of weighting factors and the results obtained are presented in Fig. 2. For convenience of comparison, other factors that affect the retrofit project remain the same. In particular, the budget is kept at $10000 and the tax incentive program is taken into account during these optimization processes. In Fig. 2, it can be seen that the energy savings increase and the payback period decreases when their corresponding weighting factors grow. For instance, the percentage of energy savings of the building retrofit project increases from 1.9% to 17.2% when the value of its corresponding

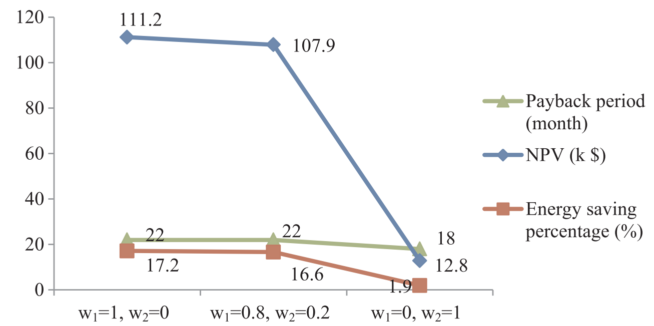


Fig. 2. Optimal results obtained by optimization model I with different weighting factors.

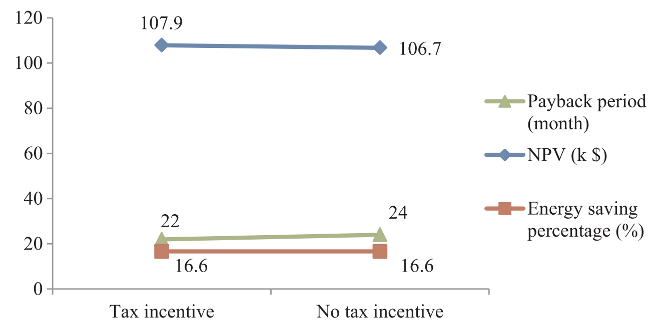


Fig. 3. Impact of tax incentive on the optimal results obtained by optimization model II with $w_1 = 0.8$, $w_2 = 0.2$, β = "$"10, 000.

Table 10
Results of applying optimization model II with different budgets.

Description	Budget1	Budget2	Budget3	Budget4
β ($)	10000	25000	45000	200000
C_{r2} ($)	9860	24925	44936	199593
r_1	1	1	49	69
r_2	1	1	49	65
r_3	1	1	49	69
r_4	1	1	49	65
r_5	1	1	49	65
r_6	1	1	49	69
u_1	56	23	23	23
u_2	1	24	23	22
u_3	1	24	24	22
u_4	64	24	23	22
u_5	1	24	23	22
u_6	64	24	23	22
(v, k)	1	1	13	17
p	4	2	2	3
N_{pv}	0	0	17	163
T_{p2} (month)	22	22	27	60
ES_2 (kWh)	594086	1504742	1875121	2531403
E_p	0.916	0.616	0.494	0.277
RSD of T_{p2}	2.00%	1.95%	3.55%	2.69%
RSD of ES_2	1.74%	3.47%	3.64%	0.64%
RSD of E_p	0.36%	2.54%	3.62%	1.79%

weighting factor w_1 changes from zero to one. The payback period of the project decreases from 22 months to 18 months when the value of its corresponding weighting factor w_2 increases from zero to one. In view of the results in Fig. 2, it can be concluded that optimization model I gives decision makers the flexibility of obtaining the desired result according to their preferences on energy savings or payback period.

To investigate the impact of the tax incentive program on the building retrofit project, the optimization problem is solved by optimization model I without considering the tax incentive and the results are presented in Fig. 3. In Fig. 3, one finds that considering the tax incentive program in the optimization process results in a slightly shorter payback period and higher net present value. This verifies that the tax incentive program is capable of further shortening the payback period of the building retrofit project.

3.3.2. Results analysis of optimization model II

The optimal solutions obtained by optimization model II with different budgets are provided in Table 10, in which r_1, r_2, r_3, r_4, r_5 and r_6 represent the retrofit options from Table 11 for the envelope systems of the six floors and the HVAC system of the building. u_1, u_2, u_3, u_4, u_5 and u_6 represent the retrofit options from Table 12 for the lighting systems of the six floors.

The same trend reported in Section 3.3.1 is observed in Table 10. For example, retrofit plans obtained by optimization model II with budgets of $10000, $25000, $45000 and $200000 can help the building under study to get a D, C, B and A rating from the EPC, respectively. The RSD values of T_{p2}, ES_2 and E_p are less than 5%, which verifies the stability of optimization model II in finding optimal retrofit plans for buildings.

3.4. Comparison of the two models

Following the two simplified methods, the number of decision variables of compiling a systematic retrofit plan for a whole building is reduced from hundreds (even more) to a small value compared with that of [32]. Both methods reduce the complexity of solving whole-building retrofit problems and eliminate the need for a comprehensive energy audit.

From a theoretical point of view, model I and model II differ in resolution of the grouping of items to be retrofitted. Model I essentially groups all items on a floor as a virtual item, whereas model II treats subsystems such as lighting systems and HVAC systems, as individual items. More detailed grouping in model II contributes to better utilization of the available investment, as discussed earlier. It must be pointed out at this stage that the differences in the grouping method adopted by the two models will be studied further in future research on how to design an optimal model and a corresponding grouping method that results in an acceptable precision and confidence level of the model predicted energy savings while effectively reducing the complexity of the retrofit optimization problem. This is particularly relevant because it was shown by researchers from the same research group that similar grouping methods will not affect the final performance of the retrofit planning problem significantly [33]. In other words, theoretical comparison of the two simplification models presented in this study is still an open research question and the design of a method to select an optimization model considering its precision and complexity at the same time is being actively investigated currently.

For practical applications, conclusions on how to select the two developed models for a specific application are drawn as detailed below, according to the findings of the case study.

For small-scale building retrofit problems, model II is more accurate and performs better than model I. This is because model II allows the retrofit options for all the subsystems in the building to be different. Its flexibility promotes better utilization of the available investment. This can be verified by dividing the C_r by β in Tables 9 and 10. The results show that the utilization rate of the budget is between 98.6% and 99.8% with model II, while the same rate ranges from 92.6% to 99.3% with model I. In fact, the results in Tables 9 and 10 indicate that model II produces better results than model I in terms of absolute energy savings. For instance, 33 MWh extra energy is saved by model II with a budget of $10000 compared to model I.

With respect to building retrofit problems with a large number of floors involved, model I is simpler and more effective than model II. This is because the dimension of the optimization problem is much less for model I compared with that of model II, especially when a large number of floors are involved. There are only five decision variables in optimization model I, whereas the number of decision variables in optimization model II is $2F + 4$, which depends on the number of floors in the building. When the number of floors increases, the number of decision variables of model I will remain unchanged, while that of model II will increase rapidly. Therefore, solving retrofit problems for buildings with a large number of floors using model II is relatively more difficult compared to using model I. In addition, the solution obtained with method II might sometimes be very poor when a large number of decision variables are involved because of the inefficiency of existing algorithms to solve integer programming problems.

In practical applications, one first need to obtain information of the target building to be retrofitted, such as its existing energy-consuming systems, and conduct a notch test of retrofitting a certain item by a particular alternative (this can also be taken from similar projects). Then, the developed models in this paper can be directly used with the obtained parameters to solve for the optimal retrofit plan. The idea of the simplified models has already been used in hundreds of M&V energy saving projects undertaken by the Center of M&V at the University of Pretoria, such as the energy efficiency lighting projects [8]. In addition, the proposed models are useful to help decision makers obtain optimal retrofit plans for a building portfolio consisting of multiple buildings, which is a common challenge in practice.

4. Conclusion

In this study, two simplified optimization models are proposed to reduce the complexity of systematic whole-building retrofit planning problems considering both the envelope components and indoor appliances. The two retrofit models aim at saving energy and achieving desired green building ratings by implementing energy-efficient interventions in the most cost-effective way. The simplification is done by using a grouping method and measured and verified energy savings of sample retrofits. The simplification not only reduces the complexity of the retrofit optimization problem in terms of technical difficulty and computational load of solving the problem, but also helps to obviate the need for an expensive detailed energy audit to support the retrofit planning. The two models proposed are tested with a case study and both are shown to be effective in achieving the objectives of this study. The simplified optimization methods are suitable for reducing complexity and eliminating a detailed energy audit of all building retrofit optimization problems.

Appendix A

Tables 11–14.

Table 11
Notch test data of retrofitting a floor's envelope and the building's HVAC system.

$r(v, e)$	Chiller	Heat pump	Window	Wall	$ES^{mix}(r)$ (kWh)	$C^{mix}(r)$ ($)
1	0	0	0	0	0	0
2	0	0	0	1	67	2544
3	0	0	0	2	75	3289
4	0	0	0	3	83	10815
5	0	0	1	0	2393	3456
6	0	0	1	1	2460	6000
7	0	0	1	2	2468	6745
8	0	0	1	3	2476	14271
9	0	0	2	0	2018	5530
10	0	0	2	1	2085	8074
11	0	0	2	2	2093	8819
12	0	0	2	3	2101	16345
13	0	0	3	0	2144	6705
14	0	0	3	1	2211	9249
15	0	0	3	2	2219	9994
16	0	0	3	3	2227	17520
17	0	1	0	0	290	7920
18	0	1	0	1	320	10464
19	0	1	0	2	323	11209
20	0	1	0	3	327	18735
21	0	1	1	0	2656	11376
22	0	1	1	1	2685	13920
23	0	1	1	2	2689	14665
24	0	1	1	3	2693	22191
25	0	1	2	0	2280	13450
26	0	1	2	1	2309	15994
27	0	1	2	2	2313	16739
28	0	1	2	3	2317	24265
29	0	1	3	0	2406	14625
30	0	1	3	1	2435	17169
31	0	1	3	2	2439	17914
32	0	1	3	3	2443	25440
33	0	2	0	0	279	7425
34	0	2	0	1	310	9969
35	0	2	0	2	314	10714
36	0	2	0	3	318	18240
37	0	2	1	0	2646	10881
38	0	2	1	1	2677	13425
39	0	2	1	2	2681	14170
40	0	2	1	3	2684	21696
41	0	2	2	0	2270	12955

(continued on next page)

Table 11 (*continued*)

$r(v, e)$	Chiller	Heat pump	Window	Wall	$ES^{mix}(r)$ (kWh)	$C^{mix}(r)$ ($)
42	0	2	2	1	2301	15499
43	0	2	2	2	2305	16244
44	0	2	2	3	2308	23770
45	0	2	3	0	2396	14130
46	0	2	3	1	2427	16674
47	0	2	3	2	2431	17419
48	0	2	3	3	2434	24945
49	1	0	0	0	4870	8580
50	1	0	0	1	4924	11124
51	1	0	0	2	4930	11869
52	1	0	0	3	4937	19395
53	1	0	1	0	5391	12036
54	1	0	1	1	5445	14580
55	1	0	1	2	5452	15325
56	1	0	1	3	5458	22851
57	1	0	2	0	5315	14110
58	1	0	2	1	5369	16654
59	1	0	2	2	5376	17399
60	1	0	2	3	5382	24925
61	1	0	3	0	5342	15285
62	1	0	3	1	5396	17829
63	1	0	3	2	5402	18574
64	1	0	3	3	5409	26100
65	1	1	0	0	5160	16500
66	1	1	0	1	5177	19044
67	1	1	0	2	5179	19789
68	1	1	0	3	5181	27315
69	1	1	1	0	5655	19956
70	1	1	1	1	5671	22500
71	1	1	1	2	5673	23245
72	1	1	1	3	5675	30771
73	1	1	2	0	5577	22030
74	1	1	2	1	5594	24574
75	1	1	2	2	5596	25319
76	1	1	2	3	5598	32845
77	1	1	3	0	5604	23205
78	1	1	3	1	5620	25749
79	1	1	3	2	5622	26494
80	1	1	3	3	5624	34020
81	1	2	0	0	5149	16005
82	1	2	0	1	5167	18549
83	1	2	0	2	5169	19294
84	1	2	0	3	5171	26820
85	1	2	1	0	5645	19461
86	1	2	1	1	5663	22005
87	1	2	1	2	5665	22750
88	1	2	1	3	5667	30276
89	1	2	2	0	5568	21535
90	1	2	2	1	5586	24079
91	1	2	2	2	5588	24824
92	1	2	2	3	5590	32350
93	1	2	3	0	5594	22710
94	1	2	3	1	5612	25254
95	1	2	3	2	5614	25999
96	1	2	3	3	5616	33525
97	2	0	0	0	4701	7590
98	2	0	0	1	4756	10134
99	2	0	0	2	4762	10879
100	2	0	0	3	4769	18405
101	2	0	1	0	5288	11046
102	2	0	1	1	5342	13590
103	2	0	1	2	5349	14335
104	2	0	1	3	5355	21861
105	2	0	2	0	5201	13120
106	2	0	2	1	5256	15664
107	2	0	2	2	5262	16409
108	2	0	2	3	5269	23935
109	2	0	3	0	5231	14295
110	2	0	3	1	5286	16839
111	2	0	3	2	5292	17584
112	2	0	3	3	5299	25110
113	2	1	0	0	4992	15510
114	2	1	0	1	5009	18054
115	2	1	0	2	5011	18799
116	2	1	0	3	5013	26325

(*continued on next page*)

Table 11 (*continued*)

$r(v, e)$	Chiller	Heat pump	Window	Wall	$ES^{mix}(r)$ (kWh)	$C^{mix}(r)$ ($)
117	2	1	1	0	5551	18966
118	2	1	1	1	5568	21510
119	2	1	1	2	5570	22255
120	2	1	1	3	5572	29781
121	2	1	2	0	5463	21040
122	2	1	2	1	5481	23584
123	2	1	2	2	5483	24329
124	2	1	2	3	5485	31855
125	2	1	3	0	5493	22215
126	2	1	3	1	5510	24759
127	2	1	3	2	5512	25504
128	2	1	3	3	5514	33030
129	2	2	0	0	4981	15015
130	2	2	0	1	4999	17559
131	2	2	0	2	5001	18304
132	2	2	0	3	5004	25830
133	2	2	1	0	5541	18471
134	2	2	1	1	5560	21015
135	2	2	1	2	5562	21760
136	2	2	1	3	5564	29286
137	2	2	2	0	5454	20545
138	2	2	2	1	5472	23089
139	2	2	2	2	5474	23834
140	2	2	2	3	5477	31360
141	2	2	3	0	5483	21720
142	2	2	3	1	5502	24264
143	2	2	3	2	5504	25009
144	2	2	3	3	5506	32535

Table 12
Notch test data of retrofitting the lighting system of one floor.

u	Light 1	Light 2	Light 3	$ES^{lig}(u)$ (kWh)	$C^{lig}(u)$ ($)
1	0	0	0	0	0
2	0	0	1	8110	2544
3	0	0	2	7649	1696
4	0	0	3	7373	1357
5	0	1	0	7016	1487
6	0	1	1	15126	4031
7	0	1	2	14665	3183
8	0	1	3	14388	2844
9	0	2	0	6169	1558
10	0	2	1	14279	4102
11	0	2	2	13818	3254
12	0	2	3	13542	2915
13	0	3	0	5443	1158
14	0	3	1	13553	3702
15	0	3	2	13092	2854
16	0	3	3	12816	2514
17	1	0	0	10644	1254
18	1	0	1	18755	3798
19	1	0	2	18294	2950
20	1	0	3	18017	2611
21	1	1	0	17660	2741
22	1	1	1	25770	5285
23	1	1	2	25309	4437
24	1	1	3	25033	4098
25	1	2	0	16813	2812
26	1	2	1	24924	5356
27	1	2	2	24463	4508
28	1	2	3	24186	4169
29	1	3	0	16088	2412
30	1	3	1	24198	4956
31	1	3	2	23737	4108
32	1	3	3	23460	3768
33	2	0	0	9884	1354
34	2	0	1	17994	3898
35	2	0	2	17533	3050
36	2	0	3	17257	2711

(*continued on next page*)

Table 12 (*continued*)

u	Light 1	Light 2	Light 3	$ES^{lig}(u)$ (kWh)	$C^{lig}(u)$ ($)
37	2	1	0	16900	2841
38	2	1	1	25010	5385
39	2	1	2	24549	4537
40	2	1	3	24273	4198
41	2	2	0	16053	2913
42	2	2	1	24163	5457
43	2	2	2	23702	4609
44	2	2	3	23426	4269
45	2	3	0	15327	2512
46	2	3	1	23437	5056
47	2	3	2	22977	4208
48	2	3	3	22700	3869
49	3	0	0	6463	663
50	3	0	1	14573	3207
51	3	0	2	14112	2359
52	3	0	3	13836	2019
53	3	1	0	13478	2149
54	3	1	1	21588	4693
55	3	1	2	21128	3845
56	3	1	3	20851	3506
57	3	2	0	12632	2221
58	3	2	1	20742	4765
59	3	2	2	20281	3917
60	3	2	3	20004	3578
61	3	3	0	11906	1820
62	3	3	1	20016	4364
63	3	3	2	19555	3516
64	3	3	3	19279	3177

Table 13
Notch test data of retrofitting the roof considering the HVAC retrofit.

v,k	Chiller	Heat pump	Roof	$ES^{rof}(k)$ (kWh)	$C^{rof}(k)$ ($)
1	0	0	0	0	0
2	0	0	1	81	2189
3	0	0	2	122	2790
4	0	0	3	131	11927
5	0	1	0	83	0
6	0	1	1	119	2189
7	0	1	2	137	2790
8	0	1	3	141	11927
9	0	2	0	80	0
10	0	2	1	118	2189
11	0	2	2	137	2790
12	0	2	3	141	11927
13	1	0	0	29	0
14	1	0	1	94	2189
15	1	0	2	128	2790
16	1	0	3	134	11927
17	1	1	0	112	0
18	1	1	1	132	2189
19	1	1	2	143	2790
20	1	1	3	145	11927
21	1	2	0	109	0
22	1	2	1	131	2189
23	1	2	2	142	2790
24	1	2	3	144	11927
25	2	0	0	28	0
26	2	0	1	94	2189
27	2	0	2	127	2790
28	2	0	3	134	11927
29	2	1	0	111	0
30	2	1	1	132	2189
31	2	1	2	142	2790
32	2	1	3	145	11927
33	2	2	0	108	0
34	2	2	1	130	2189
35	2	2	2	142	2790
36	2	2	3	144	11927

Table 14
Notch test data of installing one solar panel.

p	$ES^{pv}(p)$ (kWh)	$C^{pv}(p)$ ($)
1	0	0
2	393	593
3	408	870
4	402	1043

References

[1] Huang Y, Niu J, Chung T. Study on performance of energy-efficient retrofitting measures on commercial building external walls in cooling-dominant cities. Appl Energy 2013;103:97–108.
[2] Diakaki C, Grigoroudis E, Kolokotsa D. Towards a multi-objective optimization approach for improving energy efficiency in buildings. Energy Build 2008;40(9):1747–54.
[3] Jafari A, Valentin V. An optimization framework for building energy retrofits decision-making. Build Environ 2017;115:118–29.
[4] Ma Z, Cooper P, Daly D, Ledo L. Existing building retrofits: methodology and state-of-the-art. Energy Build 2012;55:889–902.
[5] Santamouris M, Balaras C, Dascalaki E, Argiriou A, Gaglia A. Energy conservation and retrofitting potential in hellenic hotels. Energy Build 1996;24(1):65–75.
[6] Shao Y, Geyer P, Lang W. Integrating requirement analysis and multi-objective optimization for office building energy retrofit strategies. Energy Build 2014;82:356–68.
[7] Menassa CC, Baer B. A framework to assess the role of stakeholders in sustainable building retrofit decisions. Sustain Cities Soc 2014;10:207–21.
[8] Ye X, Xia X, Zhang J. Optimal sampling plan for clean development mechanism energy efficiency lighting projects. Appl Energy 2013;112:1006–15.
[9] Tong Z, Chen Y, Malkawi A. Estimating natural ventilation potential for high-rise buildings considering boundary layer meteorology. Appl Energy 2017;193:276–86.
[10] Chua K, Chou S, Yang W, Yan J. Achieving better energy-efficient air conditioning – a review of technologies and strategies. Appl Energy 2013;104:87–104.
[11] Tong Z, Chen Y, Malkawi A. Defining the influence region in neighborhood-scale cfd simulations for natural ventilation design. Appl Energy 2016;182:625–33.
[12] Tong Z, Chen Y, Malkawi A, Liu Z, Freeman RB. Energy saving potential of natural ventilation in China: the impact of ambient air pollution. Appl Energy 2016;179:660–8.
[13] Gowri K. Green building rating systems: an overview. ASHRAE J 2004;46(11):56–8.
[14] Haapio A, Viitaniemi P. A critical review of building environmental assessment tools. Environ Impact Assess Rev 2008;28(7):469–82.
[15] Illankoon ICS, Tam VW, Le KN, Shen L. Key credit criteria among international green building rating tools. J Clean Prod 2017;164:209–20.
[16] Sinou M, Kyvelou S. Present and future of building performance assessment tools. Manage Environ Qual: An Int J 2006;17(5):570–86.
[17] Tronchin L, Tommasino MC, Fabbri K. On the cost-optimal levels of energy-performance requirements for buildings: a case study with economic evaluation in Italy. Int J Sustain Energy Plan Manage 2014;3:49–62.
[18] Energy performance certificates for buildings; 1st ed. South African Bureau of Standards; 2014. SANS 1544.
[19] Tronchin L, Fabbri K. Energy performance certificate of building and confidence interval in assessment: an Italian case study. Energy Policy 2012;48:176–84.
[20] International performance measurement and verification protocol: concepts and options for determining energy and water savings. Tech Rep; Efficiency Valuation Organization (EVO); 1997.
[21] Carstens H, Xia X, Yadavalli S. Measurement uncertainty in energy monitoring: present state of the art. Renew Sustain Energy Rev 2018;82:2791–805.
[22] Qian D, Guo J. Research on the energy-saving and revenue sharing strategy of ESCOs under the uncertainty of the value of Energy Performance Contracting Projects. Energy Policy 2014;73:710–21.
[23] Castro-Lacouture D, Sefair JA, Flórez L, Andrés L. Optimization model for the selection of materials using a LEED-based green building rating system in Colombia. Build Environ 2009;44(6):1162–70.
[24] Michael M, Zhang L, Xia X. An optimal model for a building retrofit with LEED standard as reference protocol. Energy Build 2017;139:22–30.
[25] Asadi E, da Silva MG, Antunes CH, Dias L. Multi-objective optimization for building retrofit strategies: a model and an application. Energy Build 2012;44:81–7.
[26] Güçyeter B, Günaydın HM. Optimization of an envelope retrofit strategy for an existing office building. Energy Build 2012;55:647–59.
[27] Edeisy M, Cecere C. Envelope retrofit in hot arid climates. Proc Environ Sci 2017;38:264–73.
[28] Fan Y, Xia X. A multi-objective optimization model for energy-efficiency building envelope retrofitting plan with rooftop PV system installation and maintenance. Appl Energy 2017;189:327–35.
[29] Kang L, Liu Y. Multi-objective optimization on a heat exchanger network retrofit with a heat pump and analysis of CO_2 emissions control. Appl Energy 2015;154:696–708.
[30] Carstens H, Xia X, Ye X. Improvements to longitudinal clean development mechanism sampling designs for lighting retrofit projects. Appl Energy 2014;126:256–65.
[31] Wang B, Xia X. Optimal maintenance planning for building energy efficiency retrofitting from optimization and control system perspectives. Energy Build 2015;96:299–308.
[32] Fan Y, Xia X. Energy-efficiency building retrofit planning for green building compliance. Build Environ 2018;136:312–21.
[33] Wang B. A control system framework for building energy retrofitting and maintenance planning. PhD thesis; University of Pretoria; 2017.
[34] Ye X, Xia X. Optimal metering plan for measurement and verification on a lighting case study. Energy 2016;95:580–92.
[35] Xia X, Zhang J. Mathematical description for the measurement and verification of energy efficiency improvement. Appl Energy 2013;111:247–56.
[36] Said GAENA, Mahmoud AM, El-Horbaty ESM. A comparative study of meta-heuristic algorithms for solving quadratic assignment problem. Int J Adv Comput Sci Appl 2014;5:1–6.
[37] Deep K, Singh KP, Kansal M, Mohan C. A real coded genetic algorithm for solving integer and mixed integer optimization problems. Appl Math Comput 2009;212(2):505–18.
[38] Juan YK, Gao P, Wang J. A hybrid decision support system for sustainable office building renovation and energy performance improvement. Energy Build 2010;42(3):290–7.
[39] Yokota T, Gen M, Li YX. Genetic algorithm for non-linear mixed integer programming problems and its applications. Comput Indus Eng 1996;30(4):905–17.
[40] Martinot E, Cabraal A, Mathur S. World Bank/GEF solar home system projects: experiences and lessons learned 1993–2000. Renew Sustain Energy Rev 2001;5(1):39–57.
[41] Mondal AH, Klein D. Impacts of solar home systems on social development in rural Bangladesh. Energy Sustain Develop 2011;15(1):17–20.
[42] Kreith F, Goswami DY, editors. Energy management and conservation handbook. CCC Press; 2007.
[43] Marler RT, Arora JS. The weighted sum method for multi-objective optimization: new insights. Struct Multidiscip Optim 2010;41(6):853–62.
[44] Kim I, de Weck O. Adaptive weighted sum method for multiobjective optimization: a new method for Pareto front generation. Struct Multidiscip Optim 2006;31(2):105–16.
[45] Malatji EM, Zhang J, Xia X. A multiple objective optimisation model for building energy efficiency investment decision. Energy Build 2013;61:81–7.
[46] The application of the National Building Regulations Part XA: Energy usage in buildings; 1st ed. South African Bureau of Standards; 2011. SANS 10400-XA.
[47] Zarnd P, Polgr I. A theoretical study on the relative standard deviation of TLD systems. Nucl Instrum Methods Phys Res 1983;205(3):525–9.

Journal of Cleaner Production 201 (2018) 526–541

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

A design approach for multiple drive belt conveyors minimizing life cycle costs

Mukalu Sandro Masaki, Lijun Zhang*, Xiaohua Xia

Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria, 0002, South Africa

ARTICLE INFO

Article history:
Received 2 February 2018
Received in revised form
2 August 2018
Accepted 4 August 2018
Available online 9 August 2018

Keywords:
Belt conveyor
Life cycle cost
Multiple drive
Optimal design
Component sizing
Energy management

ABSTRACT

Energy efficiency of belt conveyors has recently gained in importance worldwide. While significant research efforts were consecrated to the operational aspects, the literature study shows the design optimization problem was scarcely investigated in the past. Among the various type of belt conveyors, the multi-drive technology is now increasingly acknowledged as involving further cost saving opportunities as a result of the possible reduction of the belt weight. In this paper, a multi-drive belt conveyor sizing model that aims to minimize the life cycle cost of the conveyor is presented. The effectiveness of the proposed approach in improving their economic benefits over the single-drive conveyors has been established through extensive simulations on a practical case study. The robustness of the best design solution against the variation in the inflation rate have been also validated.

1. Introduction

Energy shortage is a major concern to many countries around the world. With 83% of electricity generated by coal-fired power plants (Eskom, 2017), statistics from the largest South African energy company, Eskom, indicated that while the entire mining sector consumed 15% of its annual electricity supply, approximately 23% of this consumption was used for material transportation purposes only (Eskom, 2010). Amongst the existing technologies, belt conveyors are largely used for bulk material transfer over short and medium distances because of their low energy consumption per tonne of material transported in comparison to other alternatives (Zhang and Xia, 2010; Tapp, 2000; Darling, 2011). A significant number of belt conveyors are, however, either oversized or inadequately operated, resulting in poor energy efficiency and economic performances (Dalgleish and Grobler, 2003; Zhang and Xia, 2010; Bindzár and Malindžák, 2008). As a result, any improvement in energy efficiency achieved at the design or operation stage of a conveyor can reduce its capital investment and/or operating expenditures.

Generally, energy efficiency activities can be clustered into four categories, namely initiatives focusing on technology, operation, equipment, and performance efficiencies (Xia and Zhang, 2010, 2011, 2016). With indicators such as feasibility, life cycle cost, and return on investment, technology efficiency refers to the efficiencies of energy conversion, processing, transmission, and usage. Equipment efficiency is a measure of the energy output of isolated individual equipment with respect to given technology design specifications. Typical indicators include capacity and maintenance. Operation efficiency focuses on the degree of coordination of the different components of an energy system. Physical coordination, time coordination and human coordination are the indicators usually considered at this level. Performance efficiency is a measure of the global efficiency of the energy system, and is evaluated by external but deterministic indicators such as the production, cost, and environmental footprint. Readers interested in more detailed definitions of technology, operation, equipment, and performance efficiencies are referred to reference (Xia and Zhang, 2010).

Findings from the literature indicate that most of the previous efforts to improve belt conveyor's energy efficiency were done at the equipment level, operation level, and technology level. Equipment efficiency activities include the development of energy-efficient belting materials (Association for Rubber Products Manufacturers Inc, 2011; Dejchanchaiwong et al., 2016), application of energy-efficient motors and variable speed drives (de Almeida et al., 2003), and monitoring and maintenance of

* Corresponding author.
E-mail address: lijun.zhang@up.ac.za (L. Zhang).

https://doi.org/10.1016/j.jclepro.2018.08.040

conveyor components based on risks of failure and expert system tools (Petrović et al., 2014; Mazurkiewicz, 2015). At the operation efficiency level, initiatives include cost-effective load shifting (Middelberg et al., 2009), adaptive belt speed control (Jeftenić et al., 2010; Ristic et al., 2012; He et al., 2017), material scheduling under the time-of-use electricity tariff (Zhang and Xia, 2010; Luo and Shen, 2015; Luo et al., 2015) and critical peak pricing (Mathaba et al., 2012), and the optimal power flow between the electric drives (Windmann et al., 2015). The impacts of the settings of certain design parameters on the future power consumption of a belt conveyor were also investigated (Schützhold et al., 2014). Technology efficiency activities involve strategies for selecting idler rolls (Reicks and Alspaugh, 2008), advanced design of troughing idler sets (Tapp, 2000; Maton, 2003), and modelling of motion resistance components (Lodewijks, 2003; Reicks et al., 2012; Wheeler, 2006).

In practice, for a given transportation task, different designs of belt conveyor will usually result in different cost implications regarding both investment and operation expenditures. Possibilities to minimize the entire life cycle cost through the optimal dimensioning of belt conveyor components were reported in the literature for single drive conveyor belts (Roberts, 1981; Wheeler and Alspaugh, 2008). Belt conveyors using multiple drive design were, however, not covered in the studies previously reported.

Since the early applications in the United States and Germany in the seventies, the distributed drive technology has today reached maturity. Multiple drive conveyor systems are already widely used in underground coal mines and are increasingly being considered in the mining plan for future developments (Alspaugh, 2003). Fig. 1 illustrates the evolution of the belt tension profile before and after one, and subsequently, two drive stations are inserted in the upper stretch of a single drive conveyor. The decline of the maximum belt tension achieved by increasing the number of intermediate drive stations may allow conveyor designers to shift towards less resistant belt products and thereby reducing the weight of the belt and the supporting structure.

The resulting economic benefit of this practice is, however, subject to the interactions between the various design parameters relevant to the capital and operational costs of the belt conveyor components. In particular, to achieve good performance for a specified transportation task, the number of drive stations should be set taking into account the influence of their respective sizes and

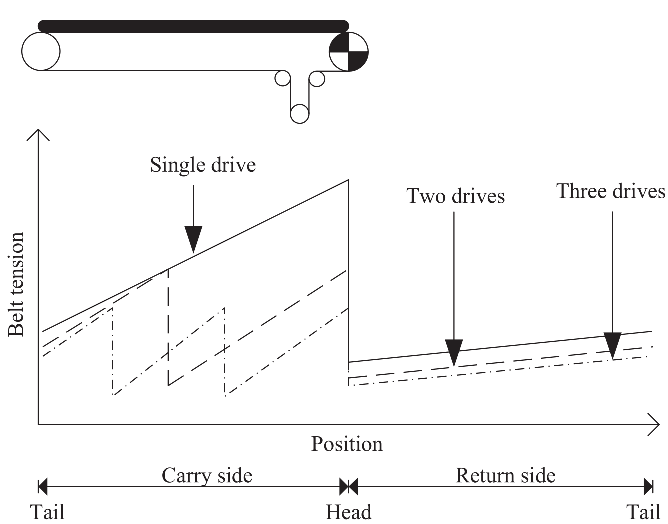


Fig. 1. Belt tension profile v.s. number of driving stations (adapted from Alspaugh (2003)).

distribution along the belt path, and the belt speed, among others. To the best of the authors' knowledge, no previous study dedicated to the economic design of multiple drive belt conveyors was reported in the literature. The rest of this paper presents a contribution on the cost-effective component sizing model for multiple drive belt conveyors. The proposed approach intends to optimally determine the most important design parameters in order to minimize the life cycle cost of such a conveyor system while satisfying various design and operational constraints. This introduces a method for economic design of multi-drive conveyors for plant owners to make the best use of capital investments and to reduce operating cost of the belt conveyors. In particular, solution of the presented model yields the optimal sizes of components of a multi-drive conveyor that will result in the minimum capital and operating costs over the entire lifespan of the conveyor subject to design constraints. Therefore, the model developed will be a handy and powerful tool to help plant owners to design the most cost-effective conveyor solutions to their needs. It can also be used as a decision support tool for plant owners when comparing different investment options for material transportation using belt conveyors.

The rest of this paper is organized as follows. A brief introduction to multi-drive belt conveyor is given in Section 2. The cost-effective design problem for multi-drive conveyors is formulated in Section 3, followed by the detailed mathematical modelling of the belt conveyor in Section 4. Section 5 presents a case study to demonstrate the effectiveness of the optimal component sizing model developed. The robustness of the cost-effective conveyor designs against possible fluctuations in the inflation rate during the project is also discussed. Section 6 concludes this study.

Nomenclature	
α_i	Wrap angle of drive pulleys in the i-th drive station [°]
β	Equivalent angle of slope of the material [°]
γ_{belt}	Specific mass of the belt [kg/m^2]
δ_j	Inclination angle of the belt section j [°]
$\eta_{gear,i}$	Efficiency of gear reducers in the i-th drive station
$\eta_{mot,i}$	Efficiency of motors in the i-th drive station
λ	Troughing angle [°]
μ	Friction factor between the drive pulley and the conveyor belt
μ_1	Friction factor between belt and material conveyed
μ_2	Friction factor between the lateral chutes and the material transferred
μ_3	Friction factor between the belt cleaning device and the belt
ξ_o	Number of idler rolls per set on the carry side
ξ_u	Number of idler rolls per set on the return side
ρ	Material density [kg/m^3]
a	Constant factor for the calculation of clear width of lateral chutes
A_{Gr}	Effective contact area between belt cleaning device and belt [m^2]
A_{th}	Theoretical cross section of fill [m^2]
$A_{conveyor}$	Annual equivalent cost of the belt conveyor [USD/year]
A_{belt}	Annual equivalent cost of the belt [USD/year]
$A_{carryidler}$	Annual equivalent cost of all the carry idler rolls [USD/year]
A_{energy}	Annual equivalent energy cost [USD/year]
A_{eq}	Annual equivalent cost of an equipment [USD/year]
$A_{gear,i}$	Annual equivalent cost of each gear reducer in the i-th drive station [USD/year]
$A_{motor,i}$	Annual equivalent cost of each motor in the i-th drive station [USD/year]
j	Belt section index
$A_{returnidler}$	Annual equivalent cost of all the return idler rolls [USD/year]
B	Belt width [m]
$\mathfrak{B}$	Set of the recommended width of belt
b	Usable belt width [m]
B_f	Dynamic load factor related to bearing life
b_{Sch}	Clear width of lateral chutes [m]
$C_{eq,0}$	First cost of the item of an equipment purchased at the year zero [USD]

(continued on next page)

(*continued*)

C_f	Belt flap factor
C_{Rank}	Rakine coefficient
C_{Schb}	Constant factor for additional resistance between material transferred and lateral chutes
c_{Tr}	Drive pulleys constant coefficient related to the type longitudinal tension members of the belt
$C_{w,i}$	Combined warp factor of the drive pulleys in the *i*-th drive station
$c_1, \ldots, c_{22}$	Initial cost coefficients
$\mathfrak{D}$	Set of the recommended diameters of idler roll
$\mathfrak{d}$	Set of the recommended shaft diameters of idler roll
$\mathfrak{D}_{tr}$	Set of the recommended diameters of drive pulley
d_{Gk}	Thickness of the longitudinal tension members of the belt [m]
D_j	Diameter of idler rolls in the belt section *j* [m]
d_j	Shaft diameter of idler rolls in the belt section *j* [m]
D_o	Diameter of idler rolls in the upper stretch [m]
d_o	Shaft diameter of idler rolls in the upper stretch [m]
D_u	Diameter of idler rolls in the lower stretch [m]
d_u	Shaft diameter of idler rolls in the lower stretch [m]
e_o	Unit cost of energy at the year zero of the project [USD/kWh]
f_j	Hypothetical friction factor in the belt section *j*
$F_{Auf\,j}$	Resistance due to the acceleration of the material in the loading zone of the belt section *j* [N]
F_{min}	Minimum belt tension in steady-state operating conditions [N]
$F_{G,j}$	Gradient resistance in the belt section *j* [N]
$F_{Gr,j}$	frictional resistance due to belt cleaning devices situated in the belt section *j* [N]
$F_{H,j}$	Primary resistance in the belt section *j* [N]
F_0	Belt tension at each side of the tail pulley [N]
$F_{N,j}$	Secondary resistance in the belt section *j* [N]
$F_{S,j}$	Special resistance in the belt section *j* [N]
$F_{s,o}$	Static load on the central carry idler roll in a three-idler troughing configuration [N]
$F_{s,u}$	Static load on a flat return idler in the lower stretch [N]
$F_{Schb,j}$	frictional resistance between belt and lateral chutes in the acceleration zone of the belt section *j* [N]
$F_{T1,i}$	Tight side tension of the first drive pulley in the *i*-th drive station [N]
$F_{T2,i}$	Slack side tension of the second drive pulley in the *i*-th drive station [N]
F_{TU}	Belt tension on both sides of the take-up device [N]
$F_{W,j}$	Total resistance to movement in the belt section *j* [N]
F_0	Belt tension at the conveyor tail [N]
g	Gravitational acceleration [m/s^2]
H	Lifting height [m]
h_{rel}	Maximum belt sag related to spacing between idler rolls
i	Integer index
i_d	Interest rate on debt
i_e	After-tax return required on equity funds with zero inflation rate
$i_{f,j}$	Inflation modified rate of return of the year *j* of the project
i_f^0	Time value of money with all cash flows converted from inflated value to constant year zero value
j	Integer index
K	Total length of the belt along the conveyor path [m]
k_b	Constant factor for calculation of the total length of the acceleration path
k_N	Nominal breaking strength of the belt related to belt width [N/m]
k_{eq}	Equivalent annual cost coefficient of an equipment
$k_{t,rel}$	Relative reference endurance strength of the belt
$k_1, \ldots, k_6$	Equivalent annual cost coefficients
L	Horizontal transport distance [m]
l_b	Total length of the acceleration path [m]
L_f	Dynamic load factor related to lump size of the material transferred
l_j	Length of the belt section *j* [m]
$L_{o,j}$	Length of the belt section *j* in the upper stretch [m]
$l_{M,o}$	Length of the shell of a carry idler roll [m]
l_o	Spacing between idler rolls in the upper stretch [m]
$L_{u,j}$	Length of the belt section *j* in the lower stretch [m]
l_u	Spacing between idler rolls in the lower stretch [m]
M	Expected lifetime of each item of an equipment [year]
m'_G	Linear mass of the belt [kg/m]
$m'_{L,j}$	Linear mass of the transferred material in the belt section *j* [kg/m]
$m_{R,j}$	Mass of the rotating parts of each idler situated in the belt section *j* [kg]

(*continued*)

$m'_{R,j}$	Linear mass of the rotating parts of idlers per running meter in the belt section *j* [kg/m]
m_1	Belt weight model coefficient [kg/m^2]
m_2	Belt weight model coefficient [s^2/m^2]
m_3	Steelcord diameter model coefficient [m]
m_4	Steelcord diameter model coefficient [m]
m_5	Steelcord diameter model coefficient [m]
N	Number of intermediate drive stations
N_o	Number of belt sections in the upper stretch
N_u	Number of belt sections in the lower stretch
n_1	Dynamic speed load factor model coefficient
n_2	Dynamic speed load factor model coefficient [s/m]
o	Upper stretch
P_i	Rated power of motors in the *i*-th drive station [kW]
PED_{eq}	Present equivalent of depreciation of the items of an equipment [USD]
PEF_{eq}	Present equivalent of all the first cost of an equipment [USD]
PEV_{eq}	Present equivalent of all the salvage value of an equipment [USD]
p_{Gr}	Pressure between the belt cleaning device and the belt [N/m^2]
Q	Material flow rate [kg/s]
q_f	Remaining proportion of the initial value of an equipment at the end of the expected lifetime
q_i	Remaining proportion of the initial value of the item *i* of an equipment at the end of its actual lifetime
R	Number of items of an equipment to be purchased over the project lifetime
r_{avg}	Average general inflation rate over the project duration
r_d	Proportion of debt capital
$r_{e,j}$	Annual escalation rate of energy during the year *j*
$r_{eq,j}$	Annual cost escalation rate of an equipment during the year *j*
r_j	General inflation rate during the year *j*
S_f	Dynamic load factor related to belt speed
S_0	Belt safety factor related to the splicing conditions
S_1	Belt safety factor related to the expected lifetime and the operation conditions
t	Income tax rate
t_a	Operating hours per annum [h]
T_i	Rated torque of gear reducers in the *i*-th drive station [kNm]
u	Lower stretch
v	Conveyor speed [m/s]
$v_{0,j}$	Initial speed of the material in the direction of belt travel in the belt section *j* [m/s]
w	Width of the contact area between belt and belt cleaning device [m]
X_i	Purchase year of the item *i* of an equipment
Y_i	Year of decommissioning of the item *i* of an equipment
y_1, y_2	Belt length model coefficients [m]
Z	Project lifetime [year]
$z_1, \ldots, z_4$	Idler roll mass model coefficients

2. Description of multi-drive belt conveyors

Fig. 2 illustrates a typical modern uphill multi-drive belt conveyor that aims to transfer a bulk material of density ρ with a flow rate Q over a distance L with a lift height H. It consists basically of an upper stretch and a lower stretch subsequently identified by the subscripts o and u, respectively. The upper stretch carries the bulk material from the loading point situated at the tail pulley and along the conveyor path to the unloading points positioned along the conveyor path and at the head pulley. On the other hand, the lower stretch consists of the empty belt that circulates from the head pulley to the tail pulley. Apart from these two pulleys, one or several pairs of drive pulleys mounted in tandem are positioned along the upper stretch of the conveyor. One pair of drive pulleys and one pair of idler pulleys are positioned in the lower stretch as indicated in Fig. 2. Each drive pulley is connected to a motor-gear reducer system mounted at its shaft. The rest of pulleys rotate freely and are driven by the belt. The arc of contact between the belt and a drive pulley is referred to as the wrap angle and is denoted by α.

The drive station is the unit comprising a pair of drive pulleys mounted in tandem together with their associated motor-gear reducer systems. In particular, the drive stations situated in the upper stretch are identified as intermediate drive stations. Within a drive station, the drive pulley 1 and drive pulley 2 refer to, respectively, the first and second pulleys when following the belt travel direction in Fig. 2. For illustration purposes, Fig. 2 shows a conveyor system comprising 4 drive stations including 3 intermediate drive stations. A general design will comprise $N+1$ ($N=1,2,\ldots$) drive stations with N positioned as intermediate drive stations. These drive stations will then be numbered from 1 to $N+1$ starting at the drive station near the tail pulley and ending at the drive station located in the lower stretch.

A belt section refers any portion of belt nestled between any two different pulleys. Like in standard DIN 22101 (DIN 22101, 2011), within each stretch, the edges of each belt section is identified by means of a unique index specified in ascending order starting at the tail pulley identified by default as edge 0. A belt section is subsequently designated by the greatest index between its two edges. For example, the belt conveyor in Fig. 2 has 7 belt sections in the upper stretch and 5 belt sections in the lower stretch. The belt sections 3 in the upper and lower stretches of the conveyor are also pointed out ig. 2. In more general case, the number N_o of belt sections in the upper stretch will depend on N, while the number N_u of belt sections in the lower stretch will be determined by the number of pairs of idler pulleys installed. Although the number of idler pulleys can vary from one conveyor system to another, this study only considers multi-drive belt conveyors with a single pair of idler pulleys. N_u will be therefore equal to 5 in the following. Thereafter, $L_{o,i}$ and $L_{u,j}$ will denote the lengths of, respectively, the belt section i in the upper stretch and the belt section j in the lower stretch.

For the purpose of supporting the belt sections that extend over long distances, carrying and return idler rolls are mounted underneath the belt in these belt sections as described in Fig. 2. Fig. 3 (a) also shows the carrying idler rolls also supporting the bulk material in transit along the belt conveyor in the case of a three-idler roll troughing configuration. In this figure, B denotes the belt width, b denotes the usable belt width, β denotes the equivalent angle of slope of the material, λ denotes the troughing angle and $l_{M,o}$ denotes the length of the shell of a carry idler roll.

The resultant longitudinal force measured at a specific point along the conveyor path is referred to as the belt tension at this point and is noted by F. Fig. 3(b) shows the belt tension components around a drive pulley. The belt tension at the belt run-on point on the drive pulley is referred to as the tight side tension, and is noted F_{T1}. On the other side, the belt tension at the belt run-off point on the drive pulley is referred to as the slack side tension, and is noted F_{T2}. Moreover, F_{Tr} denotes the peripheral force applied by the drive pulley on the belt. Analogously, $F_{T1,i}$, $F_{T2,i}$ and $F_{Tr,i}$ will denote, respectively, the tight side tension, the slack side tension and the peripheral force of the i-th drive station. The belt strength is specified by the nominal breaking strength of the belt related to belt width, k_N, which corresponds to the minimum rupture force of the belt per unit of belt width. Lastly, long belt conveyors usually require to be fitted with a tensioning equipment also referred to as take-up device (not shown in Fig. 2) so as to prevent belt slipping on drive pulleys. The belt tension on each side of the take-up device is noted F_{TU}.

3. Problem formulation

A given conveying operation can be described by L, H, Q, ρ and β. For such a material transfer task, a large variety of multi-drive belt conveyors can be envisaged, designs of which will generally lead to different cost implications over the project lifetime. The goal is therefore to identify the design solution that results in the lowest life cycle cost. To facilitate the comparison of belt conveyor designs, the equivalent annual cost of a belt conveyor $A_{conveyor}$ is adopted as the performance indicator instead of directly inspecting the life cycle costs.

Therefore for a given N, the general formulation of the optimization problem that allows to determine the design solution with the minimum $A_{conveyor}$ is stated as

$$\begin{aligned} \min_X \quad & A_{conveyor} \\ \text{s.t.} \quad & G(X)=0, \\ & H(X)\geq 0, \end{aligned}$$

where X denotes the set of design parameters and G and H denote, respectively, the functions of equality and inequality constraints relating to the belt dynamics and design conditions as detailed in subsection 4.3. Although N can be treated as a decision variable, it adds much complexity of the model. The direct comparison of the minimum $A_{conveyor}$ obtained for different N will therefore lead to the most cost-effective design solution in terms of N and X.

By keeping the two driving subsystems of each i-th drive station ($i=1,\ldots,N+1$) identical in all respects, the set X considered in this study includes: the rated power of each motor in the i-th drive station P_i, the rated torque of each gear reducer in the i-th drive station T_i; the diameter of each drive pulley in the i-th drive station $D_{tr,i}$; the wrap angle of each drive pulley in the i-th drive station α_i; $L_{o,j}$ of the belt sections j ($j=1,3,\ldots,N_o$) not nestled between drive pulleys, the belt width, B, the belt speed, v, the spacing between idler rolls in the upper stretch, l_o, the spacing between idler rolls in the lower stretch, l_u, the diameter of idler rolls in the upper stretch, D_o, the diameter of idler rolls in the lower stretch, D_u, the shaft diameter of idler rolls in the upper stretch, d_o, the shaft diameter of idler rolls in the lower stretch, d_u, k_N and F_{TU}.

Regarding the calculation of $A_{conveyor}$, the various costs incurred throughout the belt conveyor lifetime can be grouped into capital costs and operating costs. Since each of these categories comprises several expense items incurred at different points of time during

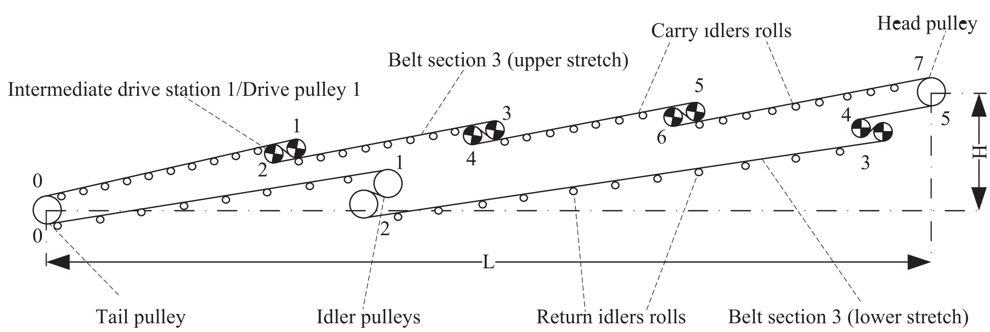
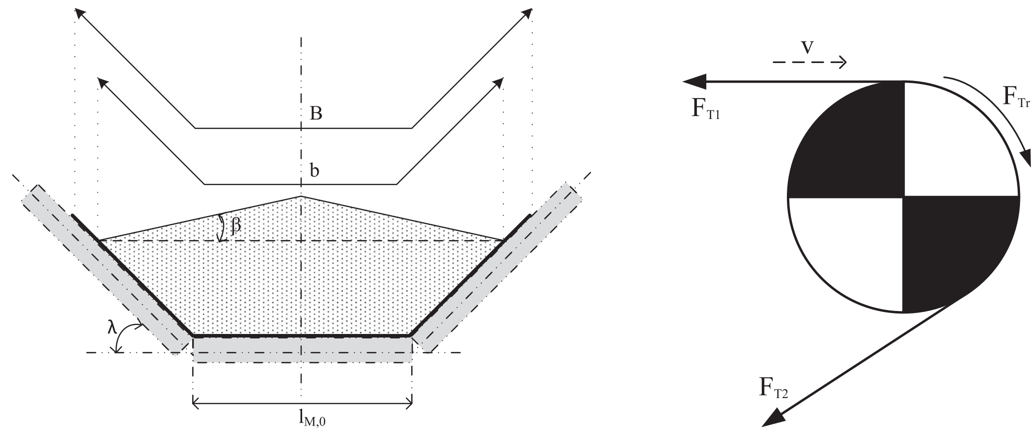


Fig. 2. Multiple drive belt conveyor layout (adapted from Alspaugh (2003)).



(a) Theoretical cross section of fill (b) Belt tensions around a drive pulley

Fig. 3. Cross section of fill and mechanic around the drive pulley.

the project life, $A_{conveyor}$ will therefore consist of the sum of the equivalent annual costs of each cost item involved. The next subsections give a brief discussion of the cost items relating to the type of belt conveyor investigated in this study.

3.1. Operating costs

The operating costs of a belt conveyor include energy cost, maintenance cost and labor cost. While the relationships between the design parameters and the expenditures for maintenance and labor were not investigated in the past, it can be expected that the number and size of drive stations, the conveyor speed and belt size will influence the maintenance cost. On the other hand, the wage and number of workers in a mine are usually driven by the plant size, the production, and the local legislation (Darling, 2011; Roberts, 1981). Accordingly, no influence on the labor is expected from the design parameters selected in this study. As a result of the difficulties experienced in accessing modelling data for the maintenance part, only with the energy cost is subsequently considered.

The equivalent annual energy cost A_{energy} of a multi-drive belt conveyor with N intermediate drive stations, with each drive station consisted of two drive pulleys and driving systems, is given by:

$$A_{energy} = k_1 e_o t_a \sum_{i=1}^{N+1} 2P_i \Big/ \eta_{mot,i}, \tag{1}$$

where k_1 denotes the equivalent annual energy cost coefficient, e_o denotes the unit cost of energy at the year zero of the project, t_a denotes the operating hours per annum, and $\eta_{mot,i}$ denotes the efficiency of motors of the i-th drive station. As detailed in Appendix A.1, the calculation of k_1 takes into account of several factors, including the general inflation rate, the annual escalation rate of energy during the project and the tax rate.

3.2. Capital costs

The capital costs of the conveyor components, including the belting material, the electric motors, the gear reducers, the carry idlers and the return idlers, are considered at this stage. Although, in practice, the cost for the supporting structure is significant among the conveyor components, it is not investigated because of its high dependency on the geographic characteristics of the mine.

The initial cost function of belt conveyor components considered in this study is derived from Roberts (1981) and Wheeler and Alspaugh (2008). Further analysis on idler roller purchase costs indicated that the shaft diameter affects the product price in addition to the idler roller's diameter and length. The cost functions of these items were revised accordingly. As a result, the annual equivalent cost of the conveyor components are given as follows

$$A_{belt} = k_2 BK\left(c_1 + c_2 k_N^{c_3}\right), \tag{2}$$

$$A_{motor,i} = k_3\left(c_4 + c_5 P_i^{c_6}\right), \tag{3}$$

$$A_{gear,i} = k_4\left(c_7 + c_8 T_i^{c_9}\right), \tag{4}$$

$$A_{carryidler} = \xi_o k_5 \sum_{j=1}^{N_o} \frac{L_{o,j}}{l_o}\left(c_{10} + c_{11} d_o^{c_{12}} + c_{13} D_o^{c_{14}} + c_{15} B^{c_{16}}\right), \tag{5}$$

$$A_{returnidler} = \xi_u k_6 \sum_{j=1}^{N_u} \frac{L_{u,j}}{l_u}\left(c_{17} + c_{18} d_u^{c_{19}} + c_{20} D_u^{c_{21}} + c_{22} B^{c_{23}}\right). \tag{6}$$

where K denotes the total length of the belt along the conveyor path, k_2 to k_6 denote the equivalent annual cost coefficients of belt conveyor components c_1 to c_{23} denote the initial cost coefficients, ξ_o denotes the number of carry idler rolls per set (e.g. $\xi_o = 3$ a three-idler troughing configuration), and ξ_u denotes the number of return idler rolls per set (e.g. $\xi_u = 1$ in a flat return configuration).

The values of c_1 to c_{23} are determined based on suppliers' price data. Besides the economic parameters relevant to k_1, the calculation of the equivalent annual cost coefficient of a given belt conveyor component implies taking into account also the annual escalation rate of the initial costs of this equipment, the first costs of the first item and its replacements purchased during the project life, their related expected lifetimes, their respective annual depreciation rates and their respective salvage values. k_2 to k_6 are obtained using the procedure disclosed in Appendix A.2. For its part, K in equation (2) can be approximated by

$$K = 2L/\cos\delta + y_1 N + y_2,$$

where δ denotes the inclination angle of the conveyor system, y_1 and

y_2 denote constant coefficients that account for, respectively, the wrapping of the belt around the drive pulleys and a reserve factor.

The equivalent annual cost of a multiple drive belt conveyor with N intermediate drive stations is therefore given by

$$A_{conveyor} = A_{energy} + A_{belt} + 2\sum_{i=1}^{N+1} A_{motor,i} + 2\sum_{i=1}^{N+1} A_{gear,i} + A_{carryidler} + A_{returnidler}. \quad (7)$$

The cost of the conveyor, $A_{conveyor}$, varies as a function of the design parameters of the components that form parts of the system. The system design must also accommodate any technical requirements relevant to the technology and the intended application. A multi-drive belt conveyor sizing model that aims to minimize the life cycle of the conveyor is developed in Section 4.

4. Mathematical model of the multi-drive belt conveyors

The force analysis, the power balance requirement, and the operational constraints that should be satisfied to ensure the proper and safe operation of the multi-drive belt conveyor are presented in this section.

4.1. Motion resistance modelling

According to DIN 22101 standard (DIN 22101, 2011), the overall resistance to the belt movement $F_{W,j}$ that occurs within a belt section j consists of the primary resistance, the secondary resistance, the gradient resistance and the special resistance.

The primary resistance $F_{H,j}$ combines together the running resistance forces caused by indentation of the belt cover on the idler rolls, the flexure of the belt between the idler rolls, and the rotational resistance of idler rolls. The resulting opposition force to the belt movement in the section j is approximated by:

$$F_{H,j} = l_j f_j \left[m'_{R,j} + \left(m'_G + m'_{L,j}\right)\cos\delta_j\right]g, \quad (8)$$

where l_j denotes the length of the belt section, f_j denotes the hypothetical friction factor, g denotes the gravitational acceleration, $m'_{R,j}$ denotes the total mass of the rotating parts of idler rolls per running meter, m'_G denotes the linear mass of the belt, $m'_{L,j}$ denotes the linear mass of the material transferred, and δ_j denotes the belt section inclination angle. Since the system is unloaded in the return side, the resistance factor due to the conveyed material $m'_{L,j}$ will not apply to the belt sections situated in the lower stretch. The total mass $m_{R,j}$ of the rotating parts of a idler roll is approximated by:

$$m_{R,j} = z_1 D_j^{z_2} B^{z_3} + z_4 d_j,$$

where z_1 to z_5 are the model coefficients, D_j denotes the shell diameter of idler rolls in the belt section j, and d_j denotes the shaft diameter of idler rolls in the same belt section.

The secondary resistance $F_{N,j}$ in a belt section j entails the frictional resistances $F_{Auf,j}$, $F_{Schb,j}$, and $F_{Gr,j}$. These resistances are modeled as follows:

$$F_{Auf,j} = \frac{Q}{\rho}\left(v - v_{0,j}\right), \quad (9)$$

$$F_{Schb,j} = C_{Schb} C_{Rank}\left[\frac{2Q}{\left(v + v_{0,j}\right)\rho^2} - \left(b_{Sch}^2 - l_{M,o}^2\right)\frac{\tan\lambda}{4}\right]^2 \frac{\rho g l_b \mu_2}{b_{Sch}^2}, \quad (10)$$

$$F_{Gr,j} = \mu_3 p_{Gr} A_{Gr}. \quad (11)$$

In equations (9)–(11), $v_{0,j}$ denotes the initial speed of the material in the direction of the belt travel, C_{Schb} denotes a constant factor for the additional resistance between material loaded and lateral chutes, C_{Rank} denotes Rankine coefficient, b_{Sch} denotes the clear width of lateral chutes, l_b denotes the total length of the acceleration path, λ denotes the troughing angle, μ_2 denotes the friction factor between lateral chutes and material transferred, μ_3 denotes the friction factor between belt cleaning device and belt, p_{Gr} denotes the pressure between belt cleaning device and belt, and A_{Gr} denotes the effective contact area between belt cleaning device and belt. A_{Gr}, b_{Sch} and l_b are determined as follows

$$A_{Gr} = wB,$$

$$b_{Sch} = a l_{M,o},$$

$$l_b = \frac{k_b\left(v^2 - v_{0,j}^2\right)}{2g\mu_1},$$

where w denotes the width of the contact area between belt and belt cleaning device, μ_1 denotes the friction factor between belt and material conveyed and a and k_b are constant coefficients.

In addition to the normal distribution of the secondary resistance components as in the single drive conveyor systems, the inertia resistance and the frictional resistance between lateral chutes and the belt given, respectively, by (9) and (10), also occur in each belt section situated in the downstream of an intermediate drive station.

The gradient resistance $F_{G,j}$ caused by the lifting of the belt and the material in a belt section j is given by

$$F_{G,j} = l_j \sin\delta_j\left(m'_G + m'_{L,j}\right)g, \quad (12)$$

in which $\delta_j > 1$ for uphill belt travel and $\delta_j < 1$ for downhill belt travel.

The special resistance component $F_{S,j}$ concerns the remaining resistances that apply only to particular conveyor designs. It includes the camber resistance, the resistance due to any lateral transfer equipment positioned along the conveyor path and the frictional resistance between lateral chutes and transferred material beyond the loading zones.

The overall resistance to the belt movement in a belt section j is therefore given by

$$F_{W,x,j} = F_{H,x,j} + F_{N,x,j} + F_{G,x,j} + F_{S,x,j},$$

where the subscript x is replaced by o for the upper stretch or by u for the lower stretch. The overall resistance to the movement F_W can be therefore expressed as follows:

$$F_W = \sum_{j=1}^{N_o} F_{W,o,j} + \sum_{j=1}^{N_u} F_{W,u,j}.$$

4.2. Power balance of the belt conveyor

The following condition ensures the power balance over the entire multi-drive conveyor system:

$$2\sum_{i=1}^{N+1} P_i \eta_{gear,i} - vF_W = 0, \tag{13}$$

where $\eta_{gear,i}$ denotes the efficiency of the gear reducers in the i-th drive station.

The power balance within the driving subsystems of each drive station is guaranteed by verifying:

$$\frac{2T_i v}{D_{tr,i}} = \eta_{gear,i} P_i, \quad i = 1,\dots,N+1. \tag{14}$$

Going a step further from equation (13), the force balance at the drive stations can be determined. First, the minimum belt tension need to be determined. For $N_u = 3$ and depending on the magnitude of the gradient resistances in the belt sections 1 and 3 in the lower stretch, the minimum belt tension F_{min} will normally occur either at the tail pulley or at the slack side the drive station situated in the return side. In case the take-up device is fitted at the point of minimum belt tension, the belt tension F_0 at the tail pulley is given by

$$F_0 = \begin{cases} F_{TU}, & \text{if } F_{min} = F_0, \\ F_{TU} + F_{W,o,1} + F_{W,o,3}, & \text{if } F_{min} = F_{T2,N+1}. \end{cases}$$

$$A_{th} = \left[l_{M,o} + (b - l_{M,o})\cos\lambda\right]^2 \frac{\tan\beta}{4} + \left(l_{M,o} + \frac{b - l_{M,o}}{2}\cos\lambda\right)\frac{b - l_{M,o}}{2}\sin\lambda.$$

Following the direction of the belt movement, the tight side tension of a i-th drive station is calculated by subtracting the total driving force due to all the drive stations located between the tail pulley and this drive station from the sum of F_0 and the total resistance of all the belt sections situated between the tail pulley and the tight side of the drive station concerned. The tight side tension of each of the $N+1$ drive stations of a multi-drive conveyor is therefore obtained by

$$F_{T1,i} = \begin{cases} F_0 + F_{W,o,i}, & \text{if } i = 1, \\ F_0 + \sum_{k=1}^{2i-1} F_{W,o,k} - 2\sum_{k=1}^{i-1} P_k \eta_k / v, & \text{if } 2 \le i \le N, \\ F_0 + \sum_{k=1}^{2i-1} F_{W,o,k} + F_{W,u,N_u} - 2\sum_{k=1}^{i-1} P_k \eta_k / v, & \text{if } i = N+1. \end{cases}$$

As shown in Fig. 1, the maximum belt tension under steady operating conditions is reduced in multi-drive belt conveyors through the equalization of the tight side belt tensions of all the drive stations installed in the conveyor system (Alspaugh, 2003; Nuttall, 2007). Hence, the following condition must be satisfied:

$$F_{T1,i} = F_{T1,1}, \quad i = 2,\dots,N+1. \tag{15}$$

Within a drive station, the slack side tension is equal to the difference between the tight side tension and the total tensile force transmitted by its gear reducers:

$$F_{T2,i} = F_{T1,i} - 2P_i \eta_{gear,i} \Big/ v, \quad i = 1,\dots,N+1.$$

To guarantee the effective transmission of the driving forces from the drive pulleys to the belt, the slack side tension of each drive station should verify:

$$F_{T2,i} - 2C_{w,i} P_i \eta_{gear,i} \Big/ v \ge 0, \quad i = 1,\dots,N+1, \tag{16}$$

where the combined wrap factor $C_{w,i}$ of a i-th drive station can be obtained by (Conveyor Equipment Manufacturers Assoc, 1997)

$$C_{w,i} = \frac{1}{e^{2\mu\alpha_i} - 1}, \quad i = 1,\dots,N+1,$$

by assuming that the friction factor μ between drive pulley and conveyor belt is constant for the entire conveyor.

4.3. Design constraints

4.3.1. Material transportation requirements

The conveyor must transport a required flow rate of material over a specific distance. The following equation ensures the required material flow

$$Q = \rho A_{th} v, \tag{17}$$

where the theoretical cross section of fill A_{th} for a three-idler troughing configuration shown in Fig. 3(a) is given by

The following condition needs to be satisfied to ensure the desired the transportation distance L of the material

$$\sum_{k=1,3,\dots}^{N_o} L_{o,k} - \sum_{i=1}^{N} D_{tr,i} = L/\cos\delta, \tag{18}$$

4.3.2. Safety and endurance requirements

The design constrains relating to the operation safety and endurance of the conveyor are discussed here. To ensure operational safety, the following conditions apply to the belt tension at the tail pulley and the slack side tension of intermediate drive stations in order to limit the belt sag in the upper stretch below a specified value h_{rel} (DIN 22101, 2011):

$$F_0 \ge \frac{g(\rho A_{th} + B\gamma_{belt})l_o}{8h_{rel}}, \tag{19}$$

$$F_0 + \sum_{j=1}^{k} F_{W,o,j} - 2\sum_{r=1}^{m} \frac{\eta_r P_r}{v} \ge \frac{g(\rho A_{th} + B\gamma_{belt})l_o}{8h_{rel}}, \tag{20}$$

for $k = 2, 4, \ldots, N_o - 1$, $m = \left\lceil \frac{k-1}{2} \right\rceil$, and where γ_{belt} denotes the specific mass of the belt. The function $\lceil . \bullet \rceil$ is the ceiling function which rounds a real number upwards to the nearest integer. Similarly, the belt sag in the lower stretch is maintained below the same value by applying the following condition at the spot of the minimum belt tension:

$$F_{TU} \geq \frac{gB\gamma_{belt} l_u}{8h_{rel}}. \tag{21}$$

Belt manufacturers' product datasheets provide the following type of relation between γ_{belt} and k_N

$$\gamma_{belt} = m_1 + m_2 k_N,$$

where m_1 and m_2 are the model coefficients.

The nominal breaking strength of the belt related to belt width and the maximum belt tension, which coincides with the tight side tension of the drive stations, should satisfy

$$\frac{k_{t,rel} k_N}{S_0 S_1} \geq \frac{F_{T1,1}}{B}. \tag{22}$$

where $k_{t,rel}$ denotes the relative reference endurance strength of the belt, S_0 denotes the belt safety factor related to the splicing conditions, and S_1 denotes the belt safety factors related to the expected lifetime, the operational conditions and the dynamics of the conveyor.

The following condition ensures that the strength of the longitudinal tensile members in the belt core endures over the expected lifetime of the belt

$$D_{tr,i} \geq c_{Tr} d_{Gk}, \quad i = 1, \ldots, N + 1, \tag{23}$$

where c_{Tr} denotes a constant factor that depends on the type of the longitudinal tensile members and d_{Gk} denotes their thickness.

In case of steelcord belts, the manufacturers' product datasheets provide the following relation between d_{Gk} and k_N:

$$d_{Gk} = m_3 + m_4 k_N^{m_5},$$

where m_3, m_4 and m_5 are the model coefficients.

As per the SANS 1313 standard, the admissible load-carrying capacities $F_{max,o}$ and $F_{max,u}$ of, respectively, the carry and return idler rolls are specified in relation to the diameters of their shafts and the belt width (Frittella and Curry, 2009). To further prevent risks of premature failure, the idler rolls in the upper stretch are also subjected to the following condition (Handling Sandvik Materials, 2013):

$$S_f B_f L_f F_{s,o} \leq F_{max,o}, \tag{24}$$

where S_f, B_f , and L_f denote the dynamic load factors related to, respectively, the belt speed, the bearing life, and the lump size of the material transported. $F_{s,o}$ denotes the static load on the central idler roll, which is determined by

$$F_{s,o} = \left\{ \gamma_{belt} l_{M,o} + \frac{1}{2} \rho l_{M,o} \left[\frac{1}{2} l_{M,o} \tan\beta + (b - l_{M,o})(\sin\lambda + \cos\lambda \tan\beta) \right] \right\} g l_o.$$

The variation of S_f with respect to v is described by (Handling Sandvik Materials, 2013):

$$S_f = n_1 + n_2 v,$$

where n_1 and n_2 are the model coefficients.

Similarly, the following condition applies to idler rolls situated in the lower stretch

$$S_f B_f C_f F_{s,u} \leq F_{max,u}, \tag{25}$$

where C_f denotes the belt flap factor and the static load $F_{s,u}$ on a flat return idler is given by

$$F_{s,u} = \gamma_{belt} l_{M,u} g l_u.$$

While complying again with the SANS 1313 standard, the rotation speed of each idler roll should not exceed the limit of 750 rpm (Frittella and Curry, 2009). This composes a constraint on the following relationship between the conveyor speed v and the idler roll diameters D_o and D_u

$$\frac{60v}{\pi D_o} \leq 750, \tag{26}$$

$$\frac{60v}{\pi D_u} \leq 750. \tag{27}$$

4.3.3. Standardization requirements

In case the use of identical equipment and settings is required for supply chain and operational motivations, the following constraints will apply along with the previous design conditions:

$$P_i = P_1, \tag{28}$$

$$T_i = T_1, \tag{29}$$

$$D_{tr,i} = D_{tr,1}, \tag{30}$$

$$\alpha_i = \alpha_1, \tag{31}$$

for $i = 2, \cdots, N + 1$.

4.3.4. Boundary limits

Lastly, the design parameters are subject to the following boundary limits:

$$0 \leq P_i \leq P_{max} \tag{32}$$

$$0 \leq T_i \leq T_{max} \tag{33}$$

$$D_{tr,i} \in \mathfrak{D}_{tr} \tag{34}$$

$$\alpha_{min} \leq \alpha_i \leq \alpha_{max} \tag{35}$$

$$L_{min} \le L_{o,j} \le L_{max} \tag{36}$$

$$B \in \mathfrak{B} \tag{37}$$

$$0 \le v \le v_{max} \tag{38}$$

$$0 \le k_N \le k_{N,max} \tag{39}$$

$$0 \le F_{TU} \le F_{TU,max} \tag{40}$$

$$l_{o,min} \le l_o \le l_{o,max} \tag{41}$$

$$l_{u,min} \le l_u \le l_{u,max} \tag{42}$$

$$D_o \in \mathfrak{D} \tag{43}$$

$$D_u \in \mathfrak{D} \tag{44}$$

$$d_o \in \mathfrak{d} \tag{45}$$

$$d_u \in \mathfrak{d} \tag{46}$$

In the above equations, the subscripts *min* and *max* denote, respectively, the lower and upper limits of the related design parameters, $\mathfrak{D}_{tr}$ denotes the set of recommended diameters of the drive pulleys, $\mathfrak{B}$ denotes the set of recommended belt width, $\mathfrak{D}$ denotes the set of recommended diameters of the idler rolls and $\mathfrak{d}$ denotes the set of recommended shaft diameters of the idler rolls.

Based on the above development, the optimization problem that minimizes the life cycle cost of the multi-drive belt conveyors fitted with N intermediate drive stations is

$$\min_X \quad \text{Eq. (7)}$$
$$\text{s.t. Eqs. (13)–(46)}$$

For readability reasons, the full optimization program is reproduced in Appendix B.

5. Case study

A simulation based case study is presented in this section to demonstrate the effectiveness of the optimal multi-drive belt conveyor design model proposed.

5.1. Simulation setup

The requirement is to design a multi-drive belt conveyor capable of transporting a certain bulk material with a flow rate of 3500 t/h over a distance of 2500 m with an inclination of 1 in 100. The description of this transportation task along with the technical parameters are listed in Table 1. In practice, the value of the hypothetical friction factor is affected by several factors, including the belt tension, conveyor speed, diameters of idler rolls and their spacing as explained in the DIN 22101 standard (DIN 22101, 2011). While it usually varies between 0.010 and 0.040, no approach to set f is mentioned in case its affecting parameters are independently varied between their usual limits. Accordingly, a fixed value of 0.03 is adopted in this study as mentioned in Table 1. For a certain design task, this should be picked by the plant designer who has knowledge about this factor. Further, the case study assumes a unique loading point at the tail pulley, a unique unloading point at the head pulley and a single belt cleaning device installed downstream of the head pulley. The standard values of $F_{max,o}$ and $F_{max,u}$ applicable to, respectively, three-idler troughing configurations and flat return idler rolls for common belt widths and shaft diameters as considered in this case study are disclosed in Frittella and Curry (2009).

The cost implications of the optimally designed multi-drive conveyor is compared with that of an optimally designed single drive conveyor for a fair comparison. In particular, an optimization model for the design of single drive belt conveyors with a unique head drive pulley was developed. This model is a modification of the design model for multiple drive belt conveyors presented

Table 1
Technical parameters of the case study.

Parameter	Value	Unit	Parameter	Value	Unit
Transport parameters			m_2	$8.174 \cdot 10^{-3}$	
L	2500	m	m_3	1.002	
H	25	m	m_4	0.0124	
Q	3500	t/h	m_5	0.771	
ρ	1280	kg/m^3	S_0	1.1	
β	20	°	S_1	1.7	
L_f	1		*Drive station parameters*		
h_{rel}	1	%	μ	0.3	
Resistance parameters			η_{gear}	0.9	
a	1.25		η_{mot}	0.95	
$C_{Schb}C_{Rank}$	1		*Idler roll parameters*		
f	0.03		n_1	0.714	
g	9.81	m/s^2	n_2	0.089	
k_b	1.1		B_f	0.80	
$v_{0,j}$	0	m/s	C_f	1.25	
μ_1	0.6		*Carry idler roll parameters*		
μ_2	0.6		$z_{1,o}$	139.39	
μ_3	0.65		$z_{2,o}$	1.722	
p_{Gr}	0.065	N/mm^2	$z_{3,o}$	1.025	
e	0.031	mm	$z_{4,o}$	80.51	kg/m
Belt parameters			*Return idler roll parameters*		
λ	35	°	$z_{1,u}$	172	
c_{Tr}	145		$z_{2,u}$	1.287	
$k_{t,rel}$	0.45		$z_{3,u}$	1	
m_1	13.823		$z_{4,u}$	124.99	kg/m

earlier in this study with the number of intermediate drive stations N set to zero. Certain design conditions among (17)–(14) and (15)–(27) were also modified to reflect the absence of intermediate stations and the use of a unique drive pulley in the belt conveyor.

5.2. Economic parameters and assumptions

Table 2 displays the economic parameters and assumptions considered in the simulation study. The straight line depreciation method is adopted for all the conveyor components and the depreciation costs remaining at the end of the project lifetime are written off. Further the annual cost escalation rate of the different equipment is assumed equal to the inflation rate given in Table 1. While the lifetimes of the belts, motors and gearbox follow the recommendations of the US Bureau of Economic Analysis (BEA) (US Bureau of Economic Analysis), that of the idler rolls is fixed at 40 000 h in accordance with SANS 1313 standard (Frittella and Curry, 2009). The salvage values are however assumed by the authors. Using the calculation method described in Appendix A, the following equivalent annual cost coefficients were obtained: $k_1 = 1.653$, $k_2 = 0.138$, $k_3 = 0.055$, $k_4 = 0.055$, $k_5 = 0.148$, $k_6 = 0.148$.

With regard to the equipment prices, the initial cost coefficients indicated in Table 2 were determined from information provided by suppliers in South Africa or abroad, with an estimate of the shipping costs for the latter case. The interested reader is referred to reference (Masaki, 2017) for calculation details. Note that while the relative differences between the energy and component costs may vary by country, this study aims to provide the basic principles by which economic design of multiple drive belt conveyors may be achieved.

The lower limits, upper limits, and set of possible values that apply to the various design parameters are summarized in Table 3.

Table 3
Boundary limit values.

Parameter	Unit	Min	Max	Set
Parameter	Unit	Min	Max	Set
P_i	kW	0	2000	–
T_i	kNm	0	950	–
$D_{tr,i}$	m	–	–	0.1-0.16-0.2-0.25-0.315-0.4-0.5-0.63-0.8-1-1.25-1.4-1.6-1.8-2-2.2
α_i	°	180	240	–
$L_{o,j}$	m	0	2500	–
B	m	–	–	0.6-0.75-0.9-1.05-1.2-1.35-1.5-1.8-2-2.2-2.4
v	m/s	0	10	–
k_N	kN/m	0	3000	–
F_{TU}	kN	0	500	–
l_o	m	1	2	–
l_u	m	1	4.5	–
D_o, D_u	mm	–	–	63-76-89-102-108-127-133-152-159-194
d_o, d_u	mm	–	–	25-30-35-40

5.3. Results and discussions

The above optimization problems of the economic design of single and multi-drive conveyors are formulated as mixed integer non-linear programming (MINLP) problems and were solved using the MIDACO solver, which is a general-purpose solver based on an extended evolutionary ant colony optimization algorithm (Schlüter et al., 2009). For a given N, an MINLP problem is generated and subsequently solved by the optimizer in order to determine the most cost-effective conveyor design for the considered number of intermediate drive stations. Then, a different N is set and the resulting MINLP problem is solved again. At the end, all solutions are compared to each other to determine the best design in terms of N and other parameters.

Table 2
Economic parameters of the case study.

Description	Value & Unit
General parameters	
project lifetime, Z	20 years
running time, t_a	12 h/day over 300 days par annum
initial energy cost, e_o	0.071 USD/kWh
annual escalation rate of energy costs, r_e	11.19%
inflation rate	5.6%
tax rate	28%
proportion of debt capital	0%
after tax return required on equity funds with 0% inflation rate	5%
Belt parameters	
expected lifetime	16 years
salvage value	0%
initial cost coefficients, c_1; c_2; c_3	25.965; 0.0014; 1.313
belt length coefficients, y_1; y_2	3 m; 20 m
Motors	
expected lifetime	16 years
salvage value	8%
initial cost coefficients, c_4; c_5; c_6	248.12; 69.062; 1.013
Gear reducers	
expected lifetime	16 years
salvage value	10%
initial cost coefficients, c_7; c_8; c_9	5699.3; 1563.1; 1.081
Carry idler rolls	
expected lifetime	11 years
salvage value	0%
initial cost coefficients, c_{10}; c_{11}; c_{12}; c_{13}; c_{14}; c_{15}; c_{16}	-69.77; $1.312 \cdot 10^{-5}$; 3.096; 4.828; 1; $1.079 \cdot 10^{-4}$; 1.829
Return idler rolls	
expected lifetime	11 years
salvage value	0%
initial cost coefficients, c_{17}; c_{18}; c_{19}; c_{20}; c_{21}; c_{22}; c_{23}	-30.5; 0.565; 0.676; 0.571; 1.054; 0.567; 0.676

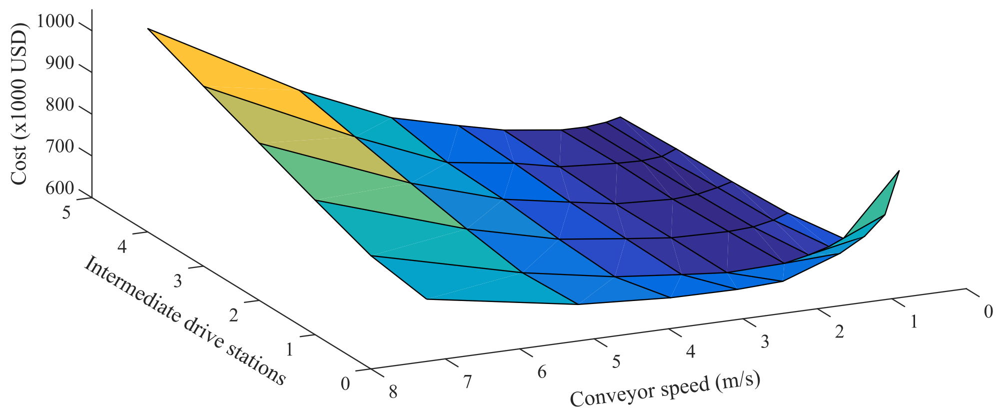


Fig. 4. Optimal belt conveyor cost vs Belt speed vs Intermediate drives.

Following Subsection 4.3.4, the design parameters $D_{tr,i}$, B, D_o, D_u, d_o and d_u were specified as discrete variables in the solver because of the limited number of recommended sizes (DIN 22101, 2011; Frittella and Curry, 2009). Treated as of integer type during the internal optimization process, the values of these parameters in each of the solutions generated are first mapped to the corresponding actual sizes of the sets prescribed in Table 3 prior to evaluating the objective and constraint functions.

For illustration purposes, Fig. 4 displays the equivalent annual cost of conveyors obtained for all the possible widths of belt and a number of intermediate drive stations limited at 5. The conveyor designs with zero intermediate drive station corresponds to the single drive belt technology. This figure shows that, in general, the economic benefits of the multi-drive technology will be more effective at low conveyor speed, while the single drive design will be the most beneficial option at high conveyor speed. A larger impact of the conveyor speed is also noted on the multiple drive conveyors compared to the single drive belt conveyors.

The synthesis of the lowest equivalent annual cost with respect to number of intermediate drive stations fitted is shown in Fig. 5 along with their respective energy and capital costs. This figure shows that the belt conveyor fitted with three intermediate drive stations operating at 1.69 m/s constitutes the most cost-effective design for the considered transport operation. By adopting the most economic single-drive belt conveyor as the reference conveyor, the equivalent annual cost savings expected from the most cost-effective multi-drive conveyor is estimated at approximately 63 131 $(USD) per annum over the 20 years of the project lifetime.

Regarding operating and capital costs, Fig. 5 shows that the observed decrease in cost in comparison with the reference conveyor is primarily due to the lower energy expenses involved. Especially, within the 63 131 $(USD) of cost savings achieved, 62.44% savings come from energy costs and the other 35.56% from the capital costs. With the increase in the number of drive stations, the energy cost tends to decline, while a decrease in the capital cost is first observed, followed by a progressive increase. The variation in the energy cost is, however, slowed by the extra inertia resistance and frictional resistance brought in by every additional drive station fitted in the conveyor system. Fig. 5 also indicates that the optimal conveyor speed decreases gradually as more drive stations are fitted.

The breakdown of the power consumption per load component shown in Fig. 6 suggests that every additional drive station can assist to achieve higher energy efficiency in transportation, because of the reduction in energy consumption from the belt and the idler rolls. Accordingly, and given (8) and (12), higher cost savings can therefore be expected for conveyors with longer transport distances.

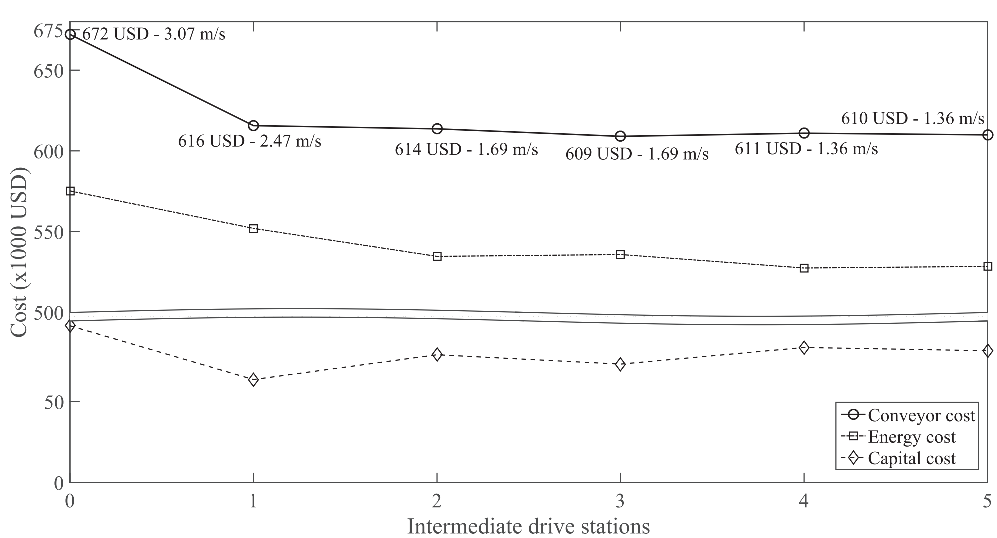


Fig. 5. Minimum belt conveyor cost per number of intermediate stations.

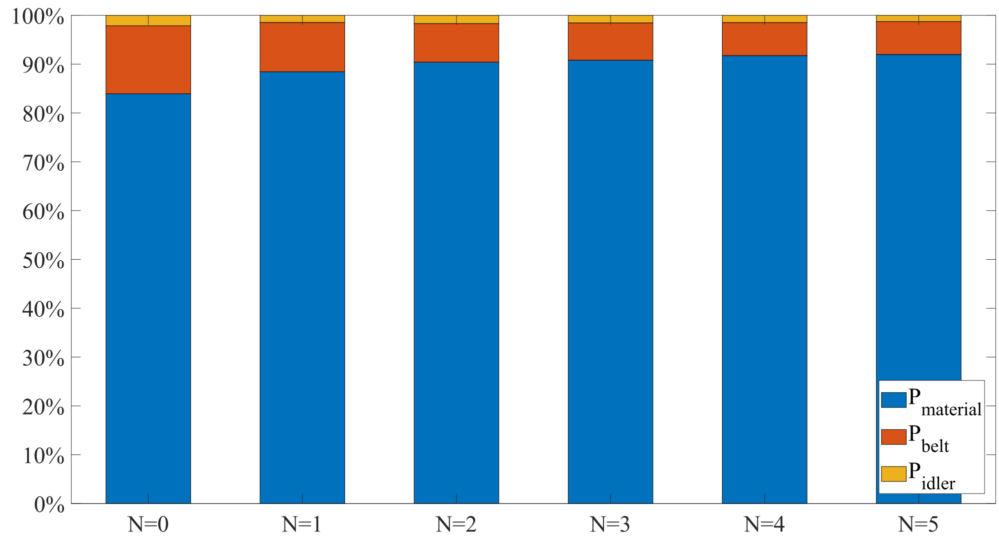


Fig. 6. Power consumption per conveyor load component.

To investigate the individual contributions of the conveyor components in the capital costs, Fig. 7 is presented. It is observed that the cost of the belt will generally form the largest portion, followed by the gear reducers. Beyond a single intermediate drive station, the cost savings induced by the use of lighter belts are balanced and gradually defeated by the need for larger belts. For the motors, Fig. 7 indicates that their cost is fairly stable, irrespective of the number of drive stations. Lastly, the belt width has a greater impact on the cost of carry idler rolls than that of the return idler rolls.

In order to evaluate the validity of the previous simulation results in case of the inflation rate fluctuation throughout the project lifetime, two additional scenarios were simulated. The first scenario considers net decrease of 5% in the initial inflation rate accompanied by a stochastic factor restrained at ±0.1%. An average inflation rate of 5.2% with a minimum of 4.8% and a maximum 5.8% was observed in this case, which resulted in the following equivalent annual cost coefficients: $k_1 = 1.716$, $k_2 = 0.138$, $k_3 = 0.053$, $k_4 = 0.052$, $k_5 = 0.145$, $k_6 = 0.145$. The second scenario consisted of a net increase of 5% in the initial inflation rate, accompanied by another stochastic factor also restrained at ±0.1%. An average inflation rate of 5.8% with a minimum of 5.4% and a maximum 6.4% was observed in this case, which resulted in the following equivalent annual cost coefficients: $k_1 = 1.613$, $k_2 = 0.137$, $k_3 = 0.057$, $k_4 = 0.057$, $k_5 = 0.147$, $k_6 = 0.147$.

Fig. 8 shows a synthesis of the most economic results with respect to the number of intermediate drive stations fitted in the conveyor system. The optimal design solution that involves 3 intermediate drive stations with a belt speed at 1.694 m/s appears to remain the most advantageous option under the three scenarios. Detailed performance of the optimal design under each scenario are given in Table 4. The comparison of the three cost-effective design solutions shows that, apart from α_i, k_N, F_{TU}, l_o and l_u, the rest of design parameters maintain the same values under the three different scenarios. Further tests focusing on α_i showed that greater values can also apply to the original most cost-effective multi-drive

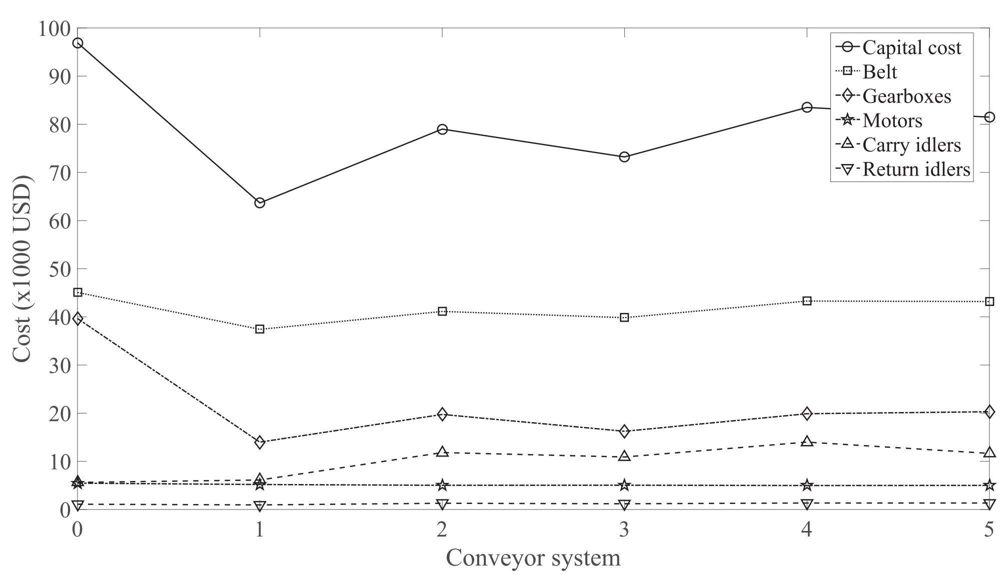


Fig. 7. Detailed capital cost per belt conveyor component.

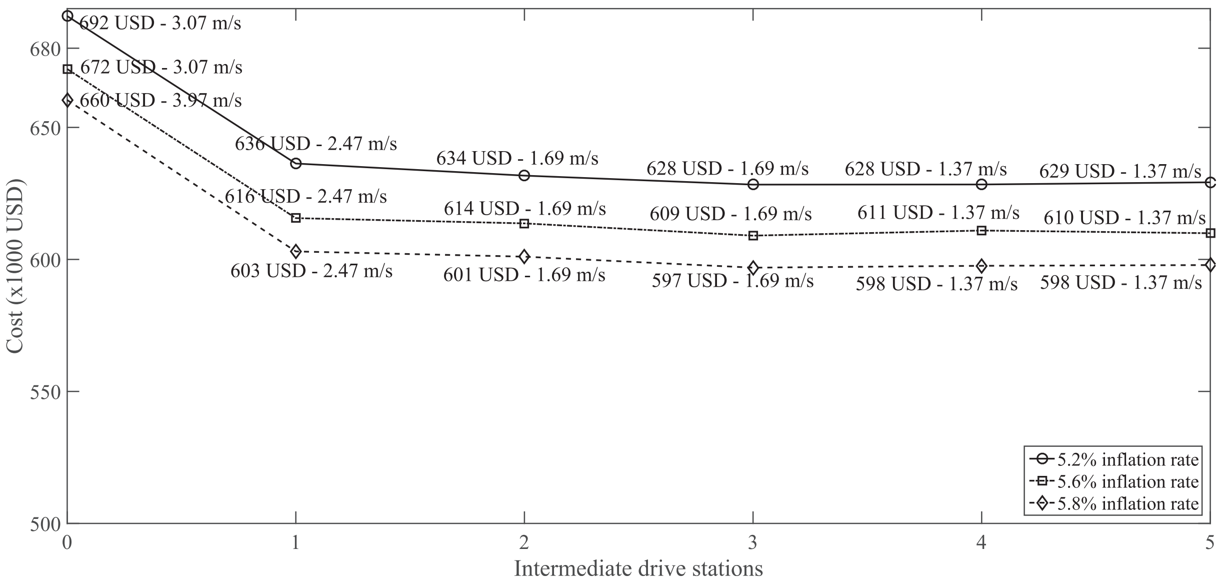


Fig. 8. Variation of the cost-effective belt conveyors with the inflation rates.

belt conveyor, that is under 5.6% inflation rate, and the cheapest conveyor system under 5.2% inflation rate without affecting their respective economic performance. Such an increase on α_i will benefit the design condition (16) on the slack side tension of drive stations. Further the comparison of k_N under the 5.2% and 5.8% inflation rate fluctuation scenarios with the fixed inflation scenario at 5.6% per year indicates a decrease of, respectively, 7.81% and 8.76% in this parameter in case the presumed fluctuations of inflation rate occur in course of the project. Also in comparison with the original most cost-effective multi-drive belt conveyor, it is observed a decrease of l_o by 19.01% and 21.13% and an increase of l_u by 15.38% and 15.38% under the 5.2% and 5.8% inflation rate fluctuation scenarios, respectively. This suggests that the additional expenses induced by the use of the original belt (greater k_N) and the original return idler roll spacing (smaller l_u) under the fluctuating inflation scenarios will be partially offset by gains resulting from the reduction in the quantity of carry idler rolls installed (greater l_o). In view of the preceding analysis, the original most cost-effective multi-drive belt conveyor is fairly robust in case of limited fluctuations of the inflation rate.

Table 5 summarizes the operational environmental footprint of the cost-effective conveyors in the event they are supplied by a coal-fired power plant (Eskom, 2017). It shows that the design solution with four intermediate stations will ensure the lowest emission of CO_2 and particulates and water consumption due to electricity generation, followed by the multi-drive conveyors fitted with five and three intermediates stations. In case priority is given to the economic aspects, the implementation of the most cost-effective system fitted with three intermediate station will result in a yearly reduction of 333.56 kg in CO_2 emissions, 101.07 ton in particulate emissions and 471.69 kl in water consumption due to electricity generation, with respect to the single drive contender. This shows that the multi-drive technology can help reduce the environmental nuisance of belt conveyors.

Table 4
Cost-effective belt conveyor designs under different inflation trends.

Parameter	Average inflation rate		
	5.2%	5.6%	5.8%
N	3	3	3
P_i, kW	151.05	151.09	151.08
v, m/s	1.69	1.69	1.69
T_i, kNm	16.05	16.05	16.05
$L_{o,1}$, m	574.63	572.98	574.77
$L_{o,3}$, m	649.14	649.7	649.1
$L_{o,5}$, m	649.14	649.7	649.1
$L_{o,7}$, m	628.41	628.94	628.35
α_i, rad	3.32	3.8	4.19
l_o, m	1.15	1.42	1.12
l_u, m	4.5	3.9	4.5
F_{TU}, kN	67.67	86.95	65.26
k_N, kN/m	526.68	571.28	521.22
B, mm	1800	1800	1800
D_o, mm	63	63	63
D_u, mm	63	63	63
d_o, mm	30	25	30
d_u, mm	30	25	30
$D_{tr,i}$, mm	400	400	400
A_{energy}, × 1000 USD/year	556.16	535.79	522.84
A_{belt}, × 1000 USD/year	39.02	39.83	38.66
A_{motor}, × 1000 USD/year	4.82	5.04	5.22
A_{gear}, × 1000 USD/year	15.51	16.24	16.82
$A_{carryidler}$, × 1000 USD/year	11.56	10.9	12.05
$A_{returnidler}$, × 1000 USD/year	1.27	1.19	1.28
$A_{conveyor}$, × 1000 USD/year	628.34	609.01	596.88

6. Conclusion

An original contribution to the cost-effective design of multiple drive belt conveyors was presented in this paper. To achieve the lowest life cycle cost for a specified material handling operation, the proposed design approach takes into account a significant

Table 5
Environmental assessment of cost-effective belt conveyors.

Design	Energy (MWh/yr)	CO_2 emiss. (kg/yr)	Partic. emiss. (ton/yr)	Water use (kl/yr)
N = 0	4917.31	4868.14	1475.19	6884.23
N = 1	4718.73	4671.54	1415.62	6606.22
N = 2	4570.60	4524.89	1371.18	6398.84
N = 3	4580.38	4534.58	1374.12	6412.54
N = 4	4466.51	4421.85	1339.95	6253.11
N = 5	4517.56	4472.38	1355.27	6324.58

number of parameters, including the number of intermediate drive stations, their distribution along the conveyor path and the conveyor speed. Simulations carried out on a practical transport operation established the validity and effectiveness of the proposed design approach. An expected annual cost saving of 63 131 \$(USD) was achieved by the most cost-effective multi-drive conveyor over the best single drive alternative. This was accompanied by a yearly reduction of 333.56 kg in CO_2 emissions, 101.07 ton in particulate emissions and 471.69 kl in water consumption due to electricity generation. The robustness of the most cost-effective conveyor designs against the fluctuation of the inflation rate was also confirmed. It is concluded that multi-drive belt conveyors are more advantageous for long distance slow speed material transportation while single drive technology is preferable for short distance fast speed applications. The scope of future works includes adding the capital costs of other conveyor components such as the supporting structure, the pulleys and the take-up device, and also the substitution of the simplified frictional resistance models from the DIN 22101 standard by advanced belt movement models.

Acknowledgements

The authors acknowledge the support for this work provided by the National Hub for Energy Efficiency and Demand Side Management and the National Research Foundation of South Africa under grant unique number 105047.

Appendix A. Calculation of the equivalent annual cost coefficients of energy and equipment

Appendix A.1. Equivalent annual energy cost coefficient

This section presents the approach on the determination of the equivalent annual energy cost coefficient in case the annual escalation rate of energy r_e and the general inflation rate r can vary from year to year. It therefore extends and also summarizes the procedure explained in the literature (Roberts, 1981; Eschenbach, 2003). In this context, k_1 is given by:

$$k_1 = \left(\frac{a}{p}\right)_Z^{i_f^0} \sum_{i=1}^{Z} \left(\frac{p}{f}\right)_i^{i_f} \prod_{j=1}^{i} (1 + r_{e,j}), \tag{A.1}$$

where $\left(\frac{a}{p}\right)_Z^{i_f^0}$ denotes the capital recovery factor of the project, $\left(\frac{p}{f}\right)_i^{i_f}$ denotes the present equivalent cost factor over i-year period of time, $r_{e,j}$ denotes the annual escalation rate of energy during the year j of the project, and Z denotes the project lifetime. The capital recovery factor is obtained by:

$$\left(\frac{a}{p}\right)_Z^{i_f^0} = \frac{i_f^0\left(1 + i_f^0\right)^Z}{\left(1 + i_f^0\right)^Z - 1}, \tag{A.2}$$

where i_f^0, the time value of money when all cash flows are converted from inflated value to constant year zero value, is given by:

$$i_f^0 = \frac{(1-t)r_d i_d - r_d r_{avg}}{1 + r_{avg}} + (1 - r_d)i_e. \tag{A.3}$$

Here, t denotes the income tax rate, r_d denotes the proportion of debt capital maintained constant by the company, i_d denotes the interest rate on debt, r_{avg} denotes the average general inflation rate over the project duration, and i_e denotes the after-tax return required on equity funds with zero inflation rate. In case the general inflation rate can vary throughout the project, the present equivalent cost factor over i-year period of time is given by:

$$\left(\frac{p}{f}\right)_i^{i_f} = \frac{1}{\prod_{j=1}^{i}\left(1 + i_{f,j}\right)}, \tag{A.4}$$

where $i_{f,j}$ denotes inflation modified rate of return of the year j of the project, which is expressed by:

$$i_{f,j} = (1-t)r_d i_d + (1 - r_d)\left[(1 + r_j)(1 + i_e) - 1\right]. \tag{A.5}$$

Here, r_j denotes the general inflation rate during the year j of the project.

The substitution of (A.2) and (A.4) into (A.1), taking into account (A.3) and (A.5), allows to determine the equivalent annual energy cost coefficient.

Appendix A.2. Equivalent annual cost of equipment

Considering an equipment (e.g. belt), one or several items can be required during the project as a function of the project duration and the expected lifetime of the equipment as well. In the rest of this section, the concept "equipment" will therefore refer to the set of items purchased throughout the project. Let k_{eq} and $C_{eq,0}$ denote, respectively, the equivalent annual cost coefficient of an equipment and the first costs of the first item purchased at the year zero of the project. The equivalent annual cost A_{eq} of this equipment can be expressed as (Roberts, 1981):

$$A_{eq} = k_{eq} C_{eq,0}. \tag{A.6}$$

It can be also obtained by multiplying the present equivalent of the capital costs PEC_{eq} of the equipment by the capital recovery factor of the project:

$$A_{eq} = \left(\frac{a}{p}\right)_Z^{i_f^0} PEC_{eq}. \tag{A.7}$$

Taking into account the first item and the replacement items purchased during the project period, the present equivalent of the capital costs of an equipment is expressed by (Roberts, 1981):

$$PEC_{eq} = \frac{PEF_{eq} - PEV_{eq} - tPED_{eq}}{1 - t}, \tag{A.8}$$

where PEF_{eq} denotes the present equivalent of the first costs of all the items, PEV_{eq} denotes the present equivalent of the salvage values of all the items, and PED_{eq} denotes the present equivalent of depreciations of all the items.

Denote M the expected lifetime of a given equipment, the total number of items to be purchased over the Z years of the project, denoted by R is given by:

$$R = \frac{Z}{M},$$

The year X_i of the purchase of the i-th item ($i = 1, \ldots, R$) is given therefore by:

$$X_i = (i - 1)M.$$

In case the inflation-modified rate of return and the annual cost escalation rate r_{eq} of the equipment vary from year to year during the project period, PEF_{eq} is given by:

$$PEF_{eq} = C_{eq,0}\left(1 + \sum_{i=2}^{R}\left(\frac{p}{f}\right)_{X_i}^{i_f}\prod_{j=1}^{X_i}(1+r_{eq,j})\right). \tag{A.9}$$

The sum in parenthesis will vanish if the first item purchased is used over the entire project duration.

On the calculation of the salvage of the equipment, let q_f denote the estimated remaining value in percentage of the first costs of the equipment after it operates over the expected lifetime. Assuming the value of equipment decreases linearly with time, the remaining value q_i of the i-th item purchased ($i = 1,\ldots,R$) after it operates over its actual lifetime with respect to the project duration is given by:

$$q_i = 1 - \frac{1-q_f}{M}\min(M, Z - M(i-1)),$$

The year Y_i of the decommissioning of the i-th item should correspond to the minimum between the year of the purchase of the next item or the project end:

$$Y_i = \min(iM, Z), \quad \text{with } i = 1,\ldots,R.$$

By taking into account all the items to be purchased during the project lifetime and the annual increase on the first costs, the present equivalent of the salvage value of the equipment is given by:

$$PEV_{eq} = C_{eq,0}\left(\left(\frac{p}{f}\right)_{Y_1}^{i_f} q_1 + \sum_{i=2}^{R}\left(\frac{p}{f}\right)_{Y_i}^{i_f} q_i \prod_{j=1}^{X_i}(1+r_{eq,j})\right). \tag{A.10}$$

The sum in parenthesis will vanish if the first item purchased is used over the whole project duration.

In order to formulate the present equivalent of depreciation under varying inflation-modified rate of return, let $\left(\frac{f}{a}\right)_{M,X_i}^{i_f}$ denote the series compound amount factor expressed as follows:

$$\left(\frac{f}{a}\right)_{M,X_i}^{i_f} = 1 + \sum_{i=2}^{M}\prod_{j=i}^{M}\left(1 + i_{f,X_i+j}\right).$$

This factor converts a uniform series of annual depreciation to a future value for an item purchased at the year X_i of the project and operated over M years

The present equivalent of depreciation of an equipment is obtained by summing up the present equivalent of the future value of the annual depreciation of all the items, taking into account the annual cost escalation rate of the equipment. Adopting the straight-line depreciation method and writing off the depreciation charges remaining at the end of the project, this results in:

$$\begin{aligned} PED_{eq} = \frac{C_{eq,0}}{M}\Bigg(&\sum_{i=1}^{R-1}\prod_{j=1}^{X_i}(1+r_{eq,j})\left(\frac{p}{f}\right)_{X_i+M}^{i_f}\left(\frac{f}{a}\right)_{M,X_i}^{i_f} \\ &+ \prod_{j=1}^{S_R}(1+r_{eq,j})\left(\frac{p}{f}\right)_{Z}^{i_f}\left(\frac{f}{a}\right)_{Z-S_R,S_R}^{i_f} \\ &+ (QM-Z)\prod_{j=1}^{S_R}(1+r_{eq,j})\left(\frac{p}{f}\right)_{Z}^{i_f}\Bigg). \end{aligned} \tag{A.11}$$

In case a unique item operates over the entire project duration, that is $R = 1$, PED_{eq} is simplified as follows

$$PED_{eq} = \frac{C_{eq,0}}{M}\left(\left(\frac{p}{f}\right)_{Z}^{i_f}\left(\frac{f}{a}\right)_{Z,0}^{i_f} + (M-Z)\left(\frac{p}{f}\right)_{Z}^{i_f}\right). \tag{A.12}$$

Keeping $C_{eq,0}$ factorized, the successive substitution of (A.9), (A.10) and (A.11) or (A.12) into (A.8), and of (A.8) into (A.7) allows to determine k_{eq} indicated in (A.6).

Appendix B. Optimization program for multi-drive belt conveyors

The optimization problem that minimizes the life cycle cost of a multi-drive belt conveyor equipped with N intermediate dive stations is given by

$$\begin{aligned} \min_{X} \quad & A_{energy} + A_{belt} + 2\sum_{i=1}^{N+1} A_{motor,i} \\ & +2\sum_{i=1}^{N+1} A_{gearreducer,i} + A_{carryidler} + A_{returnidler} \\ \text{s.t.} \quad & \rho A_{th} v = Q, \\ & \sum_{k=1,3,\ldots}^{N_o} L_{o,k} - \sum_{i=1}^{N} D_{tr,i} = L/\cos\delta, \\ & 2\sum_{i=1}^{N+1} P_i \eta_{gear,i} - vF_W = 0, \\ & \frac{2T_i v}{D_{tr,i}} = \eta_{gear,i} P_i, \quad i = 1,\ldots,N+1, \\ & F_{T2,j} \geq \frac{2C_{w,j} P_j \eta_{gear,j}}{v}, \quad j = 1,\ldots,N+1, \\ & F_0 \geq \frac{g(m'_L + m'_G) l_o}{8h_{rel}}, \\ & F_{T2,j} \geq \frac{g(m'_L + m'_G) l_o}{8h_{rel}}, \quad j = 1,\ldots,N, \\ & F_{TD} \geq \frac{g m'_G l_u}{8h_{rel}}, \\ & F_{T1,i} = F_{T1,1}, \\ & \frac{k_{t,rel} k_N}{S_0 S_1} \geq \frac{F_{T1,1}}{B}, \\ & D_{tr,j} \geq c_{Tr} d_{Gk}, \quad j = 1,\ldots,N+1, \\ & S_f B_f L_f F_{s,o} \leq F_{max,o}, \\ & S_f B_f C_f F_{s,u} \leq F_{max,u}, \\ & P_i = P_1, \\ & T_i = T_1, \\ & D_{tr,i} = D_{tr,1}, \\ & \alpha_i = \alpha_1, \\ & \frac{60v}{\pi D_o} \leq 750, \\ & \frac{60v}{\pi D_u} \leq 750, \end{aligned}$$

with the design parameters subject to the boundary limits

$$
\begin{aligned}
&0 \le P_i \le P_{max}, \quad i = 1, \dots, N+1,\\
&0 \le T_i \le T_{max}, \quad i = 1, \dots, N+1,\\
&D_{tr,i} \in \mathfrak{D}_{tr}, \quad i = 1, \dots, N+1,\\
&\alpha_{min} \le \alpha_i \le \alpha_{max}, \quad i = 1, \dots, N+1,\\
&L_{min} \le L_{o,j} \le L_{max}, \quad j = 1, \dots, N_o,\\
&B \in \mathfrak{B},\\
&0 \le v \le v_{max},\\
&0 \le k_N \le k_{N,max},\\
&0 \le F_{TU} \le F_{TU,max},\\
&l_{o,min} \le l_o \le l_{o,max},\\
&l_{u,min} \le l_u \le l_{u,max},\\
&D_o \in \mathfrak{D},\\
&D_u \in \mathfrak{D},\\
&d_o \in \mathfrak{d},\\
&d_u \in \mathfrak{d}.
\end{aligned}
$$

References

Alspaugh, M.A., 2003. The evolution of intermediate driven belt conveyor technology. Bulk. Solids Handl. 23 (3), 168–172.

Association for Rubber Products Manufacturers Inc., Indianapolis, US, 2011. Conveyor and Elevator Belt Handbook.

Bindzár, P., Malindžák, D., 2008. Number of conveyor belts optimization regarding to its type and logistical parameters in mining industry. Acta Montan. Slovaca 13 (4), 524–531.

Conveyor Equipment Manufacturers Assoc., Naples, US, 1997. Belt conveyors for Bulk Materials, fifth ed.

Dalgleish, A., Grobler, L., 2003. Measurement and verification of a motor sequencing controller on a conveyor belt. Energy 28 (9), 913–927.

Darling, P. (Ed.), 2011. SME mining Engineering Handbook, third ed. Society for Mining Metallurgy & Exploration, Englewood.

de Almeida, A.T., Fonseca, P., Bertoldi, P., 2003. Energy-efficient motor systems in the industrial and in the services sectors in the European Union: characterisation, potentials, barriers and policies. Energy 28 (7), 673–690.

Dejchanchaiwong, R., Arkasuwan, A., Kumar, A., Tekasakul, P., 2016. Mathematical modeling and performance investigation of mixed-mode and indirect solar dryers for natural rubber sheet drying. Energy Sustain. Develop. 34, 44–53.

DIN 22101, 2011. Continuous Mechanical Handing Equipment - Belt Conveyors with Carrying Idlers - Calculation of Operating Power and Tensile Force.

Eschenbach, T., 2003. Engineering Economy: Applying Theory to Practice, second ed. Oxford University Press, New York.

Eskom, February 2010. The Energy Efficiency Series. Towards an Energy Efficient Mining Sector. http://www.eskom.co.za/sites/idm/Documents/121040ESKD_Mining_Brochure_paths.pdf.

Eskom, March 2017. Integrated Report 2017.

Frittella, A., Curry, S., 2009. Conveyor idlers - sans 1313 and selection procedures. In: Proceedings of Beltcon 15 Conference, Johannesburg, South Africa, pp. 1–14.

Handling Sandvik Materials, Sandviken, Sweden, 2013. Design Criteria for Product Selection - Pulleys, Idlers, and Materials Handing Systems.

He, D., Pang, Y., Lodewijks, G., 2017. Green operations of belt conveyors by means of speed control. Appl. Energy 188, 330–341.

Jeftenić, B., Mihailović, I., Bebić, M., Ristić, L., Jevtić, D., Rašić, N., Štatkić, S., 2010. Energy efficiency in transportation of bulk material with frequency controlled drives. In: 14th International Conference on Power Electronics and Motion Control, Ohrid, Macedonia, pp. 107–113.

Lodewijks, G., 2003. Determination of rolling resistance of belt conveyors using rubber data: fact or fiction? Bulk. Solids Handl. 23 (6), 384–391.

Luo, J., Shen, Y., 2015. Energy efficiency optimization of belt conveyor for material scheduling problem. In: 2015 IEEE International Conference Information and Automation, Lijiang, China, pp. 122–127.

Luo, J., Huang, W., Zhang, S., 2015. Energy cost optimal operation of belt conveyors using model predictive control methodology. J. Clean. Prod. 105, 196–205.

Masaki, M.S., May 2017. A Cost-effective Design Approach for Multiple Drive Belt Conveyor Systems (Masters' thesis). University of Pretoria, South Africa.

Mathaba, T., Xia, X., Zhang, J., 2012. Optimal scheduling of conveyor belt systems under critical peak pricing. In: 10th International Power and Energy Conference, Ho Chi Minh City, Vietnam, pp. 315–320.

Maton, A.E., 2003. The effects of idler selection on conveyor belt power consumption. Bulk. Solids Handl. 22 (1), 46–49.

Mazurkiewicz, D., 2015. Maintenance of belt conveyors using an expert system based on fuzzy logic. Arch. Civ. Mech. Eng. 15 (2), 412–418.

Middelberg, A., Zhang, J., Xia, X., 2009. An optimal control model for load shifting - with application in the energy management of a colliery. Appl. Energy 86 (7–8), 1266–1273.

Nuttall, A.J.G., 2007. Design Aspects of Multiple Driven Belt Conveyors (Ph.D. thesis). TU Delft, Delft University of Technology.

Petrović, D.V., Tanasijević, M., Milić, V., Lilić, N., Stojadinović, S., Svrkota, I., 2014. Risk assessment model of mining equipment failure based on fuzzy logic. Expert Syst. Appl. 41 (18), 8157–8164.

Reicks, A.V., 2008. Belt conveyor idler roll behaviors. In: Alspaugh, M.A. (Ed.), Bulk Material Handling by Conveyor Belt 7, vol. 7. Society for Mining Metallurgy & Exploration, Englewood, pp. 35–40.

Reicks, A.V., Rudolphi, T.J., Wheeler, C.A., 2012. A comparison of calculated and measured indentation losses in rubber belt covers. Bulk. Solids Handl. 32 (3), 52–57.

Ristic, L.B., Bebic, M.Z., Jevtic, D.S., Mihailovic, I.D., Statkic, S.Z., Rasic, N.T., Jeftenić, B.I., 2012. Fuzzy speed control of belt conveyor system to improve energy efficiency. In: 15th International Conference on Power Electronics and Motion Control, Novi Sad, Serbia, pp. 1–7.

Roberts, A., 1981. Economic analysis in the optimization of belt conveyor systems. In: Proceedings of Beltcon 1 Conference, Johannesburg, South Africa, pp. 1–33.

Schlüter, M., Egea, J.A., Banga, J.R., 2009. Extended ant colony optimization for non-convex mixed integer nonlinear programming. Comput. Oper. Res. 36 (7), 2217–2229.

Schützhold, J., Benath, K., Müller, V., Hofmann, W., 2014. Design criteria for energy efficient belt conveyor drives. In: 2014 International Symposium on Power Electronics, Electrical Drives, Automation and Motion, Ischia, Italy, pp. 1256–1263.

Tapp, A., 2000. Conveying technology-energy saving troughing idler technology. Bulk. Solids Handl. 20 (4), 437–449.

US Bureau of Economic Analysis, BEA Rates of depreciation, service lives, declining-balance rates, and Hulten-Wykoff categories. https://bea.gov/scb/account_articles/national/wlth2594/tableC.htm.

Wheeler, C., 2006. Indentation rolling resistance of belt conveyors—a finite element solution. Bulk. Solids Handl. 26 (1), 40–43.

Wheeler, C.A., 2008. Evolutionary belt conveyor design: optimizing costs. In: Alspaugh, M.A. (Ed.), Bulk Material Handling by Conveyor Belt 7, vol. 7. Society for Mining Metallurgy & Exploration, Englewood, pp. 3–11.

Windmann, S., Niggemann, O., Stichweh, H., 2015. Energy efficiency optimization by automatic coordination of motor speeds in conveying systems. In: 2015 IEEE International Conference on Industrial Technology, Seville, Spain, pp. 731–737.

Xia, X., Zhang, J., 2010. Energy efficiency and control systems-from a POET perspective. In: Conference on Control Methodologies and Technology for Energy Efficiency, Vilamoura, Portugal, pp. 225–260.

Xia, X., Zhang, J., 2011. Modelling and control of heavy-haul trains. IEEE Contr. Syst. Mag. 31 (4), 18–31.

Xia, X., Zhang, L., 2016. Industrial energy systems in view of energy efficiency and operation control. Annu. Rev. Contr. 42, 299–308.

Zhang, S., Xia, X., 2010. Optimal control of operation efficiency of belt conveyor systems. Appl. Energy 87 (6), 1929–1937.

IET Control Theory & Applications

Research Article

Lighting retrofit and maintenance models with decay and adaptive control

ISSN 1751-8644
Received on 16th June 2017
Revised 14th November 2017
Accepted on 11th December 2017
E-First on 2nd January 2018
doi: 10.1049/iet-cta.2017.0598
www.ietdl.org

Zhou Wu[1,2] ✉, Kai Zhao[2], Xiaohua Xia[3]

[1]*Key Laboratory of Dependable Service Computing in Cyber Physical Society (Chongqing University), Ministry of Education, People's Republic of China*

[2]*School of Automation, Chongqing University, Chongqing 400044, People's Republic of China*

[3]*Department of Electrical Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa*

✉ *E-mail: wuzhsky@gmail.com*

Abstract: In lighting retrofit projects, a lamp population is subject to decay, which results in significantly deteriorated energy efficiency (EE) and reduced cost saving. Incremental retrofit and maintenance are studied to overcome the decay in the population, so that EE performance can be sustained. Current models of natural decay cannot reflect the interactive dynamics of incremental retrofit and maintenance, so a new decay model is proposed for these interventions. Using a control approach, a multiple-input and multiple-output state equation is formulated. Adaptive control laws are designed to cope with unknown parameters of the proposed model, and to achieve stable performance improvement. This new model is verified, based on empirical data, and the results of adaptive control indicate that the number of working lamps can be maintained as a required value.

1 Introduction

The deployment of energy efficiency (EE) programmes, as a kind of demand-side management (DSM), is one of the most useful alternative solutions for reducing power demand and greenhouse gas emissions [1–3]. With around 40% of the total demand, the building sector will have a great potential to reduce total demand, therefore improving the building EE becomes urgent [3, 4]. Since the start of this century, many policies and projects concerning building retrofit (also referred to as innovation or refurbishment in the literature) have been initiated all over the world to improve building EE, as building retrofit is currently the most feasible and practical way to reduce the demand of the building sector.

Many building retrofit projects are relevant to lighting retrofit [5–7]. Due to easy accessibility and energy saving, light retrofit projects are promoted in various EE incentive programmes, such as clean development mechanisms (CDM) [8], white tradable certificate schemes [9], DSM, and performance contracting [10]. In lighting retrofit projects, energy-efficient lamps, such as compact fluorescent light (CFL) and light-emitting diodes (LED), are used to retrofit less efficient incandescent lights. In general, building EE retrofit (BEER) refers to changing out-of-date facilities in existing buildings through innovative and efficient technologies for lighting, water heating, ventilation/cooling/heating, building envelope, and other energy-consuming systems [11, 12]. There are a large number of these energy systems, and their EE performance has highly complicated correlations. Therefore, designing an optimal retrofit strategy for minimal building energy consumption is a difficult task of BEER, especially BEER on a large scale.

In a recent study [13], the large-scale BEER was defined, modelled, and optimised in a time-building-technology framework. The large-scale BEER was unveiled in three dimensions, i.e. time, building, and technology. In the building dimension, different types of building, such as, office, commercial/residential/industrial buildings, school, and hospital, will be assigned different priorities for retrofit in a large-scale BEER project. In the technology dimension, different types of technology will have different priorities for retrofit. In the time dimension, incremental retrofit can be done every year, and investment is also assigned a different amount each year. In this framework of large-scale BEER, there remain several open issues that require further study, such as decay and maintenance.

In this paper, decay and maintenance in lighting retrofit projects will be studied, because lighting projects are representative and relatively simple to model. Lighting projects involve large populations that are suitable for statistical models. For example, energy-saving control strategies are proposed to minimise the energy consumption of multi-group lighting sources [14, 15]. For the LED lighting systems, lumen depreciation is studied by diagnosing individual LED failures using a photosensor system [16].

For lighting retrofit projects, one practical issue in the time dimension is to model the decay of the population with multiple interventions, such as incremental retrofit and maintenance [17]. The decay model of once-off retrofit has been studied in lighting retrofit projects, in which accurate decay models provide basis knowledge to design retrofit plans, and cost-effective metering plans [4, 18, 19]. In fact, the population of installed efficient lights are subject to deterioration due to certain factors, such as flickering, lamp burnout, and ballast failure, so the number of working lamps is dynamically decreasing. Consequently, the performance of energy saving, financial payback, and carbon emission deteriorates over time. In many projects, maintenance is essentially required in the contract, so that the failed facilities can be repaired or replaced to overcome the deterioration in EE performance. In case of both incremental retrofit and maintenance, the population decay model should be re-formulated to solve practical problems in EE applications, such as lighting retrofit, measurement and verification [20–22], energy reliability [23–25], and distributed generation [26]. Such kinds of decay become more complicated than natural decay in the following three respects.

First, there is an aggregate population of installed lights with different working time. For example, the population of installed lights in the first year has a decay curve different from that of the population of installed lights in the second and subsequent years. Second, in the case of both incremental retrofit and maintenance, interaction of multiple variables is involved in the decay model. The new intervention of maintenance has brought new characteristics to the decay model. Maintenance will change the average EE performance and the average working time, and consequently the decay curve will become non-singular. Third, the parameters of the decay model are usually unknown, although they can be estimated through additional tests on a population of similar lamps.

The other issue in the time dimension is making long-term plans for incremental retrofit and maintenance even if the decay model is known. The retrofit plans have been intensively studied by using empirical [27] and multi-criteria methods [28, 29]. In approaches to optimisation, several conflicting objectives, such as EE, financial payback, carbon emission, and other technical, economic, ecological, social, aesthetical concerns, have been optimised in the design of retrofit plans. For stable performance improvement, maintenance plans for lighting and other building facilities have become a recent focus in this research area. For lighting retrofit projects, optimal maintenance planning is proposed to optimise the number of lights to replace the failed lamps, so that the EE lighting project achieves sustainable performance in terms of maximal energy savings and the cost-benefit ratio [30]. For general BEER projects, corrective maintenance planning for building energy systems, such as lighting, monitoring, water heating, and oven, is proposed to design optimal maintenance plans for maximising energy saving and minimising the internal rate of return [31, 32].

However, current planning schemes in the literature cannot be extended in retrofit projects with multiple interventions, as they have neglected the interaction between incremental retrofit and maintenance. To the best of our knowledge, there are few studies of combined planning for incremental retrofit and maintenance, especially in the large-scale BEER projects. As the parameters in the decay model are unknown, the parameters should be estimated at the early stage of implementation process, which brings extra challenges for planning. Therefore, in this paper, adaptive control is studied for the planning problem of lighting projects with unknown parameters. The stability of adaptive control can ensure stable EE performance of these retrofit projects. Due to the closed-loop mechanism, uncertainty about lamp decay or lumen degradation can be attenuated.

The contributions of this paper are three-fold. First, a mathematical model is built for lighting projects with incremental retrofit and maintenance. The decay of the lamp population has a logistic-like curve that is related to the number of retrofitted lamps. The proposed model is verified based on the empirical data and the interactive dynamics can also be observed in the verification. Second, the decay process is studied by using a control approach, in which a multiple-input and multiple-output (MIMO) control system is derived, based on the proposed decay model. Third, to handle the unknown parameters in the system model, adaptive control laws are designed for planning incremental retrofit and maintenance. The stability of the proposed control laws is proven with the Lyapunov theory, which ensures that the EE performance can be sustained at a desired value.

The paper is organised as follows. Several models of natural lighting decay are introduced in Section 2. In Section 3, the decay of an aggregate population with incremental retrofit and maintenance is modelled, and the MIMO state equation is formulated. In Section 4, adaptive control laws are newly designed, and the stability of the adaptive controller is proved. In Section 5, the model is verified, and the adaptive control is tested and analysed in the simulation. Then paper is concluded in Section 6.

2 Models of natural decay

System dynamics in lighting projects, such as the CFL project and the LED project, is caused by the performance decay of the lamps. The performance decay or deterioration of working lamps affects the energy saving, financial profit, and carbon emission over the evaluation period. An exact model of the population decay is necessary to reflect the system dynamics. Irrespective of the types of light used, three kinds of decay model are commonly applied.

Let $N(t)$ denote the population size at the tth year. $N(0)$ is the size of initial population. In many natural phenomena, such as population growth and radioactive decay, quantities grow or decay at a rate proportional to their size. In other words, they satisfy the following differential equation:

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = kN(t), \tag{1}$$

where k is the decay rate. Note that (1) is called the law of natural growth if $k > 0$, and it is called the law of natural decay if $k < 0$.

The only solution of (1) is an exponential function

$$N(t) = N(0)e^{kt}. \tag{2}$$

Remark 1: Equation (2) satisfied the law of (1) as

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = N(0)(e^{kt})' = kN(0)e^{kt} = kN(t). \tag{3}$$

The exponential decay model is commonly used in different areas [33]. Normally a constant decay rate (failure rate) applies to this model, but in certain cases, the decay rate changes over time. In this lighting application, the lamp population exhibits ageing, so that old lamps are more likely to fail at any time than newly installed lamps.

As the second kind of decay model, a linear population decay, suggested in CDM guidelines, is utilised in the lighting projects [18]. In the linear model, the population is linearly decayed over the rated lifetime L as

$$N(t) = \begin{cases} N(0)\left(1 - \frac{t}{L} * \frac{100 - \rho_L}{100}\right), & t \le L \\ 0, & t > L \end{cases}, \tag{4}$$

where ρ_L is the percentage of surviving lamps left at the end of the rated lifetime L ($\rho_L = 50$ is recommended) [18]. When $t > L$, all lamps are deemed to have failed in this model.

In empirical studies on the useful life of facilities in retrofit projects [34], the decay curves are found to have logistic shapes. The Poland efficient lighting project (PELP), conducted by the World Bank through the International Finance Corporation, also indicates a logistic curve for a population of 1.2 million lamps [35]. According to studies in the South Africa context [36, 37], a general form of logistic function is formulated to fit empirical data. As the third kind of decay model, the general form is expressed as

$$N(t) = \frac{N(0)}{\gamma + e^{\beta t - K}} \tag{5}$$

where β and γ are two parameters related to characteristics of the device, and $K = -\ln(1 - r)$. As stated, this general form of the decay model is especially applicable to the engineering context.

According to the three models, the decay dynamics can be obtained for each model as

$$\frac{\mathrm{d}\Phi_1}{\mathrm{d}t} = k\Phi_1 \tag{6}$$

$$\frac{\mathrm{d}\Phi_2}{\mathrm{d}t} = -\frac{100 - \rho_L}{100L}, \quad t < L \tag{7}$$

$$\frac{\mathrm{d}\Phi_3}{\mathrm{d}t} = -\beta\Phi_3(1 - \gamma\Phi_3) \tag{8}$$

where $\Phi = N(t)/N(0)$ is the proportion of working lamps surviving at time t. Φ_1, Φ_2, Φ_3 denote the proportion calculated in each model, respectively. Note that (8) is deduced from the differentiation at both sides of (5). For each model, an example of the decay curve is plotted in Fig. 1.

In this figure, decay parameters are set as $k = -1$, $L = 5$, $K = 3$, $\beta = 0.95$, and $\gamma = 1.05$. Note that these models have been used to approximate the decay of CFL and LED. These models only fit essential factors of natural decay without any intervention, so these decay curves are all non-increasing as shown in the figure.

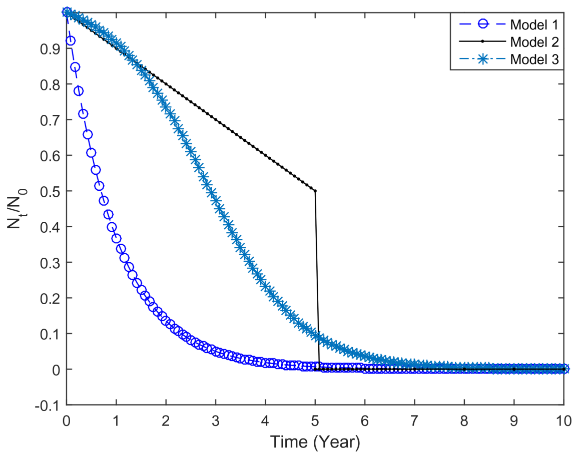


Fig. 1 *Decay curves in the three decay models ($k = -1$, $L = 5$, $K = 3$, $\beta = 0.95$ and $\gamma = 1.05$)*

3 System dynamics with multiple interventions

When only considering natural decay in the lighting retrofit project, system dynamics can be generalised as

$$\dot{\Phi} = f(\Phi), \tag{9}$$

where $f(\Phi)$ is the decay function. With respect to model 3, it follows that $f(\Phi) = -\beta\Phi(1 - \gamma\Phi)$.

In the control approach, the number of working lamps is regarded as a state variable, i.e. $x(t) = N(t)$. Based on the logistic decay models (5) and (8), the state equation can be expressed as

$$\dot{x} = -\beta x + \frac{\beta\gamma}{x(0)}x^2, \tag{10}$$

where x is the number of working lamps, and $x(0)$ is the size of initial population. Note that the control system studied is a non-linear system.

Remark 2: The state equation (10) is obtained by the differentiation at both sides of (5) as

$$\begin{aligned} \dot{x} &= x(0)\left(\frac{1}{\gamma + e^{\beta t - K}}\right)' \\ &= -x(0)\frac{\beta e^{\beta t - K}}{(\gamma + e^{\beta t - K})^2} \end{aligned}. \tag{11}$$

By substituting $e^{\beta t - K} = x(0)/x - \gamma$ in the above equation, (10) can be obtained.

The dynamics of natural decay cannot fit the practical decay dynamics well when there are multiple interventions, which include incremental retrofit and maintenance. Incremental retrofit means that retrofit does not only happen at the beginning, but also happens subsequently at multiple times. Compared with once-off retrofit, the number of retrofitted facilities in this case is incremental over time, so we call this incremental retrofit. To ensure stable performance of EE and cost saving, the broken or ill-conditioned facilities should be repaired or replaced. Maintenance means replacement or repair of retrofitted lamps that have deteriorated, and maintenance will be conducted frequently every year. It is obvious that both incremental retrofit and maintenance have certain effects on the decay dynamics (10).

However, (10) cannot be applied, when retrofit and maintenance are interacted during the whole evaluation period. At time t, a new EE facility could be used for maintenance of a retrofitted facility in a poor condition, and it could also be used for retrofitting an existing old facility. In other words, incremental retrofit will increase the population size and the number of working lamps, but maintenance only increases the number of working lamps. When different effects of incremental retrofit and maintenance are included into (10), the number of working lamps with incremental retrofit and maintenance can be expressed as

$$\dot{x} = -\beta x + \frac{\beta\gamma x^2}{x(0) + \int_{\tau=0}^{t} u_1 \,\mathrm{d}\tau} + u_1 + u_2, \tag{12}$$

where $u_1(t)$ is the number of retrofitted lamps at time t, and $u_2(t)$ is the number of lamps undergoing maintenance. Note that when there is no intervention, i.e. $u_1(t) = u_2(t) = 0$, (12) is equivalent with (10). When only maintenance is done, i.e. $u_1(t) = 0$, the population size remains the same as $x(0)$, and the number of working lamps increases by $u_2(t)$. When only retrofit is done, i.e. $u_2(t) = 0$, the population size increases by the cumulative number of retrofitted lamps, i.e. $\int_{\tau=0}^{t} u_1 \,\mathrm{d}\tau$, and the number of working lamps also increases by $u_1(t)$.

Define $x_1 = x(0) + \int_{\tau=0}^{t} u_1 \,\mathrm{d}\tau$ and $x_2 = x$. The state equation 12 can be transformed as

$$\begin{cases} \dot{x}_1 = u_1, \\ \dot{x}_2 = -\beta x_2 + \dfrac{\beta\gamma x_2^2}{x_1} + u_1 + u_2, \end{cases} \tag{13}$$

In the control approach, the system dynamics with incremental retrofit and maintenance is a standard non-linear system. Actually, $x_1(t)$ is the cumulative number of retrofitted lamps at time t, and $x_2(t)$ is the number of working lamps at time t. Note that $x_1(t) \geq x_2(t)$ is a practical constraint.

Given a sampling period t_0, the continuous control system can be written into a discrete form. The discrete system can be formulated as (14), where k represents the index of sample and $x_1(0) \geq x_2(0)$. $u_1(k)$ is the number of retrofitted lamps over the kth interval. $u_2(k)$ is the number of lamps undergoing maintenance over the kth interval. $x_1(k)$ is the cumulative number of retrofitted lamps at the kth interval, and $x_2(k)$ is the number of working lamps at the kth interval

$$\begin{cases} x_1(k+1) = x_1(k) + u_1(k)t_0, \\ x_2(k+1) = (1 - \beta t_0)x_2(k) + \beta\gamma t_0\dfrac{x_2(k)^2}{x_1(k)} + u_1(k)t_0 + u_2(k)t_0, \end{cases} \tag{14}$$

The average working time of the population is defined as the working time of all lamps divided by the population size. The average working time is related to the percentage of working lamps and the working hours of each lamp. The average working time can indicate the lumen level, which means that a lamp with more working hours will be subject to more lumen deprecation.

Theorem 1: Given a population of energy efficient lamps, maintenance will result in a shorter average working time than incremental retrofit.

Proof: Assume that the average working hours of the population at time t is $\overline{L}_t$, and the population size of retrofitted lamps is $x_1(t)$. If only n lamps is maintained, the average working hours after maintenance can be calculated as $(\overline{L}_t(x_1(t) - n))/x_1(t)$.

If only n lamps is retrofitted, the average working hours after the incremental retrofit is calculated as $(\overline{L}_t x_1(t))/(x_1(t) + n)$.

It is obvious that

$$\overline{L}_t(x_1^2(t) - n^2) < \overline{L}_t x_1^2(t). \tag{15}$$

Divide both sides with $x_1(x_1(t) + n)$, then

$$\frac{\overline{L}_t(x_1(t) - n)}{x_1(t)} < \frac{\overline{L}_t x_1(t)}{x_1(t) + n}. \tag{16}$$

The proof is completed. □

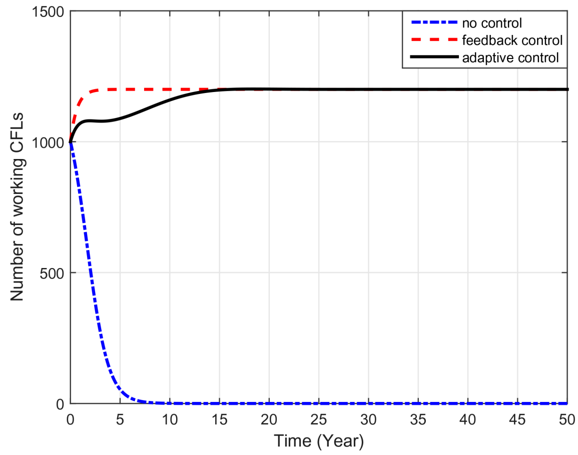


Fig. 2 *Performance of controller design ($k_1 = 1.5$ and $k_2 = 0.5$)*

Table 1 PELP empirical data on surviving rates

Year	1	2	3	4	5	6	7	8	9	10	11
Surviving rate	0.97	0.97	0.91	0.83	0.77	0.4	0.29	0.08	0.02	0.02	0.02

4 Adaptive control

The controller is necessary to keep the number of working lamps in the lighting retrofit, so the performance of EE and cost saving can be stable. As shown in Fig. 2, the number of working lamps will become 0, if there is no controller.

Assume β and γ are known, a feedback controller is required to achieve stable states, so that

$$\lim_{t \to \infty} x_1(t) = r_1, \ \lim_{t \to \infty} x_2(t) = r_2, \tag{17}$$

where r_1 is the reference value of the population size, and r_2 is the reference value of the working lamps. Note that $r_1 \geq r_2$ holds.

Define the tracking error as

$$e_1 = x_1 - r_1, \tag{18}$$

$$e_2 = x_2 - r_2. \tag{19}$$

Take the derivative of e_1 and e_2. It yields

$$\dot{e}_1 = \dot{x}_1 = u_1, \tag{20}$$

$$\begin{aligned}\dot{e}_2 = \dot{x}_2 &= -\beta x_2 + \beta\gamma\frac{x_2^2}{x_1} + u_1 + u_2 \\ &= \phi[-\beta, \beta\gamma]^{\mathrm{T}} + u_1 + u_2,\end{aligned} \tag{21}$$

where $\phi = [x_2, (x_2^2/x_1)]$ is the composite function used for simplicity.

To cancel the non-linear items in (21), a feedback controller is straightforwardly designed as

$$u_1 = -k_1 e_1, \tag{22}$$

$$u_2 = -k_2 e_2 - u_1 - \phi p, \tag{23}$$

where $k_1 > 0$ and $k_2 > 0$ are the control gains. p is the parameter vector to be determined in the controller design. In the assumption of known parameters, $p = [-\beta, \beta\gamma]^{\mathrm{T}}$ can be used to achieve stable control. In other words, the number of working lamps can be kept as the required value with the above scheme according to Theorem 2.

Theorem 2: Assume β and γ are known, the closed-loop system under the feedback controller (22) and (23) with $p = [-\beta, \beta\gamma]^{\mathrm{T}}$ is Lyapunov stable.

The proof has been given in the Appendix. As shown in Fig. 2, the feedback controller can drive the number of working lamps towards the reference value (set as 1200 for illustration). However, the parameters of the model must be known in the feedback control.

If β and γ are unknown in practical applications, it is difficult to determine proper values of p [38]. In this situation, it is necessary to design an adaptive controller for ensuring stable EE performance.

For the adaptive control, the following control scheme is proposed:

$$u_1 = -k_1 e_1 \tag{24}$$

$$u_2 = -k_2 e_2 - u_1 - \phi\hat{p} \tag{25}$$

with the adaptive law for $\hat{p}$ given by

$$\dot{\hat{p}} = \eta\phi^{T} e_2, \tag{26}$$

where $\eta > 0$ is the updating rate, and $\hat{p}$ is the estimate value of p. When the parameters in the decay model are unknown, the proposed adaptive control scheme can also keep the number of working lamps as the required value according to the following theorem.

Theorem 3: If the adaptive controller (24) and (25) with the adaptive law (27) is used, then it is ensured that the tracking error turns to zero as $t \to \infty$

The proof has been given in the Appendix. As shown in Fig. 2, the adaptive controller can also drive the number of working lamps to the reference value, although the parameters of the model are unknown.

For a discrete form, set $\hat{p}(0) = 0$, and the adaptive law for $\hat{p}$ can be expressed as

$$\hat{p}(t+1) = \hat{p}(t) + \eta\phi^{\mathrm{T}}(t)e_2(t)t_0. \tag{27}$$

The adaptive controller can be expressed as

$$u_1(t) = -k_1 e_1(t), \tag{28}$$

$$u_2(t) = -k_2 e_2(t) - u_1(t) - \phi(t)\hat{p}(t). \tag{29}$$

5 Simulation verification

A lighting retrofit project for retrofitting 1500 incandescent lamps is evaluated. After initial retrofit, the population size of retrofitted CFLs is 1000. For each incandescent lamp, the rated power is 60 W. For each CFL, the rated power is 14 W. Based on empirical CFL data in Table 1, parameters in the proposed model are assumed known as $\beta = 0.921$ and $\gamma = 0.986$ (reported in [19]) in the first two case studies. In real applications, these two parameters are usually unknown. In case 3, β and γ are unknown constants, and the performance of adaptive control will be analysed.

As reported in [30], LED decay also follows the logistic curve like that of CFL decay, so observations on the CFL project could be expected to apply to the LED project too. For simplicity, simulation verification on a LED project is omitted here.

5.1 Case 1: Comparison of different models

The proposed model is compared with existing models referred to in Section 2. As known, model 1 is the natural decay model; model 2 is the linear decay model; and model 3 is the logistic decay model. The main characteristic of the proposed model is its ability to reflect different effects of incremental retrofit and maintenance. However, models 1, 2, and 3 cannot be directly applied to describe decay dynamics in such interventions. For fair comparison, the CFL population of incremental retrofit or maintenance (at fourth year) is regarded as an independent population in the three existing models. Then decay curves of three existing models can be plotted as shown in Fig. 3. In the first two tests, 200 CFLs are used for

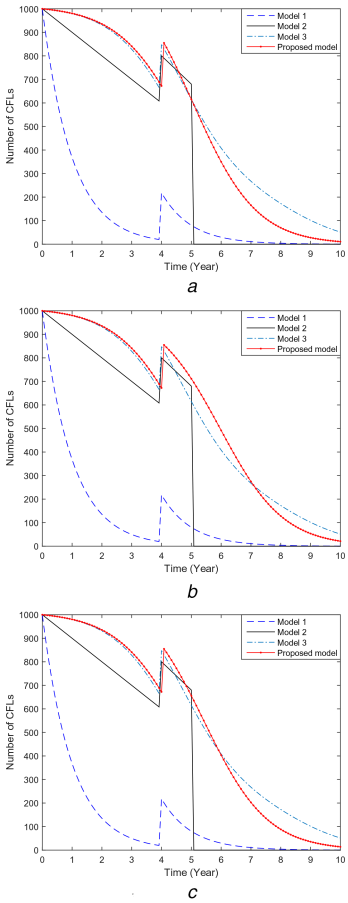


Fig. 3 *Model comparisons*
(a) Decay curves after the fourth-year retrofit, ***(b)*** Decay curves after the fourth-year maintenance, ***(c)*** Decay curves after combined retrofit and maintenance

retrofit and maintenance, respectively. The decay curves are plotted in Figs. 3*a* and *b*.

In comparison, another test is conducted where 100 CFLs are used for retrofit and 100 CFLs are used for maintenance, as shown in Fig. 3*c*. It can be noticed that decay curves of models 1, 2, and 3 are the same in the three tests, but decay curves of the proposed model are different. The reason is the fact that the decay dynamics change as incremental retrofit and maintenance are done. These changes are not considered in existing models, but considered in the proposed model. The detailed effects of incremental retrofit and maintenance will be analysed in the following case studies.

5.2 Case 2: Comparison of intervention factors

In the second case, the effects of the intervention time are first evaluated. Hundred new CFLs are used to replace the broken CFLs for maintenance at the fourth, sixth, and eighth year, respectively. The decay curves are plotted in Fig. 4*a*. It can be noticed that the maintenance in the early years shows better results than in the subsequent years. The maintenance at the fourth year results in the slowest decay.

In comparison, 100 CFLs are used to replace another 100 incandescent lamps for incremental retrofit at the fourth, sixth, and eighth year, respectively. As shown in Fig. 4*b*, the same conclusion can be drawn that the incremental retrofit in the early years shows better performance than in the latter years. However, the time effects in the incremental retrofit are not as significant as those in the maintenance. In the case of incremental retrofit, the decay curves of the fourth and sixth years are overlapped after the sixth year. One possible reason is the fact that the population of the incremental retrofit is larger, so that the average effect is less.

Two kinds of interventions, i.e. maintenance and incremental retrofit, are evaluated in the proposed model. In the fourth year, 100 CFLs are used to replace the broken CFLs for maintenance. The decay curve is shown as 'M' in Fig. 5*a*. In comparison, 100

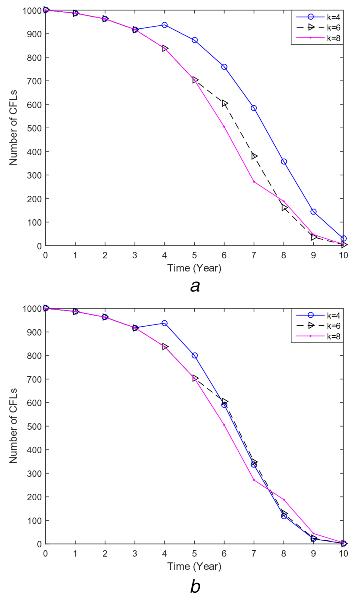


Fig. 4 *Decay curves with intervention at the fourth, sixth, and eighth years, respectively*
(a) Maintenance, ***(b)*** Incremental retrofit

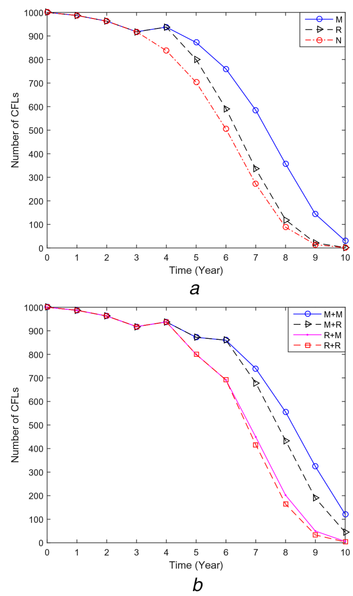


Fig. 5 *Decay curves of different interventions*
(a) Separate intervention, ***(b)*** Multiple interventions

new CFLs are used to replace another 100 incandescent lamps for incremental retrofit at the same time. The decay curve is shown as 'R' in the figure. The natural decay without any intervention is shown as 'N' in the figure. It can be noticed that interventions can postpone the decay process, as the post-intervention population of CFLs has a slower decay rate than the pre-intervention population. It can also be observed that the decay rate of maintenance is slower than that of incremental retrofit as shown in the figure, and that the overall population after maintenance has fewer average working hours than the population after incremental retrofit, which matches the statement of Theorem 1.

Furthermore, multiple interventions are also evaluated in this case study. At the fourth and seventh years, maintenance and incremental retrofit could be chosen by decision makers. The decay curves of different combinations are plotted in Fig. 5*b*. In the figure, '$M + M$' means that maintenance is conducted at the fourth and seventh year, respectively; '$M + R$' means maintenance is conducted at the fourth year and incremental retrofit is conducted at the seventh year; '$R + M$' means that incremental retrofit is conducted at the fourth year and maintenance is conducted at the seventh year; '$R + R$' means that incremental retrofit is conducted

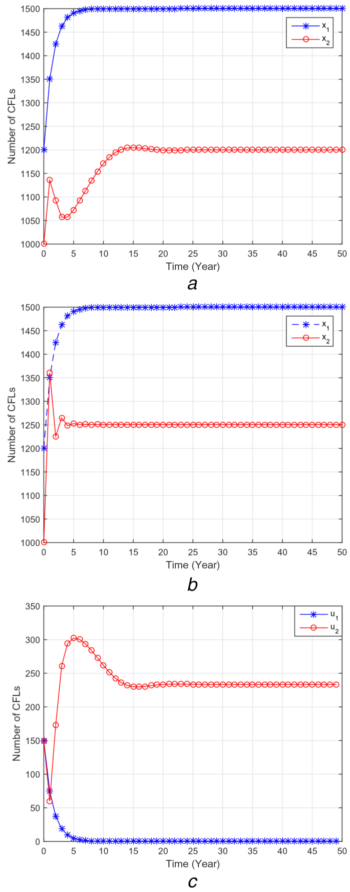


Fig. 6 *Comparison of adaptive control and feedback control*
(a) State profiles in the adaptive control, ***(b)*** State profiles in the feedback control, ***(c)*** Input variables in the adaptive control

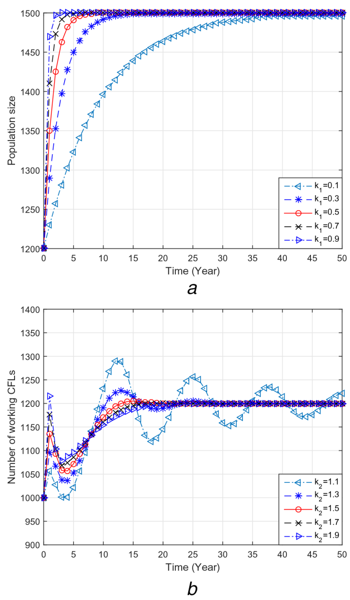


Fig. 7 *Effects of the control gains*
(a) Population size under different k_1, ***(b)*** Number of working CFLs under different k_2

at the fourth and seventh years, respectively. According to Theorem 1, the same observation can be made that the decay of '$M + M$' has the slowest rate, and that the decay of '$R + R$' has the fastest rate.

5.3 Case 3: Performance of adaptive control

In the adaptive controller, the control gains are set as $k_1 = 0.5$ and $k_2 = 1.5$, and the updating rate is $\eta = 1 * 10^{-7}$. The reference values

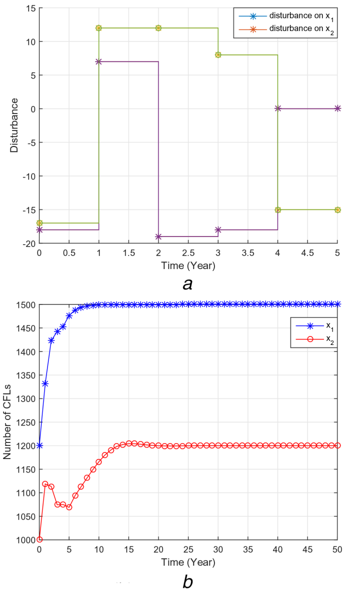


Fig. 8 *Effects of the state uncertainty*
(a) Profiles of the state disturbance, ***(b)*** Profiles of the state variables

of x_1 and x_2 are $r_1 = 1500$ and $r_2 = 1200$, respectively. In other words, the population size of CFLs is expected to be 1500, and the number of working CFLs is expected to be 1200. The control laws $u_1(t)$ and $u_2(t)$, i.e. retrofit and maintenance plans, follow (24) and (25) designed in the adaptive control.

For the adaptive control, the profiles of state variables x_1 and x_2 are plotted in Fig. 6*a*. In the adaptive control, the steady-state errors converge to 0 at finite time. As shown in the figure, it is indicated that $x_1(9) = 1500$ and $x_2(20) = 1200$. In comparison, the state profiles of feedback control, in which $p = [-0.9, 0.81]$, are also given in Fig. 6*b*. It can be noticed that the steady-state error is present in the feedback control. For the adaptive control, the profiles of input variables are plotted in Fig. 6*c*. It can be observed that the maintenance has a constant value and no retrofit is required when $t \geq 20$.

The parameters in the controller are converging to $p_1 = -0.1210$ and $p_2 = -0.0919$. With respect to EE, energy saving is related with the number of working CFLs and daily burning hours. If the average daily burning hour is 5 h, energy saving in the first year is 83,950 kWh. Energy saving in the first 5, 10, and 20 years is 448,380, 915,730, and 1,920,700 kWh, respectively. After 20 years, annual energy saving is constant at 100,740 kWh.

The robustness of control gains is also evaluated in this study. When k_2 and η are kept unchanged, k_1 is set at 0.1, 0.3, 0.5, 0.7, and 0.9, respectively. With these different settings, the profiles of the CFL population size are plotted in Fig. 7*a*. It can be observed that all profiles converge to the reference value, which indicates the robustness of k_1. A large value of k_1 causes x_1 to converge rapidly. When k_1 and η are kept unchanged, k_2 is set at 1.1, 1.3, 1.5, 1.7, and 1.9, respectively. With these different settings, the profiles of working CFLs have been plotted in Fig. 7*b*. It can be observed that the profiles converge to the reference value, which indicates the robustness of k_2. However, a small k_2 (e.g. $k_2 < 1.1$) causes x_2 to converge slowly with some oscillation.

The adaptive control is also tested in a case with state uncertainty. Assume that state variables experience disturbance during the first 5 years, and the disturbance values are random numbers on the scale $[-20, 20]$, as shown in Fig. 8*a*. As a result, state profiles can also converge to reference ones, as shown in Fig. 8*b*. Therefore, it can be concluded that the designed adaptive controller is stable to reject some uncertainty.

6 Conclusion

As an example of BEER, the lighting retrofit project is studied. In consideration of incremental retrofit and maintenance, a new decay model is proposed for the lighting retrofit project. Based on the characteristics of natural decay, the population decay with multiple interventions is formulated in the proposed model. In the control approach, a MIMO state equation is formulated to express the interactive dynamics based on the proposed decay model. Retrofit and maintenance plans are studied to stabilise the number of working lamps and the size of the overall population. To cope with unknown parameters of the system, an adaptive control approach is proposed to design stable plans. The stability is proven theoretically, and is tested in simulations.

Several observations were made in this study. First, maintenance could contribute more to conquer performance decay than incremental retrofit. Second, the early intervention (maintenance or retrofit) was preferred to postpone performance decay. Third, the adaptive control was robust to deliver stable EE performance. This work is challenging and important in the field of energy system and reliability. In future, stochastic models could be studied for the economic analysis, efforts on novel LED models, and control methods could also be made with regard to emergence of LED.

7 References

[1] Keane, A., Tuohy, A., Meibom, P., *et al.*: 'Demand side resource operation on the Irish power system with high wind power penetration', *Energy Policy*, 2011, **39**, pp. 2925–2934

[2] Behrangrad, M.: 'A review of demand side management business models in the electricity market', *Renew. Sustain. Energy Rev.*, 2015, **47**, pp. 270–283

[3] Wu, Z., Tazvinga, H., Xia, X.: 'Demand side management of photovoltaic-battery hybrid system', *Appl. Energy*, 2015, **148**, pp. 294–304

[4] Wang, B., Xia, X., Zhang, J.: 'A multi-objective optimization model for the life-cycle cost analysis and retrofitting planning of buildings', *Energy Build.*, 2014, **77**, pp. 227–235

[5] Mahlia, T., Said, M., Masjuki, H., *et al.*: 'Cost-benefit analysis and emission reduction of lighting retrofits in residential sector', *Energy Build.*, 2005, **37**, pp. 573–578

[6] Mahlia, T., Razak, H.A., Nursahida, M.: 'Life cycle cost analysis and payback period of lighting retrofit at the University of Malaya', *Renew. Sustain. Energy Rev.*, 2011, **15**, pp. 1125–1132

[7] Ye, X., Xia, X.: 'Optimal metering plan for measurement and verification on a lighting case study', *Energy*, 2016, **95**, pp. 580–592

[8] Michaelowa, A., Jotzo, F.: 'Transaction costs, institutional rigidities and the size of the clean development mechanism', *Energy Policy*, 2005, **33**, pp. 511–523

[9] Mundaca, L.: 'Transaction costs of tradable white certificate schemes: The energy efficiency commitment as case study', *Energy Policy*, 2007, **35**, pp. 4340–4354

[10] Mozzo, M.A.: 'Measurement and verification of savings in performance contracting', *Energy Eng.*, 1999, **96**, pp. 33–45

[11] Mei, J., Xia, X.: 'Energy-efficient predictive control of indoor thermal comfort and air quality in a direct expansion air conditioning system', *Appl. Energy*, 2017, **195**, pp. 439–452

[12] Fan, Y., Xia, X.: 'A multi-objective optimization model for energy-efficiency building envelope retrofitting plan with rooftop PV system installation and maintenance', *Appl. Energy*, 2017, **189**, pp. 327–335

[13] Wu, Z., Wang, B., Xia, X.: 'Large-scale building energy efficiency retrofit: Concept, model and control', *Energy*, 2016, **92**, pp. 1–13

[14] Yin, C., Dadras, S., Huang, X., *et al.*: 'Energy-saving control strategy for lighting system based on multivariate extremum seeking with newton algorithm', *Energy Convers. Manage.*, 2017, **142**, pp. 504–522

[15] Yin, C., Stark, B., Chen, Y., *et al.*: 'Fractional-order adaptive minimum energy cognitive lighting control strategy for the hybrid lighting system', *Energy Build.*, 2015, **87**, pp. 176–184

[16] Dong, J., Pandharipande, A., van Driel, W., *et al.*: 'Diagnosing lumen depreciation in LED lighting systems: an estimation approach', *IEEE Trans. Signal Process.*, 2012, **60**, pp. 3796–3808

[17] Royer, M.: 'Lumen maintenance and light loss factors: Consequences of current design practices for LEDs', *LEUKOS*, 2014, **10**, pp. 77–86

[18] Ye, X., Xia, X., Zhang, J.: 'Optimal sampling plan for clean development mechanism lighting projects with lamp population decay', *Appl. Energy*, 2014, **136**, pp. 1184–1192

[19] Carstens, H., Xia, X., Ye, X.: 'Improvements to longitudinal clean development mechanism sampling designs for lighting retrofit projects', *Appl. Energy*, 2014, **126**, pp. 256–265

[20] Heo, Y., Zavala, V.M.: 'Gaussian process modeling for measurement and verification of building energy savings', *Energy Build.*, 2012, **53**, pp. 7–18

[21] Walter, T., Price, P.N., Sohn, M.D.: 'Uncertainty estimation improves energy measurement and verification procedures', *Appl. Energy*, 2014, **130**, pp. 230–236

[22] Xia, X., Zhang, J.: 'Mathematical description for the measurement and verification of energy efficiency improvement', *Appl. Energy*, 2013, **111**, pp. 247–256

[23] Salata, F., de Lieto Vollaro, A., Ferraro, A.: 'An economic perspective on the reliability of lighting systems in building with highly efficient energy: a case study', *Energy Convers. Manage.*, 2014, **84**, pp. 623–632

[24] Li, N., Yang, Z., Becerik-Gerber, B., *et al.*: 'Why is the reliability of building simulation limited as a tool for evaluating energy conservation measures?', *Appl. Energy*, 2015, **159**, pp. 196–205

[25] Stadler, M., Groissb, M., Góck, G.C., *et al.*: 'Optimizing distributed energy resources and building retrofits with the strategic der-camodel', *Appl. Energy*, 2014, **132**, pp. 557–567

[26] Silvestre, M.L.D., Graditi, G., Sanseverino, E.R.: 'A generalized framework for optimal sizing of distributed energy resources in micro-grids using an indicator-based swarm approach', *IEEE Trans. Ind. Inf.*, 2014, **10**, pp. 152–162

[27] Dascalaki, E., Santamouris, M.: 'On the potential of retrofitting scenarios for offices', *Build. Environ.*, 2002, **37**, (6), pp. 557–567

[28] Kaklauskas, A., Zavadskas, E.K., Raslanas, S.: 'Multivariant design and multiple criteria analysis of building refurbishments', *Energy Build.*, 2005, **37**, (4), pp. 361–372

[29] Juan, Y.-K., Kim, J.H., Roper, K., *et al.*: 'GA-based decision support system for housing condition assessment and refurbishment strategies', *Autom. Constr.*, 2009, **18**, pp. 394–401

[30] Ye, X., Xia, X., Zhang, L., *et al.*: 'Optimal maintenance planning for sustainable energy efficiency lighting retrofit projects by a control system approach', *Control Eng. Pract.*, 2015, **37**, pp. 1–10

[31] Wang, B., Xia, X.: 'Optimal maintenance planning for building energy efficiency retrofitting from optimization and control system perspectives', *Energy Build.*, 2015, **96**, pp. 299–308

[32] Wang, B., Wu, Z., Xia, X.: 'A multistate-based control system approach toward optimal maintenance planning', *IEEE Trans. Control Syst. Technol.*, 2017, **25**, pp. 374–381

[33] O'Connor, P., Kleyner, A.: '*Practical reliability engineering*' (John Wiley & Sons, UK, 2011)

[34] 'California energy efficiency evaluation protocols: technical, methodological, and reporting requirements for evaluation professionals'. Tech. Report, California Public Utilities Commission, 2016

[35] Granda, C.: 'Case study: The IFC/GEF poland efficient lighting project (PELP)'. Proc. Right Light, vol. 4, 1997, pp. 1–8

[36] Botha-Moorlach, M., Mckuur, G.: 'A report on the factors that influence the demand and energy savings for compact fluorescent lamp doortodoor rollouts in South Africa'. Tech. Report, Eskom, South Africa, 2009

[37] Carstens, H., Xia, X., Zhang, J., *et al.*: 'Characterising compact fluorescent lamp population decay'. 2013 IEEE AFRICON, 2013, pp. 1–5

[38] Song, Y., Zhao, K., Krstic, M.: 'Adaptive control with exponential regulation in the absence of persistent excitation', *IEEE Trans. Autom. Control*, 2017, **62**, pp. 2589–2596

8 Appendix

8.1 Appendix 1: Proof of Theorem 2

Denote the Lyapunov function as $V(x_1, x_2)$. A Lyapunov function candidate is defined as

$$V = \frac{1}{2}e_1^2 + \frac{1}{2}e_2^2. \tag{30}$$

For $\forall e_1 \neq 0, \forall e_2 \neq 0$, it is obvious that $V > 0$. The derivative function can be deduced as

$$\dot{V} = \dot{e}_1 e_1 + \dot{e}_2 e_2. \tag{31}$$

Substituting (20) and (23) into the above equation, the derivative can be transformed as

$$\dot{V} = -k_1 e_1^2 - k_2 e_2^2 < 0. \tag{32}$$

According to Lyapunov stability theory, the feedback controller is stable. The proof is completed.

8.2 Appendix 2: Proof of Theorem 3

Choosing the Lyapunov function candidate as

$$V = \frac{1}{2}e_1^2 + \frac{1}{2}e_2^2 + \frac{1}{2\eta}\tilde{p}^T\tilde{p} \tag{33}$$

where $\tilde{p} = p - \hat{p}$. The derivative of V w.r.t. time is

$$\begin{aligned} \dot{V} &= e_1\dot{e}_1 + e_2\dot{e}_2 - \frac{1}{\eta}\tilde{p}^T\dot{\hat{p}} \\ &= e_1u_1 + e_2(\phi p + u_1 + u_2) - \frac{1}{\eta}\tilde{p}^T\dot{\hat{p}} \end{aligned} \tag{34}$$

Substituting the control laws (24) and (25) into (34), it can be deduced that

$$\dot{V} = -k_1 e_1^2 - k_2 e_2^2 + e_2 \phi \tilde{p} - \frac{1}{\eta} \tilde{p}^T \dot{\hat{p}} \tag{35}$$

Inserting the adaptive law (27) into (35), it can be deduced that

$$\dot{V} = -k_1 e_1^2 - k_2 e_2^2 \leq 0 \tag{36}$$

which shows that $V(t)$ is globally uniformly ultimately bounded (i.e. $V(t) \in L_\infty$), which implies that $e_1 \in L_\infty$, $e_2 \in L_\infty$, and $\tilde{p} \in L_\infty$, which further implies that $x_1 \in L_\infty$, $x_2 \in L_\infty$, and $\hat{p} \in L_\infty$. From the definition of ϕ, we have $\phi \in L_\infty$. Then from (24) and (25) and (27), it follows that $u_1 \in L_\infty$, $u_2 \in L_\infty$, and $\dot{\hat{p}} \in L_\infty$. From (20) and (21) it is seen that $\dot{e}_1 \in L_\infty$ and $\dot{e}_2 \in L_\infty$. From (36) we have

$$k_1 \int_0^t e_1^2(\tau)\, \mathrm{d}\tau + k_2 \int_0^t e_2^2(\tau)\, \mathrm{d}\tau + V(t) = V(0) \tag{37}$$

which implies that $e_1 \in L_2$ and $e_2 \in L_2$. According to Barbalat Lemma, it shows that $\lim_{t \to \infty} e_1(t) \to 0$ and $\lim_{t \to \infty} e_2(t) \to 0$ as $t \to \infty$. The proof is completed.

Journal of Materials Chemistry A

PAPER

View Article Online
View Journal | View Issue

Cite this: J. Mater. Chem. A, 2018, 6, 17481

Three dimensional modelling of the components in supercapacitors for proper understanding of the contribution of each parameter to the final electrochemical performance†

Farshad Barzegar, ‡*[a] Lijun Zhang,‡[a] Abdulhakeem Bello,[bc] Ncholu Manyala [b] and Xiaohua Xia[a]

Received 21st May 2018
Accepted 16th August 2018

DOI: 10.1039/c8ta04736g

rsc.li/materials-a

Three dimensional (3D) modelling of supercapacitors (SCs) has been investigated for the first time to have a better understanding of and study the effect of each parameter on the final electrochemical results. Based on this model, the resistance of the electrolyte, membrane, current collectors and active materials have effects on the first intersection points on the real axis (x-axis) of the Nyquist plots (equivalent series resistance (ESR)). These results indicate inward shrinking of the cyclic voltammograms (CV) due to a small change in the leakage resistance and resistance of the faradic component of materials, and they also explain the parameters that lead to the deformation of the CV from ideal behaviour. The 3D model was verified with experiments using activated carbon-based SC devices. The experimental results confirmed the 3D model results and suggested that the proposed 3D model is reliable and can be used for the proper design of SC devices.

Introduction

Storage systems with sufficient capacity and highly efficient charge and discharge characteristics are of huge and strategic importance for portable electronics and biomedical applications, as well as for short and medium-term stationary applications. For these purposes, different technologies are being developed including mechanical, thermal, physical, chemical and electrochemical energy storage systems.[1] An advanced solution is to use batteries with high energy densities, however, they suffer from low power densities, short cycle lives, safety risks and poor adaptability with flexible systems.[2] Electrochemical capacitors (ECs), also called supercapacitors, with high power densities, good cycling stabilities, and fast charge–discharge rates are new energy storage devices that have attracted attention in the scientific community.[3–5] Currently, research in the field of supercapacitors is focused on fine-tuning electrodes, electrolytes and material sections to achieve the best performance.[6–8] However, there are no studies on the effect of the resistances of each parameter on the final electrochemical performance, which could help researchers to develop and synthesize the best ECs depending on the usage.

For a better understanding of ECs, we need to understand the full behavior of each component of the EC during charge and discharge. Based on their charge storage mechanism, ECs can be classified as electric double layer capacitors (EDLCs), pseudo-capacitors or redox electrochemical capacitors (RECs) and hybrid electrochemical capacitors.[9] The EDLCs store energy by a charge separation at the electrode–electrolyte interface,[10] while REC materials not only store energy like EDLCs, but also in the appropriate potential window undergo electrochemical faradaic reactions between the electrode materials and ions.[11] Until now most researchers have tried to explain the electrical behavior of pure EDLCs[12] for ECs, however, none of the reports clearly explain the effects of the resistances of each component of ECs and how this reflects in their behavior that leads to the final stored energy. In this article, we study and provide a deep understanding of the electrical behavior of ECs and the effect of each component on the final electrochemical performance. Verification and confirmation of the proposed model was carried out experimentally with activated carbon-based materials and a KOH aqueous electrolyte in the laboratory.

Modelling of the supercapacitor

The electrical behavior of ECs can be described by the lumped-element impedance-based model. The most common models

[a]Electrical, Electronic and Computer Engineering Department, University of Pretoria, Pretoria 0002, South Africa. E-mail: farshadbarzegar@gmail.com

[b]Physics Department, Institute of Applied Materials, SARCHI Chair in Carbon Technology and Materials, University of Pretoria, Pretoria 0028, South Africa

[c]Department of Materials Science and Engineering, African University of Science and Technology, Abuja, Nigeria

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c8ta04736g

‡ The first two authors contributed equally to this work.

for the description of EDLCs are in three categories, namely a RC circuit model, a three branch RC circuit model, and a transmission line model (Fig. 1). The simple RC circuit model, as shown in Fig. 1(a), includes four ideal circuit elements: the R_S element represents series resistance which is due to the presence of an electrolyte and metallic conductors, the L element represents the series inductance that is mostly influenced by the geometry of the connectors and electrodes, C is the ideal behavior of a capacitor which stores energy by charge separation at the electrode–electrolyte interface and the R_{CT} element is present due to the process of charge transfer from the electrode to the electrolyte and leakage current. The loss in energy of EDLCs, during self-discharge, charging and discharging, gives rise to leakage of current resistance caused by equivalent series resistance (ESR).[13,14] It is necessary to extend the simple RC circuit model since it cannot be used to probe the porous nature of electrodes or show the behavior of EDLCs accurately over a frequency range.

The three branch RC circuit presented by L. Zubieta and R. Bonert[15] is shown in Fig. 1(b). It consists of the leakage resistance R_{LK} in parallel to three branches which correspond to different time constants for charge transfer. The first or immediate branch is for the time range of seconds and it consists of a resistance R_i in series with two capacitors: a voltage dependent capacitor C_{i1} and a normal capacitor C_{i0}. The delayed branch, with parameters R_d and C_d, represents time constants within the minutes range while the long term branch, containing R_l and C_l, represents time constants greater than ten minutes. The model shows a suitable connection with experimental results, however, the model has a weakness in that the circuit components lack physical meaning.[16]

The transmission line model is adopted in many reports[17,18] to precisely represent the porous nature of the electrodes in EDLCs as shown in Fig. 1(c). This model includes leakage resistance R_{LK}, solution resistance R_{el}, electrode resistance R_{ed}, inductance L_S which prevails at high frequency, and a resistance R_P that is in parallel with inductance L_P which are observed above resonant frequency.

All of the above mentioned models are incomplete models for actual ECs and cannot be used to examine the resistances of each parameter of ECs (the active material, the electrolyte, the separator *etc.*) individually and their focus is mostly on the EDLC material. Thus, in this paper, we propose a realistic three dimensional model to study and elucidate the resistances of each component of ECs individually and the influence of each component on the final behavior of ECs. Cyclic voltammetry (CV) and galvanostatic charge–discharge (GCD) modeling and measurements cannot be completely understood in the electrochemical reaction at the electrode–electrolyte interface but a complete description can be explained using electrochemical impedance spectroscopy (EIS).[19] EIS is often represented by a Nyquist plot which presents the real (x-axis) and imaginary parts of the impedance (y-axis). The Nyquist plot takes into account the different parameters (resistance, capacitance, inductance *etc.*) that are all dependent on the frequency and it is usually divided into three regions of low frequency, medium frequency and high-frequency.[20] For an ideal capacitor, the Nyquist plot of impedance is represented by a vertical line which is parallel to the imaginary axis (y axis), which is only correct for liquid mercury electrodes. The impedance of most solid electrodes deviates from purely capacitive behavior.[21]

To offer a realistic model close to the practical situation, the behavior of ECs should be described by a complex network of non-linear inductances, capacitances and resistances. Thus, Fig. S1(a) and (b)† have been proposed for realistic and accurate 2D modelling of the ideal behavior of EDLC and REC material respectively. The ECs shown in Fig. 2 and S1,† depend on several parameters, such as the capacitors that present double layer behavior (C_1 and C_2), the capacitors that present redox electrochemical behavior (C_{F1} and C_{F2}), inductance (L), resistance of the electrolyte (R_e), a current collector and the resistance of the electrode material (R_C), the membrane resistance (R_m), the faradic part of the material resistance (R_f) and the leakage resistance (R_{lk}) (that is dependent on packaging). The suggested RECs model in our simulation is based on behavior that is close to battery materials,[22,23] due to the fact that most oxide materials for RECs show a faradaic phenomenon. In reality, the presence of functional groups at the surface of the electrode materials for EDLCs cannot be overruled, and these

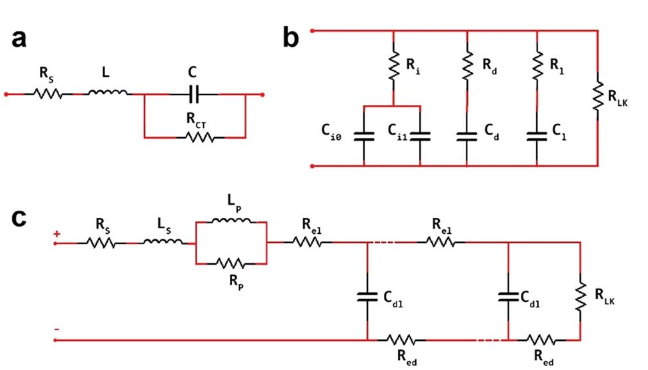


Fig. 1 EDLC: (a) a simple RC circuit model, (b) a three RC circuit model and (c) a transmission line model.

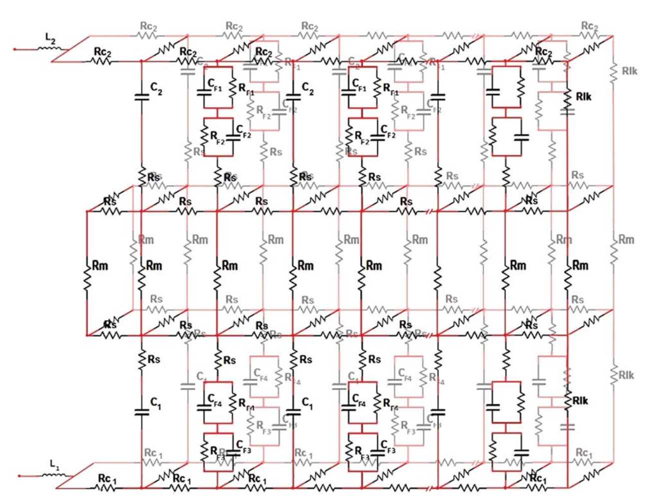


Fig. 2 3D electrical equivalent model of practical ECs.

materials show similar behavior to RECs during charge and discharge, contributing a pseudo-capacitive effect to the total electrochemical performance. Similarly, some oxide materials such as RuO and MnO_2 exhibit (pseudo) EDLC behavior [24] and this makes their models complex. To resolve these issues, the final 2D model presented in Fig. S1(c)† considers a hybrid model that takes care of all the functional groups and other parameters, bringing the model close to the practical EC behavior. Lastly, a 3D hybrid model in Fig. 2 (Fig. S1(d)†) was suggested for a realistic simulation of the behavior of EDLCs.

Methods

Simulation of the full cell supercapacitor

In order to study the performance of the passive hybrid system and to verify the analytical approach above, simulations in Matlab/Simulink were conducted using Simpower GUI. A sawtooth wave with the maximum voltage of 1 V and frequency of 0.01 was used to charge and discharge the cell in order to simulate its performance. Firstly, the 2D model was built in Simulink as shown in Fig. S2.† The two 2D models in Fig. S2† were then connected together through two resistor banks, each comprising R_e and R_C, to form the 3D model shown on the right hand side of Fig. S3.† The left hand side of Fig. S3† shows the charge/discharge control and output voltage and current measurement circuit, where Isc.mat and Vsc.mat store the simulated current and voltage profiles of the cell and Vchg.mat provides the sawtooth voltage waveform of the charge/discharge data given in Fig. S4.†

Experimental and electrode preparation of the full cell supercapacitor

Polymer based activated carbon (AC) was used in the experimental section and was prepared by the methods reported in our previous work[25] using hydrocarbons such as polyvinyl alcohol (PVA)/polyvinylpyrrolidone (PVP) as a source of carbon, through chemical activation with KOH as the active agent to produce the desired porous carbons. The porous carbon (activated carbon (AC)) that was used for the electrochemical tests had a surface area of 1063 m^2 g^{-1} and had some functional groups on the surface as explained in our previous work. The electrodes were made by combining the active materials, the conductive additive (carbon black) and polyvinylidene fluoride (PVDF) in *N*-methyl pyrrolidone (NMP) to make a slurry which was coated on a nickel foam that was graphene coated as a current collector and dried at 60 °C in an oven overnight. The device was tested in a two-electrode configuration with a microfiber filter paper (thickness of 180 μm with 11 μm pore size (particle retention)) as a separator. Three set of devices were made and tested numerous times to ascertain the reproducibility before coming to a conclusion. The reference cell and the first device consisted of the active material (activated carbon with functional groups) derived from PVA\PVP, carbon black and PVDF with weight ratios of 90%, 5% and 5% respectively, 6 M KOH, one glass microfiber filter paper separator and a graphene coated nickel foam current collector. Furthermore, the resistance of the active material (R_C) in the second device was increased by increasing the binder by 10% without the addition of a conductive agent (carbon black). The third device was made with the intention of increasing the resistance of the electrolyte (R_e), where 5 ml of a 10 wt% PVA solution was added to 5 ml of 6 M KOH. The number of the separators was also increased so as to slow the movement of ions within the electrodes which will result in increased membrane resistance (R_m). It is worth stating that it is quite difficult to control all other parameters in reality and thus all the test parameters were repeated numerous times to confirm the results obtained. Electrochemical measurements (electrochemical impedance spectroscopy (EIS) and cyclic voltammetry (CV)) were carried out using a Bio-logic VMP-300 potentiostat. The EIS measurements were conducted in the frequency range from 0.01 Hz to 100 kHz with the open circuit potential of ∼0 V.

Results and discussion

To study the effects of the resistances of each parameter separately and how they reflect in the final electrochemical performance, a reference cell was necessary. After the 3D model of the supercapacitor was designed, each parameter of the cell was assigned a specific value and the outcome of that was considered as the reference cell result. Then, to study the parameter effects individually, each parameter was changed separately while the rest of the parameters were fixed at reference values. To see the negative effect of each parameter compared to the reference cell, some of the parameters were increased and some were decreased. The modified value was just a number that was high enough to see the effect of each parameter compared to the reference cell. Fig. 3 displays the EIS plot, the phase angle *versus* frequency and the CV curves of the simulation results after changing the resistance of each parameter compared to the reference cell. In our simulation results, R_e represents the resistance of the electrolyte, R_m is the resistance of the membrane, R_C is the resistance of the current collector and the electrode materials, R_{lk} is the leakage resistance and R_{CT} is the resistance of the faradic part of the material. It is clear that each parameter had a different effect on the final result of the supercapacitor. For better understanding and clarity of the results, each parameter was plotted and investigated separately compared to the reference cell.

Fig. 4(a) and (b) present the effects of the electrolyte resistance on the final results of the EIS plot and CV curves. As observed from Fig. 4(a) a change in the resistance of the electrolyte (R_e) by a factor of 100 led to an increase in the first intersection points on the real axis (*x*-axis) of the Nyquist plots. This frequency plot corresponds to the typical time constants in most high-power applications, and for intermediate frequencies, the complex-plane plots form an angle of ∼45° with the real axis as seen in the figure. This angle is explained by the limited current penetration into the porous structures of the electrodes.[26] For lower frequencies, the spectra approach a nearly vertical line in the complex plane, which is typical of ideal capacitors. Similarly, a shift and increase in the high-frequency area and the angle of the Nyquist plots for the

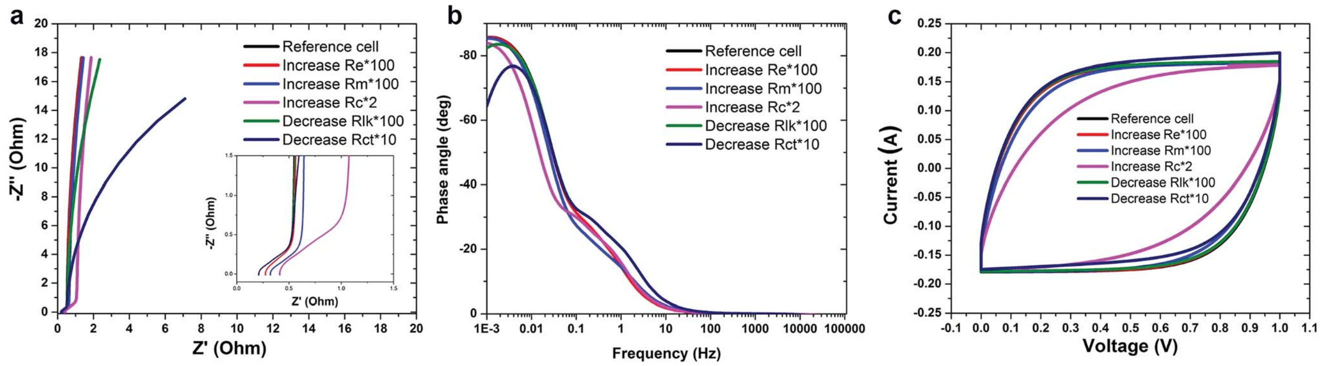


Fig. 3 (a) EIS plot, (b) the phase angle *versus* frequency and (c) CV curves of the simulation results.

intersection of the high and medium-frequency regions was also observed. The model has a highly dynamic load at the beginning as well as deeper charging and discharging. The corresponding CV and the calculated data show excellent agreement. Fig. 4(b) shows the CV curves of the simulated results based on the reference cell and the simulated cell after the R_e was increased by 100 times indicating that there was no change in the shape of the CV, and rather a small decrease in current was observed which corresponded to a slight shrinkage of the CV. This also indicates a decrease in the capacitance of the device.

Several other components contribute to the overall performance of the device, such as the membranes that prevent short-circuiting within the device. Fig. 4(c) and (d) present the effect of the membrane (separator) resistance on the final performance of the device. Fig. 4(c) shows an increase in the resistance of the membrane (R_m) by a factor of 100 and shows that the first intersection points on the real axis (*x*-axis) of the Nyquist plots increased and the whole plots shifted by the same length. The angle of the Nyquist plots for the intersection of the high and medium-frequency regions did not change. The corresponding CV in Fig. 4(d) shows a similar result with a shrunken CV indicating a capacitive decrease. It is clear that the results obtained using the proposed model give similar results to what was obtainable in the experiments that are presented at the end of the paper. Therefore, the proposed electric model can be used in designing a voltage controller and in sizing a supercapacitor for storage applications.

The resistances of the current collectors and active materials (and resistance at the interface of them) also play a crucial part in the performances of the devices. Such contributions are presented in Fig. 5. Fig. 5(a) and (b) show the results of increasing the resistance of the current collector and the electrode materials (R_C) by a factor of 2 (the parameter that was chosen for R_C for the reference cell was high at first, so an increase by a factor of 2 made it high enough to see the effect of R_C compared to the reference cell). From Fig. 5(b) which provides an enlarged view of the Nyquist plot, the remarkable deviations shown were attributed to the longer time taken for charge to reach the surface of the active material. These deviations could also affect the efficiency of the device which is influenced by the resistance of the electrodes which means that increasing the real part of the impedance with a corresponding decrease in frequency has to be taken into consideration. However, the angle of the Nyquist plots for the intersection of the high and medium-frequency regions remained the same. Fig. 5(c) shows a clear shrinkage in the CV that explains a decrease in the performance of the device, hinting that the resistance of the active material and the current collector play a crucial part in the performance of the electrochemical devices. The shrinking might be attributed to a number of factors such as, the conductivity, and the pore dimensions of the active materials and the current collectors.

Fig. 6 presents the effect of leakage resistance (R_{lk}) on the final results. Fig. 6(a) and (b) show that by decreasing the R_{lk} by 100 times, the first intersection points on the real axis (*x*-axis) of

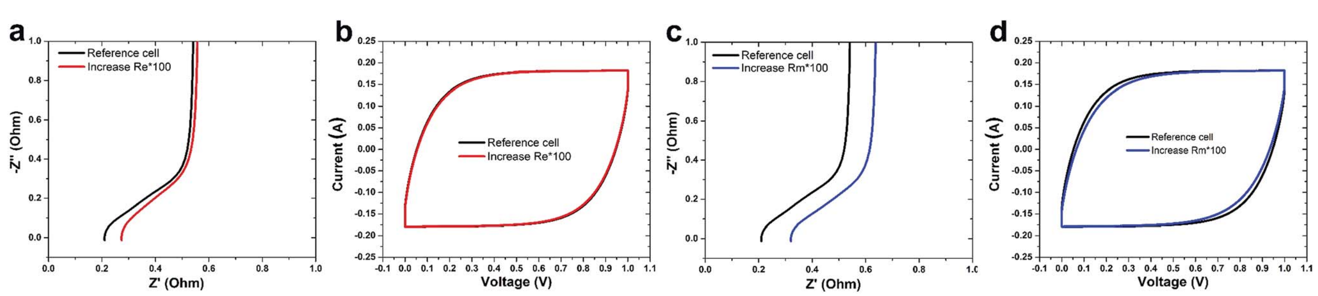


Fig. 4 Simulation results of the EIS plot and CV curves (a) and (b) when increasing the resistance of the electrolyte 100 times, and (c) and (d) when increasing the membrane resistance 100 times.

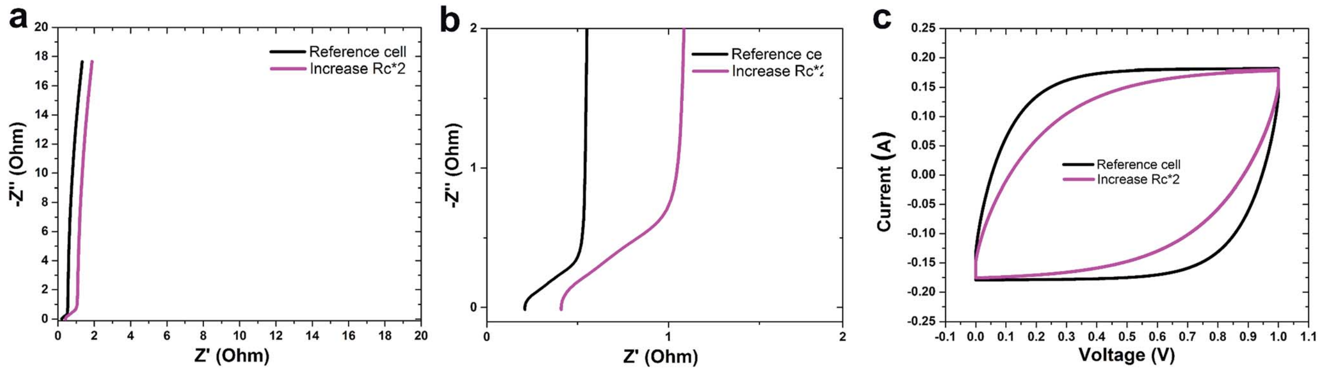


Fig. 5 Simulation result (a) and (b) of the EIS plots and (c) the CV curves when increasing the resistance of the current collectors and active materials 2 fold.

the Nyquist plots did not change. Decreasing the R_{lk} had a negative effect on the capacitance of the supercapacitor. The only part affected was the low-frequency region that shows more deviation from the vertical lines. Decreasing the R_{lk} meant that the cell had an easier way to discharge than keeping the charge at the surface. Fig. 6(c) and (d) show CV curves of the simulation results of the reference cell and the cell after decreasing R_{lk} by 100 times. For the first time, this paper reports an upward shift and a shape change (pushed inward) of the CV curve by adjusting one parameter such as the leakage resistance (Fig. 6) or the resistance of the faradic part of the materials (effect of functional groups) (Fig. 7). The initial and final points moved up and the shape of the CV curve was pushed inward a little. The shift in CV was more at the endpoint with the high current than the first point with the low current which was considered as a shift to match the initial point. As the EIS results showed, decreasing the R_{lk} had a negative effect on the capacitance of the supercapacitor, thus, by decreasing the R_{lk}, the supercapacitor needs more energy to charge than the energy given in the discharge section.

Fig. 7 presents the effect of the resistance of the faradic part of the material (R_{CT}) on the final results. As shown in Fig. 7(a) and (b), a decrease in the R_{CT} by 10 times led to no change in the intercept value on the real axis (*x*-axis) of the Nyquist plots, however, it introduced a negative effect on the faradic part of the performance of the supercapacitor. The only part that was affected was the low-frequency region that became more resistive and deviated from the vertical lines. Fig. 7(c) shows CV curves of the simulation results of the reference cell and the cell after decreasing R_{CT} by 10 times and it shows the same trend as Fig. 6(c) and (d). The initial and final points moved up and the shape of the CV curves was pushed inward a little. A decrease in the R_{CT} reduced the potential capacity of the capacitor that presented redox electrochemical behavior (C_F) and had a negative effect on the capacitance of the supercapacitor so that the EC needed more energy to charge than the energy given in the discharge section. The R_{CT} and C_F had a very close relationship to each other that was linked to the properties of the material.

Fig. 8 presents an enlarged view of the phase angle *versus* frequency of the simulations. This figure shows the effect of each parameter discussed above on the final result of the phase angle. The ideal capacitor phase angle should be -90°. The closer the phase angle is to -90°, the more similarly the device performs to an ideal capacitor.[27] Fig. 8 shows that all the EIS simulations showed a similar trend. Those parameters that provide more resistive behavior that influences the final capacity, move the phase angle far away from -90°.

For the purpose of verifying the analysis above and confirming the proposed model, experiments with an activated carbon-based supercapacitor were tested in the laboratory. In the experimental section, the parameters that it was possible to control physically in our laboratory were investigated. An investigation of the high-frequency region and effect of each resistive component of the device was carried out experimentally to ascertain and confirm the proposed three-dimensional hybrid model. We detected the same trend in Fig. 9 as the 3D

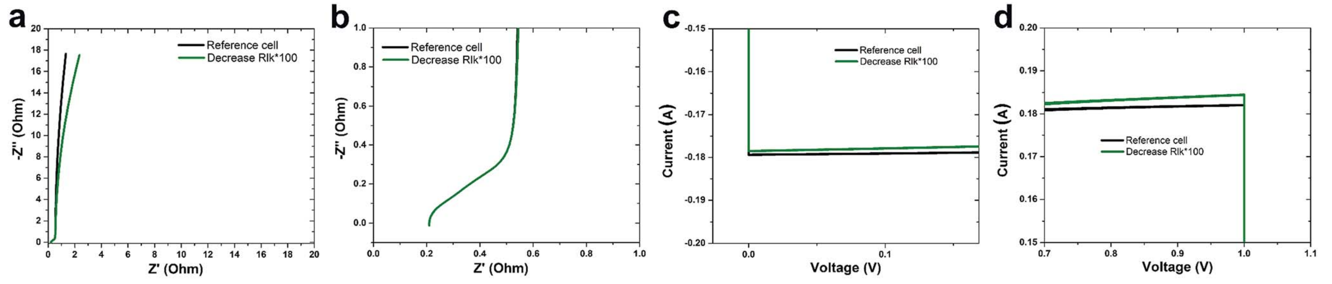


Fig. 6 Simulation results of the EIS plots (a) and (b) and CV curves (c) and (d) when decreasing the leakage resistance 100 fold.

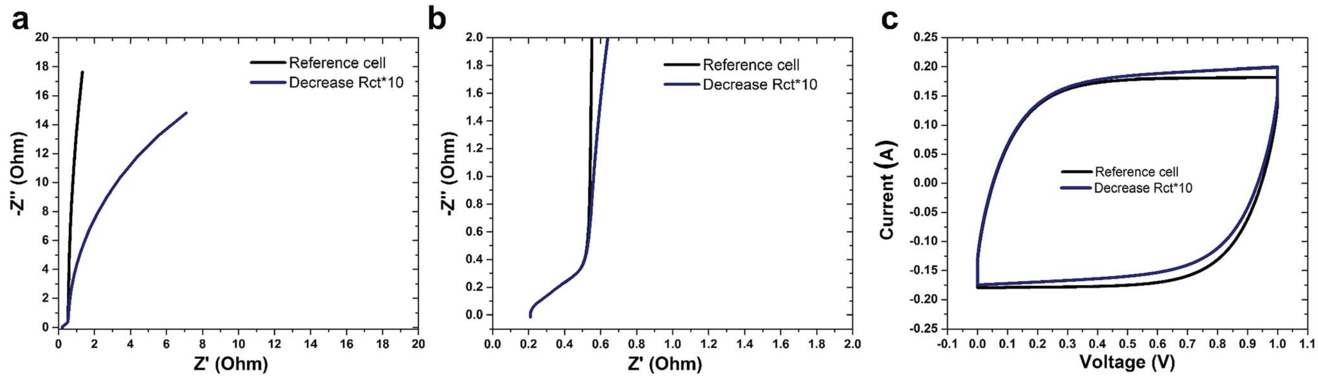


Fig. 7 Simulation result of the EIS plots (a) and (b) and CV curves (c) when decreasing resistance of the faradic part of the material 10 fold.

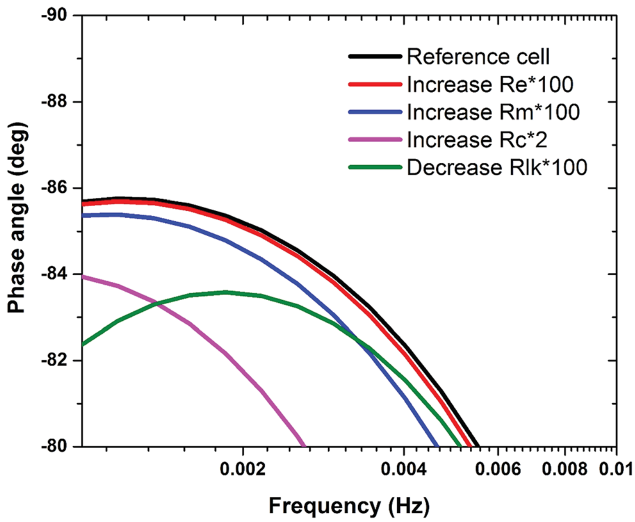


Fig. 8 An enlarged view of the phase angle *versus* frequency of the simulations.

simulation results proposed. However, as shown in Fig. 9(a), the length of the EIS increased after changing one parameter due to the fact that the mass of the material was not as completely uniform as that of the reference cell and the other cell. Controlling the mass at such a scale was very difficult with our equipment. The phase angles (Fig. 9(b) and S5†) and CV curves (Fig. 9(c)) of the experiments also followed the 3D model results that suggest that the proposed model was completely correct. By disregarding the mass effect on the capacity, the CV shape after increasing the R_e was almost the same as that of the reference cell. By increasing the R_m and R_C, clear shrinkage occurred in the CV shape as the 3D model also suggested.

Conclusion

In conclusion, a novel 3D model of a supercapacitor was presented. The results report the effect of each parameter individually for the first time in electrochemical capacitors. Based on the proposed 3D model, the resistance of the electrolyte, the membrane resistance, and the resistance of the current collectors and active materials can increase the first intersection points on the real axis (*x*-axis) of the Nyquist plots. Also, the results revealed a novel phenomenon where the initial and final points of the CV curves shift up and the shape of CV curves is pushed inward a little by changing the leakage resistance and the resistance of the faradaic part of the materials. These results can explain which parameters play a major role in deformation of the CV shape from the ideal state. The experimental results confirmed that the proposed model is completely correct, in that the change of any of the aforementioned parameters indeed increased the first intersection point on the *x*-axis of the

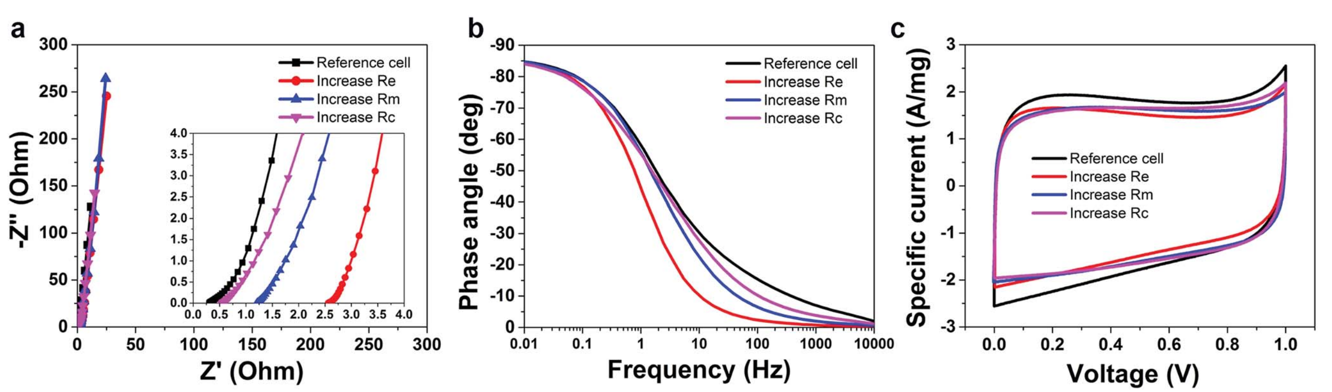


Fig. 9 (a) EIS plot, (b) the phase angle *versus* frequency and (c) CV curves at a scan rate of 20 mV s^{-1} of the real material.

Nyquist plot and also affected the shape of the CV curves. To improve the performance of ECs based on the reported results, the selected electrolyte should have the highest ionic conductivity, the chosen membrane should have a sufficient size of porosity based on the electrolyte to get the lowest ionic resistivity, the current collector should have the highest conductivity with good surface interactions with the active material and the active material should have the highest electrical conductivity. The leakage resistance depends on the packaging of the cell, so packaging methods play an important role in the final result. Finally, the faradaic part of the active material can improve the capacity of the EC by helping the choice of the best candidate with high capacitance and low resistive behaviour (if R_{CT} is high in the material, it provides capacity with low resistive behaviour).

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This research was co-sponsored by the National Research Foundation of South Africa (Grant Number: 61056). All findings, conclusions or recommendations expressed in this work are solely those of the author(s). The NRF does not accept liability in any regard.

References

1 H. Ibrahim, A. Ilinca and J. Perron, *Renewable Sustainable Energy Rev.*, 2008, **12**, 1221–1250.
2 C. Shen, Y. Xie, B. Zhu, M. Sanghadasa, Y. Tang and L. Lin, *Sci. Rep.*, 2017, **7**, 14324.
3 P. Simon and Y. Gogotsi, *Nat. Mater.*, 2008, **7**, 845–854.
4 J. Pu, F. Cui, S. Chu, T. Wang, E. Sheng and Z. Wang, *ACS Sustainable Chem. Eng.*, 2014, **2**, 809–815.
5 H. Jiang, P. S. Lee and C. Li, *Energy Environ. Sci.*, 2013, **6**, 41–53.
6 F. Barzegar, A. Bello, J. K. J. K. Dangbegnon, N. Manyala and X. Xia, *Appl. Energy*, 2017, **207**, 417–426.
7 F. Barzegar, J. K. J. K. Dangbegnon, A. Bello, D. Y. D. Y. Momodu, A. T. C. T. C. Johnson and N. Manyala, *AIP Adv.*, 2015, **5**, 097171.
8 Y. Wang, Y. Song and Y. Xia, *Chem. Soc. Rev.*, 2016, **45**, 5925–5950.
9 F. Wang, X. Wu, X. Yuan, Z. Liu, Y. Zhang, L. Fu, Y. Zhu, Q. Zhou, Y. Wu and W. Huang, *Chem. Soc. Rev.*, 2017, **46**, 6816–6854.
10 T. Purkait, G. Singh, D. Kumar, M. Singh and R. S. Dey, *Sci. Rep.*, 2018, **8**, 640.
11 D. W. Lee, J. H. Lee, N. K. Min and J.-H. Jin, *Sci. Rep.*, 2017, **7**, 12005.
12 B.-A. Mei and L. Pilon, *Electrochim. Acta*, 2017, **255**, 168–178.
13 A. B. Cultura and Z. M. Salameh, Modeling, Evaluation and Simulation of a Supercapacitor Module for Energy Storage Application, *International Conference on Computer Information Systems and Industrial Applications (CISIA)*, 2015, pp. 876–882.
14 O. Bohlen, J. Kowal and D. U. Sauer, *J. Power Sources*, 2007, **173**, 626–632.
15 L. Zubieta and R. Bonert, *IEEE Trans. Ind. Appl.*, 2000, **36**, 199–205.
16 N. Ber, J. Sabatier, O. Briat and J.-M. Vinassa, *IEEE Trans. Ind. Electron.*, 2010, **57**, 3991–4000.
17 W. Lajnef, J.-M. Vinassa, O. Briat, S. Azzopardi and E. Woirgard, *J. Power Sources*, 2007, **168**, 553–560.
18 A. Hammar, P. Venet, R. Lallemand, G. Coquery and G. Rojat, *IEEE Trans. Ind. Electron.*, 2010, **57**, 3972–3979.
19 S.-M. Park and J.-S. Yoo, *Anal. Chem.*, 2003, **75**, 455A–461A.
20 M. E. Orazem and B. Tribollet, *Electrochemical Impedance Spectroscopy*, John Wiley & Sons, 2011.
21 H.-K. Song, H.-Y. Hwang, K.-H. Lee and L. H. Dao, *Electrochim. Acta*, 2000, **45**, 2241–2257.
22 H. He, R. Xiong and J. Fan, *Energies*, 2011, **4**, 582–598.
23 A. Rahmoun, H. Biechl and A. Rosin, *Electr. Contr. Commun. Eng.*, 2013, **2**, 34–39.
24 Z. J. Han, D. H. Seo, S. Yick, J. H. Chen and K. Ostrikov, *NPG Asia Mater.*, 2014, **6**, e140.
25 F. Barzegar, A. Bello, O. O. Fashedemi, J. K. Dangbegnon, D. Y. Momodu, F. Taghizadeh and N. Manyala, *Electrochim. Acta*, 2015, **180**, 442–450.
26 S. Buller, E. Karden, D. Kok and R. W. De Doncker, in *Thirty-sixth IAS annual meeting*, IEEE, 2001, vol. 4, pp. 2500–2504.
27 K. Sheng, Y. Sun, C. Li, W. Yuan and G. Shi, *Sci. Rep.*, 2012, **2**, 247.

ISA Transactions 91 (2019) 78–89

Contents lists available at ScienceDirect

ISA Transactions

journal homepage: www.elsevier.com/locate/isatrans

Research article

Active nonlinear partial-state feedback control of contacting force for a pantograph–catenary system

Bing Zhu [a,*], Zhiling Ren [b], Wenjing Xie [c], Fengyi Guo [b], Xiaohua Xia [d]

[a] The Seventh Research Division, Beihang University, Beijing 100191, PR China
[b] Faculty of Electrical and Control Engineering, Liaoning Technical University, Huludao 125105, PR China
[c] School of Computer and Information Science, Southwest University, Chongqing 400715, PR China
[d] Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0028, South Africa

HIGHLIGHTS

- The backstepping approach is firstly applied to control design for pantograph–catenary system.
- The closed-loop system is capable of tracking not only constant reference contacting force but also time-varying periodic reference forces.
- A high-order differentiator is designed to approximate the unknown derivatives of time-varying elasticity coefficient.
- A simple observer is designed to reconstruct the un-measurable system states.

ARTICLE INFO

Article history:
Received 11 May 2018
Received in revised form 16 October 2018
Accepted 25 January 2019
Available online 12 February 2019

Keywords:
Pantograph–catenary system
Backstepping
Trajectory tracking
Observer

ABSTRACT

In this paper, a nonlinear partial-state feedback control is designed for a 3-DOF pantograph–catenary system by using backstepping approach, such that the contacting force of the closed-loop system is capable of tracking its reference profile. In the control design, the pantograph–catenary model is transformed into a triangular form, facilitating the utilization of backstepping. Derivatives of virtual controls in backstepping are calculated explicitly. A high-order differentiator is designed to estimate the unknown time derivatives of elasticity coefficient; and an observer is proposed to reconstruct the unmeasurable states. It can be proved theoretically that, with the proposed nonlinear partial-state feedback control, the tracking error of the contacting force is ultimately bounded with tunable ultimate bounds. Theoretical results are demonstrated by numerical simulations.

1. Introduction

Pantograph–catenary system is prevailingly adopted in modern railway industry to supply electricity to high-speed trains. To guarantee that the high-speed train obtains stable electricity supply from the wires, a solid contact between pantograph and catenary is of great importance [1]. As pointed out in previous researches, the loss of contact would lead to insufficient supply of energy to the high-speed train, resulting in disfunctions in acceleration, braking and communication. In another aspect, however, with over-contacting force, there would be considerably arcing phenomenon or rapid wear in both pantograph and catenary, reducing significantly the duration of the entire system. Consequently, it is significantly necessary to maintain an appropriate contacting force between pantograph and catenary.

To maintain an optimal contacting force for the pantograph–catenary system is a challenging task, since there exist couplings and periodic nonlinearities in its dynamics [2] due to the high speed of the train. Influences resulted from couplings and periodic nonlinearities become especially negative, if the train speed is large such that the frequency of catenary stiffness variation is excessively high. The optimal contacting force between the pantograph and catenary can be calculated experimentally [3]. Active PID control strategy can be applied to maintaining a constant reference contacting force [4]. Other advanced control technologies that can be employed to the pantograph–catenary system include robust optimal control [5], feedback linearization [6], output-feedback regulation [7], output-feedback control with adaptive estimation [8], model predictive control [9], and an implementation-oriented technique named wire-actuated control with contact force estimation [10].

Some typical difficulties in active control design for pantograph–catenary system include that: (1) the elasticity coefficient of the catenary is time-varying, and the parameters in its mathematical model are unknown; and (2) some system states, such

* Corresponding author.
E-mail addresses: zhubing@buaa.edu.cn (B. Zhu), lngdrzl@163.com (Z. Ren), xiewenjing@swu.edu.cn (W. Xie), fyguo64@126.com (F. Guo), xxia@up.ac.za (X. Xia).

https://doi.org/10.1016/j.isatra.2019.01.033
0019-0578/

as displacement velocities of the pantograph, are unmeasurable. For the time-varying elasticity coefficient of the catenary, it can be assumed that it is periodic, and some approximations have been proposed [10,11]; however, the approximated models cannot be directly used, because accuracy of the approximated model is un-assured, and some parameters are difficult to determine. To estimate the unmeasurable states, sliding mode observers have been proposed [7]; however, typical problems in sliding mode control (such as chattering) would arise.

Generally, there exist two types of pantograph–catenary system modeling, namely 2-DOF modeling [7,11–14] and 3-DOF modeling [5,10,15,16]. The 3-DOF model contains more dynamics (thus more accurate) than the 2-DOF model; however, it is comparatively complicated, and there exist more uncertain parameters or un-measurable states. In this paper, based on backstepping approach, a nonlinear partial-state feedback control is proposed for a 3-DOF pantograph–catenary system. The linear time-varying model of the pantograph–catenary system can be transformed into a cascaded form, and backstepping can be applied to serve as the fundamental structure of the proposed controller, such that the controller can be designed in steps for reduced-order subsystems. Another advantage of applying backstepping is to facilitate the design of observers for unmeasurable states, and to guarantee the stability of the closed-loop system. *Main contributions* of this paper include: (1) the backstepping approach is firstly applied to control 3-DOF pantograph–catenary system; (2) by using backstepping approach, the closed-loop system is capable of tracking not only constant reference contacting force but also time-varying periodic reference forces; (3) a high-order differentiator is designed to approximate the unknown derivatives of time-varying elasticity coefficient, such that usage of unknown time-varying elasticity coefficient model can be avoided; and (4) a simple observer is designed to reconstruct the unmeasurable system states. Ultimate boundedness of tracking errors of the closed-loop system can be proved. The theoretical results are validated by numerical simulations.

The layout of this paper is arranged as following. In Section 2, the mathematical model of the pantograph–catenary system is presented, and the objectives of control design are stated. In Section 3, a full-state feedback nonlinear backstepping control is described in detail, and asymptotic stability of the tracking error is proved theoretically. In Section 4, a high-order differentiator is designed to estimate the time derivatives of elasticity coefficient, and an observer is designed to reconstruct the unmeasurable system states; it is proved that the tracking error of the closed-loop system with the proposed partial-state feedback control is ultimately bounded with tunable ultimate bounds. In Section 5, main theoretical results are demonstrated by numerical simulations, and corresponding discussions are given. The final section is the conclusion.

2. Problem statement

In this section, the pantograph–catenary system is modeled into a 3-DOF time-varying linear system. In the pantograph–catenary system, as depicted by Fig. 1, the pantograph is fixed on the top of the train, and runs in high-speed with the train. A supporting force is exerted on the lower frame of the pantograph from some actuators, and to generated a contacting force between the pan-head of the pantograph and the catenary, such that electrical power can be transferred from the catenary to the train through the pantograph.

Due to the high speed of the train, there exist some considerable vibrations in the catenary, and the contact force would be negatively influenced. Excessively large contact force would lead to extreme wear of the pan-head and the catenary, while too small

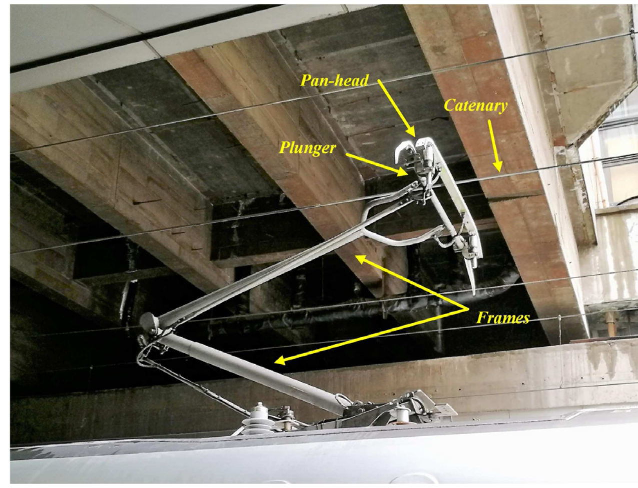


Fig. 1. Pantograph–catenary system equipped on a practical high-speed train.

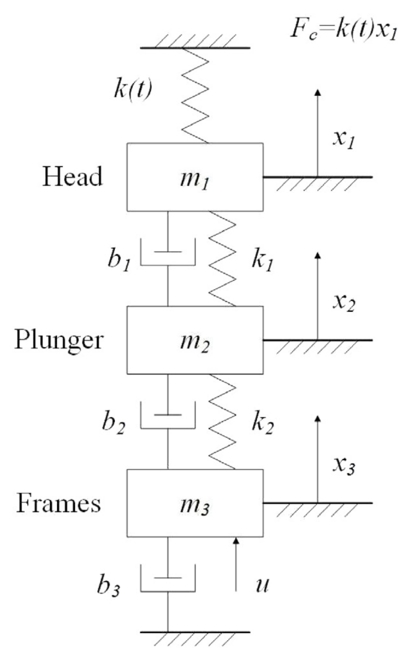


Fig. 2. Approximate structure of the pantograph–catenary system.

contact force would result in arcing phenomenon or even lost of contact. All of these situations will deteriorate the electricity supply for the train. Consequently, *the objective* of this paper is to design an active control for the pantograph, such that a proper reference contacting force can be maintained.

2.1. Mathematical model of pantograph–catenary system

The 3-DOF pantograph–catenary system is composed by a head, a plunger and frames. In this paper, the pantograph–catenary model under consideration is an under-actuated one, and it is different from the fully-actuated model in [5]. Its structure can be approximated by a mass-elasticity model, as is given in Fig. 2, where the active control u is only exerted on the lower frame. The dynamic equations of the 3-DOF pantograph–catenary system can be obtained by

$$\begin{cases} m_1\ddot{x}_1 = k_1(x_2 - x_1) + b_1(\dot{x}_2 - \dot{x}_1) - k(t)x_1, \\ m_2\ddot{x}_2 = -k_1(x_2 - x_1) - k_2(x_2 - x_3) - b_1(\dot{x}_2 - \dot{x}_1) - b_2(\dot{x}_2 - \dot{x}_3), \\ m_3\ddot{x}_3 = -k_2(x_3 - x_2) - b_2(\dot{x}_3 - \dot{x}_2) - b_3\dot{x}_3 + u, \end{cases} \quad (1)$$

where m_1 and m_2 denote the mass of the pantograph head and the plunger, respectively; m_3 denotes the gross mass of the frames; x_1, x_2 and x_3 are the positions of the pantograph head, the plunger and the frames, respectively; k_1 and k_2 denote the elasticity constants of the plunger and the frames; b_1, b_2 and b_3 are the damping constants of the pantograph head, the plunger and the frames, respectively; t denotes the continuous time; $k(t)$ is the time-varying elasticity (or stiffness) coefficient between the pantograph head and the wire; and u is the control input.

The output of the pantograph–catenary system is the contacting force between the pantograph head and the wire:

$$F_c \triangleq k(t)x_1, \tag{2}$$

where the contacting force F_c is assumed to be measured directly, and the position x_1 can be measured, indicating that the value of the elasticity coefficient can be obtained by

$$k(t) = \frac{F_c}{x_1}. \tag{3}$$

However, it is supposed in this paper that the accurate physical model of the time-varying elasticity coefficient $k(t)$ is unknown.

The time-varying elasticity coefficient can be approximated by a high-order periodic model [11]:

$$k(t) = K_0 + \sum_{i=1}^{3} K_i \cos(\frac{2i\pi}{L}Vt) + K_7 \cos(\frac{14\pi}{L}Vt), \tag{4}$$

where K_i $(i = 0, 1, 2, 3, 7)$ are constant uncertain stiffness coefficients; V is the train speed; and L is the span length.

Remark 1. The contact force can be measured directly with strain gages, accelerometers and strain gage position sensors, as proposed in [17]. It has to be acknowledged that, the measurement might be somehow inaccurate in harsh environment. The contact force can also be estimated by using numerical methods, e.g., [11], but with fully-known elasticity coefficient.

Remark 2. Nonlinearities in the pantograph are neglected in this paper, and the pantograph is assumed to be a 3-DOF mass–spring–damper system with constant elasticity and damping coefficients. Please see [18] for more details.

Remark 3. The elasticity coefficient (4) of the catenary is time-varying due to vibrations of the contacting wire, and it can be expanded as Taylor Series. In quite a lot of previous researches, higher-order terms in Taylor Series are neglected, and only the first order periodic term is considered [5,6,8,13,14]. However, some researches claim that it is inaccuracy to consider only the first periodic term in the Taylor Series [11,16]. Consequently, more higher-order terms in Taylor Series are considered in this paper to improve the accuracy of the structure. In the Taylor Series of the catenary, parameters are unmeasurable and to be estimated.

2.2. Linear time-varying representation

The system model (1) can be transformed into a linear time-varying representation:

$$\begin{cases} \dot z_1 = z_2, \\ \dot z_2 = -\frac{k_1+k(t)}{m_1}z_1 - \frac{b_1}{m_1}z_2 + \frac{k_1}{m_1}z_3 + \frac{b_1}{m_1}z_4, \\ \dot z_3 = z_4, \\ \dot z_4 = \frac{k_1}{m_2}z_1 + \frac{b_1}{m_2}z_2 - \frac{k_1+k_2}{m_2}z_3 - \frac{b_1+b_2}{m_2}z_4 + \frac{k_2}{m_2}z_5 + \frac{b_2}{m_2}z_6, \\ \dot z_5 = z_6, \\ \dot z_6 = \frac{k_2}{m_3}z_3 + \frac{b_2}{m_3}z_4 - \frac{k_2}{m_3}z_5 - \frac{b_2+b_3}{m_3}z_6 + \frac{1}{m_3}u, \end{cases} \tag{5}$$

or

$$\dot z = A(t)z + bu,$$

where $z = [z_1, z_2, z_3, z_4, z_5, z_6]^T \triangleq [x_1, \dot x_1, x_2, \dot x_2, x_3, \dot x_3]^T$, and

$$A(t)=\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -\frac{k_1+k(t)}{m_1} & -\frac{b_1}{m_1} & \frac{k_1}{m_1} & \frac{b_1}{m_1} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \frac{k_1}{m_2} & \frac{b_1}{m_2} & -\frac{k_1+k_2}{m_2} & -\frac{b_1+b_2}{m_2} & \frac{k_2}{m_2} & \frac{b_2}{m_2} \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{k_2}{m_3} & \frac{b_2}{m_3} & -\frac{k_2}{m_3} & -\frac{b_2+b_3}{m_3} \end{bmatrix},$$

$$b = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{m_3} \end{bmatrix}.$$

The output of the system can be given by

$$y = C(t)z, \tag{6}$$

where $y = F_c$, and $C(t) = [k(t), 0, 0, 0, 0, 0]$. System (5) is time-varying in existence of the time-varying elasticity coefficient $k(t)$.

2.3. Optimal contacting force with respect to mechanical wear and electrical resistance

The optimal contact force should achieve a tradeoff between material wear and electrical resistance of the pantograph–catenary system. The material wear decreases as the contact force decreases; meanwhile, decrease of the contact force would lead to larger electrical resistance between the pantograph and catenary, impeding the reliable currency transmission.

The material wear includes oxidational wear and melt wear. The oxidational wear model [19] of the contact wire can be given by

$$w_o(F_c) = f_m\left[\frac{\alpha_w \mu P_{eq}(F_c)}{L_{ox}} - \frac{A_n^{0.5}K_{ox}\left(T_m^{ox} - T_b\right)P_{eq}(F_c)^{0.5}n_{asp}^{0.5}}{L_{ox}H_0^{0.5}l_b v}\right], \tag{7}$$

where $w_o(F_c)$ denotes the wear of the contact wire, which is a function of the contacting force; f_m represents the volume fraction of molten material; α_w denotes the heat distribution coefficient; μ is the sliding friction coefficient of the contact wire material; L_{ox} denotes the latent heat of fusion per unit volume of oxide; A_n is the actual contact area; K_{ox} represents the thermal conductivity of oxide; T_m^{ox} denotes the melting temperature of the material; T_b is the bulk temperature; n_{asp} is the number of asperity in contact; H_0 denotes the hardness of material; l_b denotes the equivalent linear diffusion distance for bulk heating; and v is the actual velocity of the train. Some values of the above parameters can be found in [19,20]. The equivalent contact force $P_{eq}(F_c)$ is a function of the contact force F_c:

$$P_{eq}(F_c) = F_c + P_e = F_c + \frac{R_e}{\mu v}I^2, \tag{8}$$

where P_e is the equivalent electrical contact force; R_e denotes the electrical resistance at the contact point; and I is the electrical current transferred through the contact.

The melt wear model [19] of the contact wire is a function of the contacting force F_c, and it can be given by

$$w_m(F_c) = f_m\left[\frac{\alpha_w \mu P_{eq}(F_c)}{L_m} - \frac{A_n K_m\left(T_m - T_0\right)}{L_m l_b v}\right], \tag{9}$$

where L_m is the latent heat of fusion per unit volume for metal; K_m is the thermal conductivity of the metal material; T_m is the melting temperature of the metal material; and T_0 is the ambient temperature.

The electrical resistance with respect to contact force [21] can be calculated by

$$R_e(F_c) = \frac{\rho_1 + \rho_2}{4}\sqrt{\frac{\pi H}{F_c}}, \tag{10}$$

where ρ_1 and ρ_2 are resistance rates of the pantograph and the catenary; and H is the contact hardness of the materials.

Based on (7)–(10), the cost function to calculate the optimal contact force can be constructed by

$$J(F_c) = q_1 w_o(F_c) + q_2 w_m(F_c) + q_3 R_e(F_c), \tag{11}$$

where q_1, q_2 and q_3 are positive weight parameters for optimization. The optimal contact force can be calculated by

$$F_c^* = \arg\min_{F_c} J(F_c), \tag{12}$$

subject to (7)–(10).

The value of optimal contacting force may vary from case to case with respect to different values of parameters in various pantograph–catenary projects. Detailed value of pantograph–catenary parameters can be found in examples in [11,22,23] Generally, the optimal contacting force is often constant, and its value is around 100–120 N [23].

Remark 4. It has to be admitted that, in this paper, no systematic way of selecting weight parameters can be given. Parameter selecting has to be processed through trials. Different values of weight parameters reflect different emphasis on electricity resistance (contact force) or wear. For example, with the values of parameters provided in [19] and the weight parameters $q_1 = 0.2$, $q_2 = 1$ and $q_3 = 0.5$, it can be calculated that the optimal contacting force should be $F_c^* = 109.7$N.

Remark 5. The optimization (12) can be solved by using MATLAB function "`fmincon`".

2.4. Control objective

Suppose that, in this research, the contact force (F_c) and the positions (x_1, x_2 and x_3, or z_1, z_3 and z_5) can be measured directly. However, in practical cases, the velocities ($\dot{x}_1, \dot{x}_2$ and $\dot{x}_3$, or z_2, z_4 and z_6) are often unmeasurable. Moreover, the stiffness coefficients (K_i) are uncertain constant parameters.

Remark 6. Although corresponding devices are fairly expensive, the contact force is measurable indeed. There exist some researches on estimation of contact force without using the expensive devices [10,11].

The *objective* of this paper is to design a nonlinear control for the pantograph–catenary system with unmeasurable displacement variations and uncertain stiffness coefficients, such that the output of the system is capable of tracking a constant reference contacting force with small tracking error:

$$\lim_{t\to+\infty} |y(t) - y_r| < \epsilon, \tag{13}$$

where $y_r = F_c^*$ is the reference contacting force, and $\epsilon > 0$ is a small positive number.

3. Full-state feedback nonlinear control

In this section, a full-state feedback nonlinear control is designed to give a fundamental structure of the proposed partial-state feedback control. In the next section, differentiators and state observers will be designed to replace the uncertain parameters and unmeasurable states.

To facilitate control design, the time-varying model is transformed into a triangular form (the definition of triangular system can be referred to [24]). The nonlinear control is designed through backstepping [25], with derivatives of virtual controls calculated explicitly. Asymptotic stability of tracking errors of the closed-loop system is proved theoretically.

3.1. Model transformation

Define manifolds:

$$\xi_1 = k_1 z_3 + b_1 z_4, \tag{14}$$

$$\xi_2 = k_2 z_5 + b_2 z_6. \tag{15}$$

It follows from (5), (14) and (15) that the 3-DOF model can be transformed into a triangular form

$$\begin{cases} \dot{z}_1 &= z_2, \\ \dot{z}_2 &= -\frac{k_1 + k(t)}{m_1} z_1 - \frac{b_1}{m_1} z_2 + \frac{1}{m_1}\xi_1, \\ \dot{\xi}_1 &= k_1 z_4 + b_1\left(\frac{k_1}{m_2} z_1 + \frac{b_1}{m_2} z_2 - \frac{k_1+k_2}{m_2} z_3 - \frac{b_1+b_2}{m_2} z_4 + \frac{1}{m_2}\xi_2\right), \\ \dot{\xi}_2 &= k_2 z_6 + b_2\left(\frac{k_2}{m_3} z_3 + \frac{b_2}{m_3} z_4 - \frac{k_2}{m_3} z_5 - \frac{b_2+b_3}{m_3} z_6 + \frac{1}{m_3} u\right), \end{cases} \tag{16}$$

with internal dynamics given by

$$\begin{cases} \dot{z}_3 &= -\frac{k_1}{b_1} z_3 + \frac{1}{b_1}\xi_1, \\ \dot{z}_5 &= -\frac{k_2}{b_2} z_5 + \frac{1}{b_2}\xi_2. \end{cases} \tag{17}$$

Remark 7. It can be seen from (17) that the internal dynamics is actually a linear stable system

$$\begin{cases} \dot{z}_3 &= -\frac{k_1}{b_1} z_3, \\ \dot{z}_5 &= -\frac{k_2}{b_2} z_5. \end{cases} \tag{18}$$

plus inputs $\frac{1}{b_1}\xi_1$ and $\frac{1}{b_2}\xi_2$.

In another aspect, based on (6), it holds that

$$x_1 = z_1 = \frac{y}{k(t)}.$$

Then, an auxiliary reference profile can be defined:

$$z_{1r} \triangleq \frac{y_r}{k(t)}.$$

The objective is then to design control for the triangular system (16)–(17), such that

$$\lim_{t\to+\infty} |z_1(t) - z_{1r}(t)| < \frac{\epsilon}{\sup_{t\to+\infty}(|k(t)|)}. \tag{19}$$

3.2. Control design by using backstepping

The nonlinear control for (16)–(17) is designed step by step via backstepping in this section.

Step 1: Define tracking error $e_1 = z_1 - z_{1r}$. It follows that

$$\dot{e}_1 = \dot{z}_1 - \dot{z}_{1r} = z_2 - \dot{z}_{1r} = e_2 + \alpha_1 - \dot{z}_{1r},$$

where $e_2 \triangleq z_2 - \alpha_1$, and α_1 is the virtual control to be tracked by z_2. Design the virtual control

$$\alpha_1 = -c_1 e_1 + \dot{z}_{1r}, \tag{20}$$

where $c_1 > 0$ is a constant control gain. It then follows that

$$\dot{e}_1 = -c_1 e_1 + e_2. \tag{21}$$

Select the Lyapunov candidate $L_1 = \frac{1}{2}e_1^2$. Its time derivative can be calculated by

$$\dot{L}_1 = -c_1 e_1^2 + e_1 e_2, \tag{22}$$

where $-c_1e_1^2$ is negative definite, and e_1e_2 is to be canceled in the next step.

Step 2: The time derivative of e_2 can be calculated by

$$\begin{aligned}\dot{e}_2 =&\dot{z}_2-\dot{\alpha}_1\\ =&-\frac{k_1+k(t)}{m_1}z_1-\frac{b_1}{m_1}z_2+\frac{1}{m_1}\xi_1-\dot{\alpha}_1\\ =&-\frac{k_1+k(t)}{m_1}z_1-\frac{b_1}{m_1}z_2-\dot{\alpha}_1+e_3+\alpha_2,\end{aligned}$$

where $e_3 \triangleq \frac{1}{m_1}\xi_1-\alpha_2$, and α_2 is the virtual control to be tracked by $\frac{1}{m_1}\xi_1$. Design the virtual control

$$\alpha_2=-e_1-c_2e_2+\frac{k_1+k(t)}{m_1}z_1+\frac{b_1}{m_1}z_2+\dot{\alpha}_1, \tag{23}$$

where $c_2>0$ is a constant control gain. It then follows that

$$\dot{e}_2=-e_1-c_2e_2+e_3.$$

Select the Lyapunov candidate $L_2=L_1+\frac{1}{2}e_2^2$. Its time derivative can be calculated by

$$\dot{L}_1=-c_1e_1^2-c_2e_2^2+e_2e_3,$$

where $-c_1e_1^2-c_2e_2^2$ is negative definite, e_1e_2 in (22) is canceled, and e_2e_3 is to be canceled in the next step.

Step 3: The time derivative of e_3 can be calculated by

$$\begin{aligned}\dot{e}_3 =&\frac{1}{m_1}\dot{\xi}_1-\dot{\alpha}_2\\ =&\frac{b_1}{m_1}\left(\frac{k_1}{m_2}z_1+\frac{b_1}{m_2}z_2-\frac{k_1+k_2}{m_2}z_3-\frac{b_1+b_2}{m_2}z_4\right)\\ &+\frac{k_1}{m_1}z_4-\dot{\alpha}_2+\frac{b_1}{m_1}\frac{1}{m_2}\xi_2\\ =&\frac{b_1}{m_1}\left(\frac{k_1}{m_2}z_1+\frac{b_1}{m_2}z_2-\frac{k_1+k_2}{m_2}z_3-\frac{b_1+b_2}{m_2}z_4\right)\\ &+\frac{k_1}{m_1}z_4-\dot{\alpha}_2+e_4+\alpha_3.\end{aligned}$$

where α_3 is the virtual control to be tracked by $\frac{b_1}{m_1m_2}\xi_2$, and $e_4 \triangleq \frac{b_1}{m_1m_2}\xi_2-\alpha_3$.

Design the virtual control

$$\begin{aligned}\alpha_3=&-e_2-c_3e_3+\dot{\alpha}_2-\frac{k_1}{m_1}z_4\\ &-\frac{b_1}{m_1}\left(\frac{k_1}{m_2}z_1+\frac{b_1}{m_2}z_2-\frac{k_1+k_2}{m_2}z_3-\frac{b_1+b_2}{m_2}z_4\right),\end{aligned} \tag{24}$$

where $c_3>0$ is a constant control gain. It then follows that

$$\dot{e}_3=-e_2-c_3e_3-e_4.$$

Select the Lyapunov candidate $L_3=L_2+\frac{1}{2}e_3^2$. Its time derivative can be calculated by

$$\dot{L}_3=-c_1e_1^2-c_2e_2^2-c_3e_3^2+e_3e_4,$$

where e_3e_4 is to be backstepped in the next step.

Step 4: The time derivative of e_4 can be calculated by

$$\begin{aligned}\dot{e}_4 =&\frac{b_1}{m_1m_2}\dot{\xi}_2-\dot{\alpha}_3\\ =&\frac{b_1b_2}{m_1m_2}\left(\frac{k_2}{m_3}z_3+\frac{b_2}{m_3}z_4-\frac{k_2}{m_3}z_5-\frac{b_2+b_3}{m_3}z_6\right)\\ &+\frac{b_1k_2}{m_1m_2}z_6-\dot{\alpha}_3+\frac{b_1b_2}{m_1m_2m_3}u,\end{aligned}$$

where u is the control to be designed.

The control can be designed by

$$\begin{aligned}u=&\frac{m_1m_2m_3}{b_1b_2}\left(-e_3-c_4e_4+\dot{\alpha}_3-\frac{b_1k_2}{m_1m_2}z_6\right.\\ &\left.-\frac{b_1b_2}{m_1m_2}\left(\frac{k_2}{m_3}z_3+\frac{b_2}{m_3}z_4-\frac{k_2}{m_3}z_5-\frac{b_2+b_3}{m_3}z_6\right)\right),\end{aligned} \tag{25}$$

where $c_4>0$ is the control parameter. Select Lyapunov candidate $L_4=L_3+\frac{1}{2}e_4^2$; it follows that

$$\dot{L}_3=-c_1e_1^2-c_2e_2^2-c_3e_3^2-c_4e_4^2, \tag{26}$$

which ends the backstepping design.

3.3. Time derivatives of virtual controls

As can be seen from Section 3.2, the proposed control is given by (25), where virtual controls are given by (20) and (23). It should be noted in (23) and (25) that, before applying the proposed backstepping-based nonlinear control, derivatives of virtual controls (namely $\dot{\alpha}_1$ and $\dot{\alpha}_2$) should be calculated.

The time derivative of virtual control α_1 can be calculated by

$$\dot{\alpha}_1=-c_1\dot{e}_1+\ddot{z}_{1r}=-c_1\left(z_2-\dot{z}_{1r}\right)+\ddot{z}_{1r}, \tag{27}$$

where

$$\dot{z}_{1r}=\frac{\mathrm{d}\left(\frac{y_r}{k(t)}\right)}{\mathrm{d}t}=\frac{\dot{y}_rk-y_r\dot{k}}{k^2}, \tag{28}$$

$$\ddot{z}_{1r}=\frac{(\ddot{y}_rk-y_r\ddot{k})k^2-2k\dot{k}(\dot{y}_rk-y_r\dot{k})}{k^4}, \tag{29}$$

$$\dot{k}=-\sum_{i=1}^{3}K_i\omega_i\sin(\frac{2i\pi}{L}Vt)-K_7\omega_7\sin(\frac{14\pi}{L}Vt), \tag{30}$$

$$\ddot{k}=-\sum_{i=1}^{3}K_i\omega_i^2\cos(\frac{2i\pi}{L}Vt)-K_7\omega_7^2\cos(\frac{14\pi}{L}Vt), \tag{31}$$

$$\omega_i=\frac{2i\pi}{L}V,\quad i=1,2,3,7. \tag{32}$$

The time derivative of virtual control α_2 can be calculated by

$$\dot{\alpha}_2=-\dot{e}_1-c_2\dot{e}_2+\frac{k_1+k}{m_1}\dot{z}_1+\frac{\dot{k}}{m_1}z_1+\frac{b_1}{m_1}\dot{z}_2+\ddot{\alpha}_1, \tag{33}$$

where $\dot{e}_1$ can be calculated by

$$\dot{e}_1=\left(z_2-\dot{z}_{1r}\right), \tag{34}$$

and $\dot{z}_{1r}$ is calculated by (28); $\dot{e}_2$ can be calculated by

$$\dot{e}_2=-\frac{k_1+k}{m_1}z_1-\frac{b_1}{m_1}z_2+\frac{k_1}{m_1}z_3+\frac{b_1}{m_1}z_4-\dot{\alpha}_1, \tag{35}$$

where $\dot{\alpha}_1$ is calculated by (27)–(31); $\dot{z}_1$ and $\dot{z}_2$ can be obtained by

$$\dot{z}_1=z_2, \tag{36}$$

$$\dot{z}_2=\dot{e}_2+\dot{\alpha}_1, \tag{37}$$

where $\dot{e}_2$ is calculated by (35), and $\dot{\alpha}_1$ is calculated by (27); $\ddot{\alpha}_1$ can be calculated by

$$\ddot{\alpha}_1=-c_1\ddot{e}_1+z_{1r}^{(3)}=-c_1\left(\dot{z}_2-\ddot{z}_{1r}\right)+z_{1r}^{(3)}, \tag{38}$$

where $\dot{z}_2$ and $\ddot{z}_{1r}$ can be obtained respectively by (31) and (37). Denote $\ddot{z}_{1r}$ in (29) by

$$\ddot{z}_{1r}=\frac{\theta-\phi}{\psi},$$

where

$$\theta=(\ddot{y}_rk-y_r\ddot{k})k^2,$$

$\phi = 2k\dot{k}(\dot{y}_r k - y_r \dot{k})$,
$\psi = k^4$.

It follows that

$$z_{1r}^{(3)} = \frac{(\dot{\theta} - \dot{\phi})\psi - (\theta - \phi)\dot{\psi}}{\psi^2},$$

where

$$\dot{\theta} = \left(y_r^{(3)} k + \ddot{y}_r \dot{k} - \dot{y}_r \ddot{k} - y_r k^{(3)}\right) k^2 + 2k\dot{k}\left(\ddot{y}_r k - y_r \ddot{k}\right), \tag{39}$$

$$\dot{\phi} = 2k\dot{k}\left(\ddot{y}_r k - y_r \ddot{k}\right) + \left(\dot{y}_r k - y_r \dot{k}\right)\left(2\dot{k}^2 + 2k\ddot{k}\right), \tag{40}$$

$$\dot{\psi} = 4k^3\dot{k}. \tag{41}$$

$$\dddot{k} = \sum_{i=1}^{3} K_i \omega_i^3 \sin(\frac{2i\pi}{L} Vt) + K_7 \omega_7^3 \sin(\frac{14\pi}{L} Vt). \tag{42}$$

The time derivative of α_3 can be calculated by

$$\dot{\alpha}_3 = -\dot{e}_2 - c_3\dot{e}_3 - \frac{k_1}{m_1}\dot{z}_4 - \frac{b_1}{m_1}\left(\frac{k_1}{m_2}\dot{z}_1 + \frac{b_1}{m_2}\dot{z}_2 - \frac{k_1 + k_2}{m_2}\dot{z}_3 - \frac{b_1 + b_2}{m_2}\dot{z}_4\right) + \ddot{\alpha}_2, \tag{43}$$

where $\dot{e}_2$ can be calculated by (35); $\dot{e}_3$ can be calculated by

$$\dot{e}_3 = \frac{b_1}{m_1}\left(\frac{k_1}{m_2}z_1 + \frac{b_1}{m_2}z_2 - \frac{k_1 + k_2}{m_2}z_3 - \frac{b_1 + b_2}{m_2}z_4\right) + \frac{k_1}{m_1}z_4 - \dot{\alpha}_2 + \frac{b_1}{m_1}\frac{1}{m_2}\xi_2; \tag{44}$$

$\dot{z}_i$ $(i = 1, 2, 3, 4)$ can be calculated by using the state equations in (5).

In (43), the second-order derivative of α_2 in (43) can be calculated by

$$\ddot{\alpha}_2 = -\ddot{e}_1 - c_2\ddot{e}_2 + \frac{k_1 + k}{m_1}\ddot{z}_1 + \frac{\dot{k}}{m_1}\dot{z}_1 + \frac{\ddot{k}}{m_1}z_1 + \frac{b_1}{m_1}\ddot{z}_2 + \dddot{\alpha}_1, \tag{45}$$

where

$$\ddot{e}_1 = \dot{z}_2 - \ddot{z}_{1r}, \quad \dot{z}_2 = \dot{e}_2 + \dot{\alpha}_1, \tag{46}$$

$$\ddot{e}_2 = -\frac{k_1 + k}{m_1}\dot{z}_1 - \frac{\dot{k}}{m_1}z_1 - \frac{b_1}{m_1}\dot{z}_2 + \frac{k_1}{m_1}\dot{z}_3 + \frac{b_1}{m_1}\dot{z}_4 - \ddot{\alpha}_1, \tag{47}$$

$$\ddot{z}_1 = \dot{z}_2, \quad \ddot{z}_2 = \ddot{e}_2 + \ddot{\alpha}_1, \tag{48}$$

$$\dddot{\alpha}_1 = -c_1(\dddot{z}_2 - \dddot{z}_{1r}) + z_{1r}^{(4)}, \tag{49}$$

$$z_{1r}^{(4)} = \frac{\ddot{\theta} - \ddot{\phi}}{\psi} - \frac{2(\dot{\theta} - \dot{\phi})\dot{\psi} + (\theta - \phi)\ddot{\psi}}{\psi^2} + \frac{2(\theta - \phi)\dot{\psi}^2}{\psi^3}. \tag{50}$$

Remark 8. It is implied from (27)–(50) that the derivatives of virtual controls can be explicitly calculated from system states and the reference contacting force.

3.4. Analysis on closed-loop system

The control algorithm can be summarized as following.

Algorithm 1.

(1) Calculate the virtual control α_1 with (20) and (28).
(2) Calculate the virtual control α_2 with (23) and (27)–(31).
(3) Calculate the virtual control α_3 with (24) and (33), where $\dot{e}_1$, $\dot{e}_2$, $\dot{z}_1$ and $\dot{z}_2$ are calculated with (34)–(37), and $\ddot{\alpha}_1$ is calculated with (38)–(42).
(4) Calculate the control u with (25), where some relevant terms can be calculated by (43)–(50).

Stability of the closed-loop system with the proposed control algorithm can be given by the following proposition.

Proposition 1. *Consider the pantograph–catenary system given by (1)–(4). Its reference contacting force is constant or time-varying continuous periodic. If the control is designed by Algorithm 1, then tracking error of the closed-loop system are globally asymptotically stable, and (13) is satisfied globally.*

Proof. Consider Lyapunov candidate L_4. It satisfies

$$\beta_1\|e\|^2 \le L_4 \le \beta_2\|e\|^2,$$

where $e \triangleq [e_1, e_2, e_3, e_4]^T$, $\beta_1 = \beta_2 = \frac{1}{2}$, and $\|\cdot\|$ denotes the Euclidean norm of vector or co-vector. The time derivative of L_4 along the closed-loop system with the control algorithm given in Algorithm 1 can be calculated by

$$\dot{L}_4 = -c_1e_1^2 - c_2e_2^2 - c_3e_3^2 - c_4e_4^2 \le -\beta_3\|e\|^2, \tag{51}$$

where $\beta_3 = \min[c_1, c_2, c_3, c_4]$. Moreover,

$$\left\|\frac{\partial L_4}{\partial e}\right\| \le \beta_4\|e\|,$$

where $\beta_4 = 1$. Consequently, according to Theorem 4.10 in [25], L_4 is a Lyapunov function, and e_1, e_2, e_3 and e_4 are globally asymptotically stable.

Based on (51), it can be obtained that

$$L_4(t) \le \mathrm{e}^{-\frac{\beta_3}{\beta_2}t}L_4(0),$$

and therefore,

$$\|e_1\| \le \|e\| \le \sqrt{\frac{1}{\beta_1}L_4(t)} \le \sqrt{\frac{1}{\beta_1}\mathrm{e}^{-\frac{\beta_3}{\beta_2}t}L_4(0)}, \tag{52}$$

indicating that (13) and (19) are satisfied.

Moreover, according to Proposition 4 in Appendix, z_3 and z_5 track periodic trajectories z_{3r} and z_{5r} asymptotically, and tracking errors $e_3^z \triangleq z_3 - z_{3r}$ and $e_5^z \triangleq z_5 - z_{5r}$ satisfy (70) in Appendix, where $\mathcal{L}_2(0) = 0$ and $\|\mathcal{L}_2(e)\| \le \kappa_3^z\|e\|$ with $\kappa_3^z > 0$.

Select a Lyapunov candidate $L_0 = L_4 + \frac{1}{2\gamma_3^z}{e_3^z}^2 + \frac{1}{2\gamma_5^z}{e_5^z}^2$ for the full-state closed-loop system, where $\gamma_3^z > 0$ and $\gamma_5^z > 0$. Its time derivative can be calculated by

$$\begin{aligned}\dot{L}_0 \le &-\beta_3\|e\|^2 - \frac{k_1}{b_1\gamma_3^z}{e_3^z}^2 + \frac{\kappa_3^z}{\gamma_3^z}e_3^z\|e\| - \frac{k_2}{b_2\gamma_5^z}{e_5^z}^2 + \frac{\kappa_5^z}{\gamma_5^z}e_5^z\|e\| \\ = &-\left(\frac{1}{2}\beta_3 - \frac{{\kappa_3^z}^2 b_1}{4k_1\gamma_3^z}\right)\|e\|^2 - \left(\sqrt{\frac{k_1}{b_1\gamma_3^z}}e_3^z - \frac{\kappa_3^z}{2}\sqrt{b_1}k_1\gamma_3^z\|e\|\right)^2 \\ &-\left(\frac{1}{2}\beta_3 - \frac{{\kappa_5^z}^2 b_2}{4k_2\gamma_5^z}\right)\|e\|^2 - \left(\sqrt{\frac{k_2}{b_2\gamma_5^z}}e_5^z - \frac{\kappa_5^z}{2}\sqrt{b_2}k_2\gamma_5^z\|e\|\right)^2 \\ \le &0,\end{aligned}$$

where γ_3^z and γ_5^z can be selected appropriately such that

$$\left(\frac{1}{2}\beta_3 - \frac{{\kappa_3^z}^2 b_1}{4k_1\gamma_3^z}\right) > 0, \quad \left(\frac{1}{2}\beta_3 - \frac{{\kappa_5^z}^2 b_2}{4k_2\gamma_5^z}\right) > 0;$$

and $\dot{L}_0 = 0$ if and only if $e = 0$, $e_3^z = 0$ and $e_5^z = 0$.

Consequently, tracking errors of the closed-loop system with the proposed control are globally asymptotically stable. □

Remark 9. For more details about the principle and design process of backstepping, please see [25].

Remark 10. It should be noted that the system (1) (or (5)) is linear time-varying; consequently, the criteria of stability for time-varying system (e.g., Theorem 4.10 in [25]) should be used for closed-loop system analysis.

Remark 11. As can be seen from (52), performances of the closed-loop system can be tuned by control gains.

4. Partial-state feedback control

In practical applications, although its value can be obtained by using (3), the elasticity coefficient model (4) are usually unknown, indicating that $\dot{k}$, $\ddot{k}$, $\dddot{k}$ and $k^{(4)}$ cannot be directly calculated through the steps in Section 3. Moreover, velocities of the springs z_2, z_4 and z_6 are un-measurable; they cannot be used directly for state feedback.

In this section, it is supposed that the actual contacting force y, displacements z_1, z_3 and z_5 are measurable; differentiators and observers are designed to estimate the uncertain $\dot{k}$, $\ddot{k}$, $\dddot{k}$ and $k^{(4)}$, and un-measurable z_2, z_4 and z_6.

4.1. High-order differentiators for estimating $\dot{k}$, $\ddot{k}$ and $k^{(3)}$

It follows from (6) that $k(t)$ can be obtained by

$$k = \frac{y}{x_1}, \tag{53}$$

where $y = F_c$ and x_1 can be directly measured.

A simple high-order differentiator can be introduced to estimate time-derivatives of the elasticity coefficient:

$$\begin{cases} \dot{\zeta}_1 = \zeta_2, \\ \dot{\zeta}_2 = \zeta_3, \\ \dot{\zeta}_3 = \zeta_4, \\ \dot{\zeta}_4 = \zeta_5, \\ \dot{\zeta}_5 = R^5\left(-a_1(\zeta_1 - k(t)) - \frac{a_2}{R}\zeta_2 - \frac{a_3}{R^2}\zeta_3 - \frac{a_4}{R^3}\zeta_4 - \frac{a_5}{R^4}\zeta_5\right), \end{cases} \tag{54}$$

where a_i $(i = 1, 2, 3, 4, 5)$ and R are positive differentiator parameters to be tuned. For some recent detailed researches in differentiators, please refer to [26,27].

The time-derivatives of k are estimated by

$$\hat{k} = \zeta_1, \tag{55}$$

$$\hat{\dot{k}} = \zeta_2, \tag{56}$$

$$\hat{\ddot{k}} = \zeta_3, \tag{57}$$

$$\widehat{k^{(3)}} = \zeta_4, \tag{58}$$

$$\widehat{k^{(4)}} = \zeta_5. \tag{59}$$

Proposition 2. *With the differentiator (54), the time derivatives of the elasticity coefficient can be estimated by (56)–(59) with bounded estimation errors.*

Proof. It is obvious that (54) is an asymptotically stable linear system with a periodic input $k(t)$. Consequently, it can be claimed that ζ_1 tracks $k(t)$ with bounded tracking errors, which can be tuned arbitrarily small by assigning appropriate a_i $(i = 1, 2, 3, 4, 5)$ and R. It can be seen that ζ_i $(i = 2, 3, 4, 5)$ are time derivatives of ζ_1, and they are uniformly differentiable; as a result, they are capable of tracking time derivatives of k with bounded errors. □

Remark 12. Let $\tilde{\zeta}_i \triangleq \zeta_i - k^{(i-1)}$ $(i = 1, 2, 3, 4, 5)$, and $\tilde{\zeta} \triangleq [\tilde{\zeta}_1, \tilde{\zeta}_2, \tilde{\zeta}_3, \tilde{\zeta}_4, \tilde{\zeta}_5]^T$. It is apparent that estimation errors of the differentiator are input-to-state stable (ISS [25]) with respect to $k^{(i-1)}$ $(i = 2, 3, 4, 5)$. Moreover, there exists a positive function $L_5(\tilde{\zeta}_i)$ satisfying

$$\delta_1^d\|\tilde{\zeta}\|^2 \le L_5 \le \delta_2^d\|\tilde{\zeta}\|^2,$$
$$\dot{L}_5 \le -\delta_3^d\|\tilde{\zeta}\|^2 + \beta_d(\dot{k}, \ddot{k}, k^{(3)}, k^{(4)}),$$

where $\delta_i^d > 0$, $(i = 1, 2, 3)$, and β_d is a positive scalar satisfying $\beta_d(0, 0, 0, 0) = 0$.

Remark 13. It should be noted that $\hat{\dot{k}}$, $\hat{\ddot{k}}$, $\widehat{k^{(3)}}$, and $\widehat{k^{(4)}}$ are estimated values of $\dot{k}$, $\ddot{k}$, $k^{(3)}$, and $k^{(4)}$; they are different from derivatives $\dot{\hat{k}}$, $\ddot{\hat{k}}$, $\hat{k}^{(3)}$, and $\hat{k}^{(4)}$.

Remark 14. In this section, the high-order differentiator is applied to estimate derivatives of k. More detailed analysis on high-order differentiators can be found in [26] and [27].

4.2. Observer for z_2, z_4 and z_6

The observer for z_2, z_4 and z_6 can be designed by

$$\begin{cases} \dot{\hat{z}}_1 = \hat{z}_2 + l_1(\tilde{z}_1, \tilde{z}_3, \tilde{z}_5), \\ \dot{\hat{z}}_2 = -\frac{k_1+k(t)}{m_1}\hat{z}_1 - \frac{b_1}{m_1}\hat{z}_2 + \frac{k_1}{m_1}\hat{z}_3 + \frac{b_1}{m_1}\hat{z}_4 + l_2(\tilde{z}_1, \tilde{z}_3, \tilde{z}_5), \\ \dot{\hat{z}}_3 = \hat{z}_4 + l_3(\tilde{z}_1, \tilde{z}_3, \tilde{z}_5), \\ \dot{\hat{z}}_4 = \frac{k_1}{m_2}\hat{z}_1 + \frac{b_1}{m_2}\hat{z}_2 - \frac{k_1+k_2}{m_2}\hat{z}_3 - \frac{b_1+b_2}{m_2}\hat{z}_4 + \frac{k_2}{m_2}\hat{z}_5 \\ \qquad + \frac{b_2}{m_2}\hat{z}_6 + l_4(\tilde{z}_1, \tilde{z}_3, \tilde{z}_5) \\ \dot{\hat{z}}_5 = \hat{z}_6 + l_5(\tilde{z}_1, \tilde{z}_3, \tilde{z}_5), \\ \dot{\hat{z}}_6 = \frac{k_2}{m_3}\hat{z}_3 + \frac{b_2}{m_3}\hat{z}_4 - \frac{k_2}{m_3}\hat{z}_5 - \frac{b_2+b_3}{m_3}\hat{z}_6 + l_6(\tilde{z}_1, \tilde{z}_3, \tilde{z}_5) + \frac{1}{m_3}u, \end{cases} \tag{60}$$

where z_1, z_3 and z_5 are outputs; $\hat{z}_i$ $(i = 1, 2, 3, 4, 5, 6)$ are estimations of z_i $(i = 1, 2, 3, 4, 5, 6)$; $\tilde{z}_i \triangleq \hat{z}_i - z_i$ $(i = 1, 2, 3, 4, 5, 6)$ are estimation errors; and

$$l_1(\tilde{z}_1, \tilde{z}_3, \tilde{z}_5) = -\varpi_1\tilde{z}_1,$$
$$l_2(\tilde{z}_1, \tilde{z}_3, \tilde{z}_5) = \left(\frac{k_1 + k(t)}{m_1} - \varpi_2\right)\tilde{z}_1 - \frac{k_1}{m_1}\tilde{z}_3,$$
$$l_3(\tilde{z}_1, \tilde{z}_3, \tilde{z}_5) = -\varpi_3\tilde{z}_3,$$
$$l_4(\tilde{z}_1, \tilde{z}_3, \tilde{z}_5) = -\frac{k_1}{m_2}\tilde{z}_1 + \left(\frac{k_1 + k_2}{m_2} - \varpi_4\right)\tilde{z}_3 - \frac{k_2}{m_2}\tilde{z}_5,$$
$$l_5(\tilde{z}_1, \tilde{z}_3, \tilde{z}_5) = -\varpi_5\tilde{z}_5,$$
$$l_6(\tilde{z}_1, \tilde{z}_3, \tilde{z}_5) = -\frac{k_2}{m_3}\tilde{z}_3 + \left(\frac{k_2}{m_3} - \varpi_6\right)\tilde{z}_5,$$

where $\varpi_i > 0$ $(i = 1, 2, 3, 4, 5, 6)$.

Proposition 3. *With the observer designed by (60), observation errors $\tilde{z}_i$ $(i = 1, 2, 3, 4, 5, 6)$ are globally exponentially stable.*

Proof. Subtracting (60) by (5) yields

$$\begin{cases} \dot{\tilde{z}}_1 = -\varpi_1\tilde{z}_1 + \tilde{z}_2, \\ \dot{\tilde{z}}_2 = -\varpi_2\tilde{z}_1 - \frac{b_1}{m_1}\tilde{z}_2 + \frac{b_1}{m_1}\tilde{z}_4, \\ \dot{\tilde{z}}_3 = -\varpi_3\tilde{z}_3 + \tilde{z}_4, \\ \dot{\tilde{z}}_4 = \frac{b_1}{m_2}\tilde{z}_2 - \varpi_4\tilde{z}_3 - \frac{b_1+b_2}{m_2}\tilde{z}_4 + \frac{b_2}{m_2}\tilde{z}_6, \\ \dot{\tilde{z}}_5 = -\varpi_5\tilde{z}_5 + \tilde{z}_6, \\ \dot{\tilde{z}}_6 = \frac{b_2}{m_3}\tilde{z}_4 - \varpi_6\tilde{z}_5 - \frac{b_1+b_2}{m_3}\tilde{z}_6. \end{cases} \tag{61}$$

Select a Lyapunov candidate

$$L_6^o = \frac{\varpi_2 m_1}{2}\tilde{z}_1^2 + \frac{m_1}{2}\tilde{z}_2^2 + \frac{\varpi_4 m_2}{2}\tilde{z}_3^2 + \frac{m_2}{2}\tilde{z}_4^2 + \frac{\varpi_6 m_3}{2}\tilde{z}_5^2 + \frac{m_3}{2}\tilde{z}_6^2.$$

Its time derivative can be calculated by

$$\begin{aligned}\dot{L}_6^o &= -\varpi_1\varpi_2 m_1\tilde{z}_1^2 - b_1\tilde{z}_2^2 + b_1\tilde{z}_2\tilde{z}_4 - \varpi_3\varpi_4 m_2\tilde{z}_3^2 - (b_1+b_2)\tilde{z}_4^2 \\ &\quad + b_1\tilde{z}_2\tilde{z}_4 + b_2\tilde{z}_4\tilde{z}_6 \\ &\quad - \varpi_5\varpi_6 m_3\tilde{z}_5^2 + b_2\tilde{z}_4\tilde{z}_6 - (b_2+b_3)\tilde{z}_6^2 \\ &= -\varpi_1\varpi_2 m_1\tilde{z}_1^2 - b_1\left(\tilde{z}_2 - \tilde{z}_4\right)^2 - \varpi_3\varpi_4 m_2\tilde{z}_3^2 \\ &\quad - b_2(\tilde{z}_4 - \tilde{z}_6)^2 - b_3\tilde{z}_6^2 - \varpi_5\varpi_6 m_3\tilde{z}_5^2 \\ &\le 0\end{aligned}$$

where $\dot{L}_6^o = 0$ if and only if $\tilde{z} \triangleq [\tilde{z}_1, \tilde{z}_2, \tilde{z}_3, \tilde{z}_4, \tilde{z}_5, \tilde{z}_6]^T = 0$.

Consequently, the observation errors are globally asymptotically stable. It is apparent that (61) is a time-invariant linear system; therefore, the observation errors are globally exponentially stable. □

With the proposed observer (60), the un-measurable z_2, z_4 and z_6 can be re-constructed by $\hat{z}_2$, $\hat{z}_4$ and $\hat{z}_6$.

Remark 15. According to Lyapunov converse theorem [25], there exists a Lyapunov function L_6 for the exponentially stable linear system (61), such that

$$\delta_1^{ob}\|\tilde{z}\|^2 \le L_6^o(\tilde{z}) \le \delta_2^{ob}\|\tilde{z}\|^2,$$

$$\dot{L}_6^o(\tilde{z}) \le -\delta_3^{ob}\|\tilde{z}\|^2,$$

where $\delta_i^{ob} > 0,\ (i = 1, 2, 3)$.

Remark 16. More detailed information of tracking by using observers can be found in [28].

4.3. Stability of the closed-loop system with observers

With the observer (54) and (60), the un-measurable $\dot{k}, \ddot{k}, \dddot{k}, k^{(4)}$, z_2, z_4 and z_6 can be reconstructed, and the control algorithm can be summarized as following.

Algorithm 2.

(1) Calculate the virtual control α_1 with (20) and (28), where $\dot{k}$ should be replaced by $\hat{\dot{k}}$ in (56), and k is calculated by (53).
(2) Calculate the virtual control α_2 with (23) and (27)–(29), where $\dot{k}$ and $\ddot{k}$ should be replaced by $\hat{\dot{k}}$ in (56) and $\hat{\ddot{k}}$ in (57), respectively; and k is calculated by (53). z_2 is reconstructed by $\hat{z}_2$ in (60).
(3) Calculate the control α_3 with (25) and (33), where $\dot{e}_1$, $\dot{e}_2$, $\dot{z}_1$ and $\dot{z}_2$ are calculated with (34)–(37); $\ddot{\alpha}_1$ is calculated by (38)–(41); k is calculated by (53); $\dot{k}$, $\ddot{k}$ and $k^{(3)}$ are replaced by $\hat{\dot{k}}$ in (56), $\hat{\ddot{k}}$ in (57) and $\widehat{k^{(3)}}$ in (58), respectively; and z_2 and z_4 are reconstructed by $\hat{z}_2$ and $\hat{z}_4$ in (60).
(4) Calculate the control u with (25), where derivatives of virtual controls can be calculated by (43)–(50); k is calculated by (53); $\dot{k}, \ddot{k}, k^{(3)}, k^{(4)}$ are replaced by $\hat{\dot{k}}, \hat{\ddot{k}}, \widehat{k^{(3)}}, \widehat{k^{(4)}}$ in (56)–(59), respectively; and z_2, z_4 and z_6 are reconstructed by $\hat{z}_2, \hat{z}_4$ and $\hat{z}_6$ in (60).

Stability of the closed-loop system with uncertain parameters and unmeasurable states can be given by the following theorem.

Theorem 1. *Consider the pantograph–catenary system given by (1)–(4), where the elasticity coefficient model is unknown, and displacement variations $\dot{x}_1$, $\dot{x}_2$, and $\dot{x}_3$ are unmeasurable. Suppose that the reference contacting force is constant or continuously periodic. The control is designed by Algorithm 2, with time derivatives of $k(t)$ estimated by the differentiator (54), and with the unmeasurable z_2, z_4, and z_6 reconstructed by the observer (60). Then, tracking errors of the closed-loop system are ultimately bounded with tunable ultimate bounds, and (13) is satisfied.*

Proof. Consider that $\dot{k}$, $\ddot{k}$, $k^{(3)}$, $k^{(4)}$, z_2, z_4, and z_6 are reconstructed by (54) and (60), respectively. It follows that the tracking error dynamics can be given by

$$\begin{cases}\dot{e}_1 = -c_1e_1 + e_2 + o_1(\tilde{z}, \tilde{\zeta}), \\ \dot{e}_2 = -e_1 - c_2e_2 + e_3 + o_2(\tilde{z}, \tilde{\zeta}), \\ \dot{e}_3 = -e_2 - c_3e_3 + e_4 + o_3(\tilde{z}, \tilde{\zeta}), \\ \dot{e}_4 = -e_3 - c_4e_4 + o_4(\tilde{z}, \tilde{\zeta}),\end{cases} \tag{62}$$

where $\tilde{\zeta} \triangleq [\tilde{\zeta}_1, \tilde{\zeta}_2, \tilde{\zeta}_3, \tilde{\zeta}_4, \tilde{\zeta}_5]^T$; $o_1(\tilde{z})$, $o_2(\tilde{z})$, $o_3(\tilde{z})$ and $o_4(\tilde{z})$ are errors resulted from differentiator errors and observation errors, and they satisfy

$$o_1(0,0) = 0,\ o_2(0,0) = 0,\ o_2(0,0) = 0,\ o_2(0,0) = 0.$$

Since $\dot{k}$, $\ddot{k}$, $k^{(3)}$, $k^{(4)}$ are continuously bounded, and $\tilde{\zeta}$ and $\tilde{z}$ are bounded, there exist positive κ_{ij} $(i = 1, 2, 3, 4,\ j = 1, 2)$ such that the following expressions hold locally:

$$\begin{aligned}&\|o_1(\tilde{z}, \tilde{\xi})\| \le \kappa_{11}\|\tilde{z}\| + \kappa_{12}\|\tilde{\zeta}\|, \\ &\|o_2(\tilde{z}, \tilde{\xi})\| \le \kappa_{21}\|\tilde{z}\| + \kappa_{22}\|\tilde{\zeta}\|, \\ &\|o_3(\tilde{z}, \tilde{\xi})\| \le \kappa_{31}\|\tilde{z}\| + \kappa_{32}\|\tilde{\zeta}\|, \\ &\|o_4(\tilde{z}, \tilde{\xi})\| \le \kappa_{41}\|\tilde{z}\| + \kappa_{42}\|\tilde{\zeta}\|.\end{aligned} \tag{63}$$

It follows that the time derivative of L_4 can be calculated by

$$\begin{aligned}\dot{L}_4 &= -c_1e_1^2 - c_2e_2^2 - c_3e_3^2 - c_4e_4^2 + e_1o_1 + e_2o_2 + e_3o_3 + e_4o_4 \\ &\le -\sum_{i=1}^{4}\left(c_ie_i^2 + \kappa_{i1}\|e_i\tilde{z}\| + \kappa_{i2}\|e_i\tilde{\zeta}\|\right).\end{aligned}$$

Select the Lyapunov candidate $L_7 = L_4 + \gamma_d L_5 + \gamma_{ob} L_6$ with $\gamma_d > 0$ and $\gamma_{ob} > 0$. Its time derivative can be calculated by

$$\begin{aligned}\dot{L}_7 &\le \sum_{i=1}^{4}\left(-c_ie_i^2 + \kappa_{i1}\|e_i\tilde{z}\| + \kappa_{i2}\|e_i\tilde{\zeta}\|\right) \\ &\quad - \gamma_{ob}\delta_3^{ob}\|\tilde{z}\|^2 - \gamma_d\delta_3^d\|\tilde{\zeta}\|^2 + \gamma_d\beta_d \\ &= -\sum_{i=1}^{4}(c_i - 2\eta_i)e_i^2 + \gamma_d\beta_d - \sum_{i=1}^{4}\left(\gamma_{ob}\delta_3^{ob} - \frac{\kappa_{i1}^2}{4\eta_i}\right)\tilde{z}_i^2 \\ &\quad - \sum_{i=1}^{4}\left(\sqrt{\eta_i}e_i - \frac{\kappa_{i1}}{2\sqrt{\eta_i}}\tilde{z}\right)^2 \\ &\quad - \sum_{i=1}^{4}\left(\gamma_d\delta_3^d - \frac{\kappa_{i2}^2}{4\eta_i}\right)\tilde{\zeta}_i^2 - \sum_{i=1}^{4}\left(\sqrt{\eta_i}e_i - \frac{\kappa_{i2}}{2\sqrt{\eta_i}}\tilde{\zeta}\right)^2 \\ &\le -\sum_{i=1}^{4}(c_i - 2\eta_i)e_i^2 + \gamma_d\beta_d - \sum_{i=1}^{4}\left(\gamma_{ob}\delta_3^{ob} - \frac{\kappa_{i1}^2}{4\eta_i}\right)\tilde{z}_i^2 \\ &\quad - \sum_{i=1}^{4}\left(\gamma_d\delta_3^d - \frac{\kappa_{i2}^2}{4\eta_i}\right)\tilde{\zeta}_i^2\end{aligned}$$

where $0 < 2\eta_i < c_1$ $(i = 1, 2, 3)$; $\gamma_{ob} > 0$ and $\gamma_d > 0$ can be selected large enough, such that $\gamma_{ob}\delta_3^{ob} - \frac{\kappa_{i1}^2}{4\eta_i} > 0$ and $\gamma_d\delta_3^d - \frac{\kappa_{i2}^2}{4\eta_i} > 0$.

Then, it can be claimed that L_7 satisfies

$$\delta_1\|\bar{e}\|^2 \le L_7 \le \delta_2\|\bar{e}\|^2, \tag{64}$$

$$\dot{L}_7 \le -\delta_3\|\bar{e}\| + \gamma_d\beta_d(\dot{k}, \ddot{k}, k^{(3)}, k^{(4)}), \tag{65}$$

where $\bar{e} \triangleq [e^T, \tilde{z}^T, \tilde{\zeta}^T]^T$, and

$$\delta_1 = \min\left[\frac{1}{2}, \gamma_{ob}\delta_1^{ob}, \gamma_d\delta_1^d\right], \quad \delta_2 = \max\left[\frac{1}{2}, \gamma_{ob}\delta_2^{ob}, \gamma_d\delta_2^d\right],$$

$$\delta_3 = \min_{i=1,2,3,4}\left[c_i - 2\eta_i,\ \gamma_{ob}\delta_3^{ob} - \frac{\kappa_{i1}^2}{4\eta_i},\ \gamma_d\delta_3^d - \frac{\kappa_{i2}^2}{4\eta_i}\right].$$

Consequently, it can be concluded that $\bar{e}$ is ISS with respect to $k^{(i)}$ $(i = 1, 2, 3, 4)$.

Since $k(t)$ is periodic, it follows that $k^{(i)}$ $(i = 1, 2, 3, 4)$ are periodic and bounded, and it holds that

$$\beta_d(\dot{k}, \ddot{k}, k^{(3)}, k^{(4)}) \le \bar{\beta}_d, \tag{66}$$

where $\bar{\beta}_d > 0$ denotes the bound of β_d. It can be solved from (64) and (65) that

$$L_7(t) \le e^{-\frac{\delta_3}{\delta_2}t}\left(L_7(0) - \frac{\delta_2\gamma_d\bar{\beta}_d}{\delta_3}\right) + \frac{\delta_2\gamma_d\bar{\beta}_d}{\delta_3},$$

and therefore,

$$\|e_1\| \le \sqrt{\frac{1}{\delta_1}e^{-\frac{\delta_3}{\delta_2}t}\left(L_7(0) - \frac{\delta_2\gamma_d\bar{\beta}_d}{\delta_3}\right) + \frac{\delta_2\gamma_d\bar{\beta}_d}{\delta_1\delta_3}}, \tag{67}$$

which can be tuned by assigning appropriate δ_i $(i = 1, 2, 3)$.

Moreover, according to Proposition 5 in Appendix, z_3 (and z_5) tracks a periodic trajectory z_{3r} (and z_{5r}) asymptotically, and its tracking error $e_3^z \triangleq z_3 - z_{3r}$ ($e_5^z \triangleq z_5 - z_{5r}$) satisfies (71) in Appendix, where $\mathcal{L}_3(0) = 0$ and $\|\mathcal{L}_3(\bar{e})\| \le \beta_3^z\|\bar{e}\|$ with $\beta_3^z > 0$.

Select a Lyapunov candidate $L_0^{ob} = L_7 + \frac{1}{2\gamma_3^z}{e_3^z}^2 + \frac{1}{2\gamma_5^z}{e_5^z}^2$ for the full-state closed-loop system with observers, where $\gamma_3^z > 0$ and $\gamma_5^z > 0$. Its time derivative can be calculated by

$$\begin{aligned}
\dot{L}_0^{ob} &\le -\delta_3\|\bar{e}\|^2 - \frac{k_1}{b_1\gamma_3^z}{e_3^z}^2 + \frac{\beta_3^z}{\gamma_3^z}e_3^z\|\bar{e}\| \\
&\quad - \frac{k_2}{b_2\gamma_5^z}{e_5^z}^2 + \frac{\beta_5^z}{\gamma_5^z}e_5^z\|\bar{e}\| + \gamma_d\bar{\beta}_d \\
&= -\left(\delta_3 - \frac{{\beta_3^z}^2 b_1}{4k_1\gamma_3^z} - \frac{{\beta_5^z}^2 b_2}{4k_2\gamma_5^z}\right)\|\bar{e}\|^2 + \gamma_d\bar{\beta}_d \\
&\quad - \left(\sqrt{\frac{k_1}{b_1\gamma_3^z}}e_3^z - \frac{\beta_3^z}{2}\sqrt{b_1k_1\gamma_3^z}\|\bar{e}\|\right)^2 \\
&\quad - \left(\sqrt{\frac{k_2}{b_2\gamma_5^z}}e_5^z - \frac{\beta_5^z}{2}\sqrt{b_2k_2\gamma_5^z}\|\bar{e}\|\right)^2
\end{aligned}$$

where γ_3^z and γ_5^z can be selected large enough such that $\left(\delta_3 - \frac{{\beta_3^z}^2 b_1}{4k_1\gamma_3^z} - \frac{{\beta_5^z}^2 b_2}{4k_2\gamma_5^z}\right) > 0$. As a consequence, tracking errors of the closed-loop system with the proposed control and observers are ultimately bounded with tunable ultimate bounds. □

5. Simulations and discussion

In the simulations, values of parameters of the pantograph–catenary system are taken from [6,9,11], as listed in Table 1. The train speed is set to $V = 90$ m/s to test performances of the closed-loop system with high speed. The reference contact force is set by 100N. Initial values of system states are given by

$$[x_1(0), \dot{x}_1(0), x_2(0), \dot{x}_2(0), x_3(0), \dot{x}_3(0)]^T$$

$$= [0.005, 0, 0.01, 0, 0.01, 0]^T.$$

Table 1
Values of parameters.

Notations	Values	Notations	Values
k_1	7015.9 Nm^{-1}	L	65 m
m_1	8 kg	m_2	12 kg
b_1	120 Nsm^{-1}	b_2	30 Nsm^{-1}
V	90 ms^{-1}	K_0	7000 Nm^{-1}
K_1	3360 Nm^{-1}	K_2	650 Nm^{-1}
K_3	160 Nm^{-1}	K_7	160 Nm^{-1}
k_2	1550.1 Nm^{-1}		

Table 2
Values of control gains and observer gains.

Notations	Values	Notations	Values
c_1	12	a_1	128
c_2	36	a_2	128
c_3	108	a_3	64
c_4	324	a_4	32
R	200	a_5	4
ϖ_1	20	ϖ_2	4
ϖ_3	20	ϖ_4	4
ϖ_5	20	ϖ_6	4

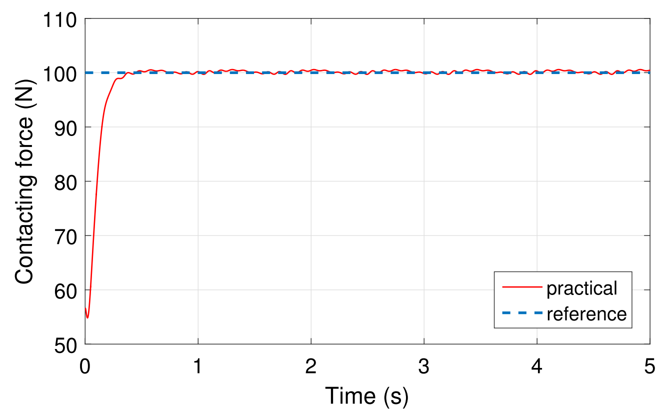


Fig. 3. Contact force with full-state feedback control: the tracking error is globally exponentially stable.

Suppose that elasticity coefficient model is fully known in priori, and z_2, z_4 and z_6 are measurable. In this case, Algorithm 1 is applied, with control gains listed in Table 2. The tracking performance of the closed-loop system with respect to a constant contacting force is illustrated by Fig. 3. As can be seen, the tracking error is globally asymptotically stable, and the transient process is satisfactory.

In more practical applications, the accurate model of elasticity coefficient $k(t)$ is unknown, and z_2, z_4 and z_6 are unmeasurable, implying that the elasticity coefficient model (4), as well as the unmeasurable z_2, z_4, and z_6, cannot be used directly in control design. In this case, Algorithm 2 is applied with control gains, differentiator parameters and observer gains listed in Table 2. Initial values of observer states are all set to zeros. It can be seen from Fig. 4 that, with the proposed partial-state feedback control algorithm, the closed-loop system is capable of tracking the reference contacting force with ultimately bounded tracking errors. The displayed bounded tracking is in significant accordance with the theoretical results. It can be seen from Figs. 5–7 that reconstructed signals $\hat{z}_2$, $\hat{z}_4$, $\hat{z}_6$ are capable of tracking their actual values exponentially. The control signal is displayed in Fig. 8, where it can be seen that the controller is fairly implementable.

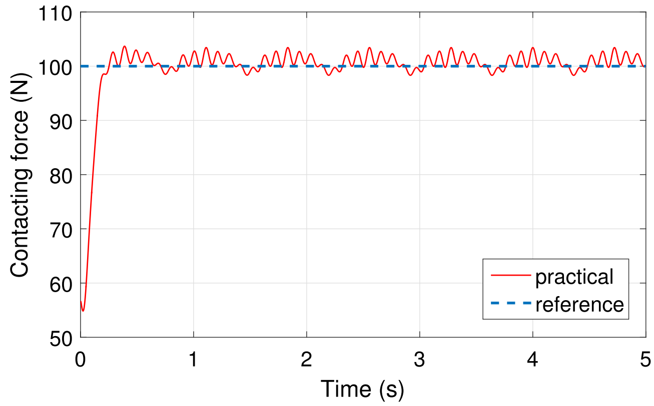


Fig. 4. Contact force with the proposed partial-state feedback control: the tracking error is ultimately bounded.

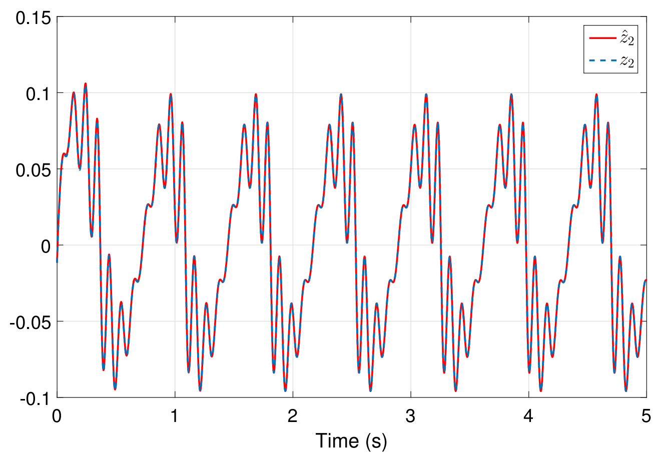


Fig. 5. Observed z_2: the observation error is globally exponentially stable.

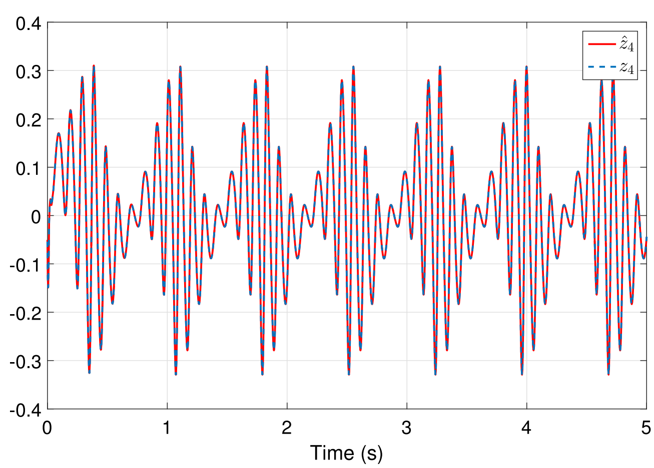


Fig. 6. Observed z_4: the observation error is globally exponentially stable.

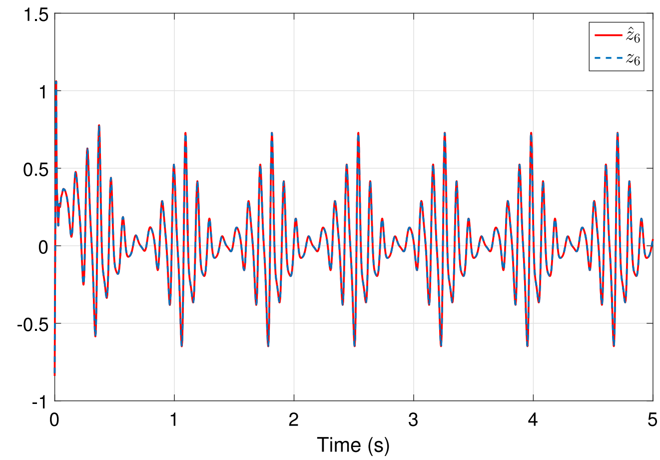


Fig. 7. Observed z_6: the observation error is globally exponentially stable.

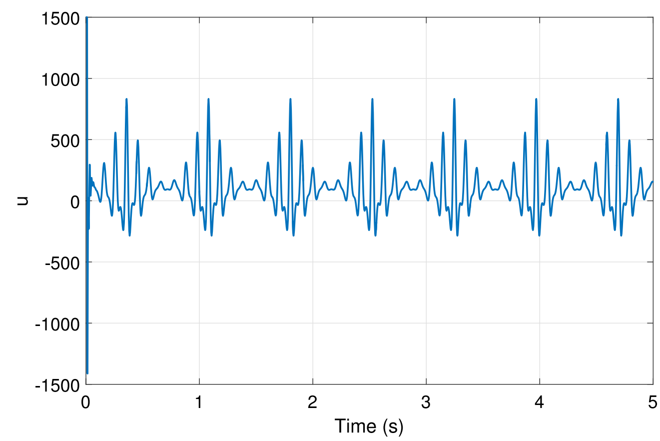


Fig. 8. Control signal of the closed-loop system with the proposed partial-state feedback.

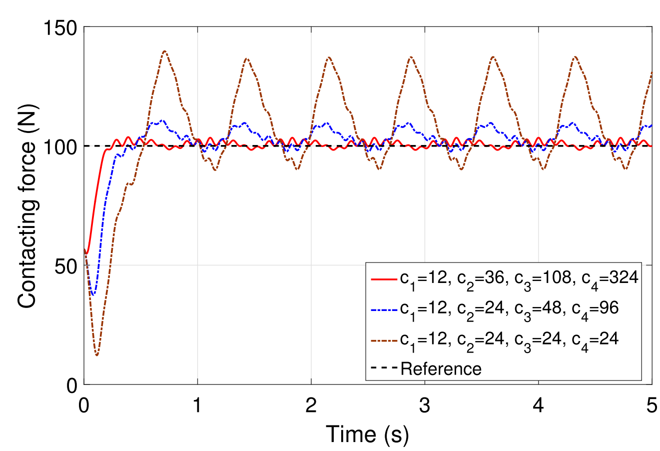


Fig. 9. Comparison of closed-loop performances with different control gains.

The ultimate bound in (67) cannot be calculated explicitly, since it is related to the differentiator error β_d in (65). The train is supposed to be operated in very high speed (90 m/s in this simulation), such that the frequency of periodic catenary stiffness $k(t)$ is very high, and the tracking error of the differentiator would be considerably large. The existence of β_d is obvious; however, its value is difficult to be determined explicitly. Even though we cannot calculate the particular value of ultimate bound explicitly, it can be tuned by control parameters. According to (67), δ_1, δ_2 and δ_3 are related directly to control gains c_1, c_2, c_3 and c_4. To illustrate, a comparison of closed-loop performance with different control gains is given in Fig. 9, where it can be seen that large control gains would lead to smaller ultimate bounds.

Remark 17. In simulation, the closed-loop system is not ideally continuous. Both the model and the controller are discretized with small sampling intervals. Consequently, the control gains cannot be increased to extremely large values to reduce the chattering. If the control gains are extremely large and the trains speed is high, there would be stability problems due to the discretization.

6. Conclusion

In this paper, a nonlinear partial-state feedback control is proposed for a 3-DOF pantograph–catenary system, such that the contact force between pantograph and catenary can track a continuous reference force. The proposed control is designed based on backstepping approach, where time derivatives of virtual controls are calculated explicitly. A high-order differentiator is designed for estimating derivatives of time-varying elasticity coefficient, and an observer is designed to reconstruct the unmeasurable spring velocities. Ultimate boundedness of tracking errors of the closed-loop system with proposed control and observer is proved rigorously. Theoretical results are demonstrated by numerical simulation.

It should be noted that the approach proposed in this paper is open for further extensions (for example, adaptive control in case of parametric uncertainties. For more details, please see the canonical design process in [24]).

Acknowledgments

This research is supported by National Natural Science Foundation of China under Grant No. 61703018, and the Young Elite Scientists Sponsorship Program by CAST, China under Grant 2017QNR C001.

Conflict of interest

There is no conflict of interest.

Appendix. Tracking performance of z_3 in Proposition 1 and Theorem 1

Proposition 4. *Suppose that the reference contacting force is constant or continuously periodic. Then, in the closed-loop system with Algorithm 1, z_3 (and z_5) tracks a periodic trajectory asymptotically.*

Proof. It follows from definition of e_3 that

$$\begin{aligned}\xi_1 &= e_3 + \alpha_2 \\ &= e_3 - e_1 - c_2 e_2 + \frac{k_1 + k}{m_1}(e_1 + z_{1r}) + \frac{b_1}{m_1}(e_2 + \alpha_1) + \dot{\alpha}_1 \\ &= \mathcal{L}_1(e) + \frac{k_1}{m_1} z_{1r} + \frac{1}{m_1} y_r + \frac{b_1}{m_1}\alpha_1 + \dot{\alpha}_1 \\ &= \mathcal{L}_1(e) + \mathcal{P}_1 + \frac{b_1}{m_1}(-c_1 e_1 + \dot{z}_{1r}) - c_1 \dot{e}_1 + \ddot{z}_{1r} \\ &= \mathcal{L}_2(e) + \mathcal{P}_2,\end{aligned} \tag{68}$$

where $\mathcal{L}_1(e)$ and $\mathcal{L}_2(e)$ are linear combinations of e_1, e_2 and e_3, and it holds that $\|\mathcal{L}_2(e)\| \le \kappa_3^z \|e\|$ with some certain $\kappa_3^z > 0$; $\mathcal{P}_1$ and $\mathcal{P}_2$ are continuously periodic terms.

Let z_{3r} be the solution of the following system:

$$\dot{z}_{3r} = -\frac{k_1}{b_1} z_{3r} + \frac{1}{b_1}\mathcal{P}_2, \tag{69}$$

which is a stable linear time-invariant system plus a periodic input. It is apparent that z_{3r} is ultimately periodic.

Let $e_3^z \triangleq z_3 - z_{3r}$. It follows from (17), (68) and (69) that

$$\dot{e}_3^z = -\frac{k_1}{b_1} e_3^z + \frac{1}{b_1}\mathcal{L}_2(e), \tag{70}$$

where e decreases exponentially according to (52). Consequently, it can be claimed that $e_3^z \to 0$, indicating that z_3 tracks a periodic trajectory asymptotically.

With similar steps, it can be proved that z_5 tracks a periodic trajectory asymptotically. □

Proposition 5. *Suppose that the reference contacting force is constant or continuously periodic. Then, in the closed-loop system with Algorithm 2, z_3 (and z_5) tracks a periodic trajectory asymptotically.*

Proof. The proof is similar to that of Proposition 4. It can be proved that ζ_1 can be expressed as the sum of $\mathcal{L}_3(\bar{e})$ and $\mathcal{P}_3$, where $\mathcal{L}_3(\bar{e})$ is a linear combination of e and $\tilde{z}$, and $\mathcal{P}_3$ is composed by continuously periodic terms. It holds that $\|\mathcal{L}_3(\bar{e})\| \le \beta_3^z \|\bar{e}\|$ with a certain $\beta_3^z > 0$. It follows that

$$\dot{e}_3^z = -\frac{k_1}{b_1} e_3^z + \mathcal{L}_3(\bar{e}), \tag{71}$$

and $e_3^z \to 0$ (since $\bar{e}$ vanishes), indicating that z_3 tracks a periodic trajectory asymptotically.

With similar steps, it can be proved that z_5 tracks a periodic trajectory asymptotically. □

References

[1] Jimenez-Octavio JR, Sanchez-Rebollo C, Carnicero A. The dependance on mechanical design in railway electrification. IEEE Electrif Mag 2013;1(1):4–10.
[2] Ambrosio J, Pombo J, Pereira M, Antunes P, Mosca A. A computational procedure for the dynamic analysis of the catenary-pantograph interaction in high-speed trains. J Theoret Appl Mech 2012;50(3):681–99.
[3] Chen Z, Wang T, Hui L, Guo F. Determination of the optimal contact load in pantograph–catenary system. Trans China Electrotech Soc 2013;28(6):86–92 [in Chinese].
[4] Sanchez-Rebollo C, Jimenez-Octavio JR, Carnicero A. Active control strategy on a cantenary-pantograph validated model. Veh Syst Dyn 2013;51(4):554–69.
[5] Lin Y, Lin C, Yang C. Robust active vibration control for rail vehicle pantograph. IEEE Trans Veh Technol 2007;56(4):1994–2004.
[6] Ide CK, Olaru S, Rodriguez-Ayerbe P, Rachid A. A nonlinear state feedback control approach for a Pantograph-Catenary system. In: Proceedings of the 17th international conference on system theory, control and computing. IEEE; 2013, p. 268–73.
[7] Pisano A, Usai E. Output-feedback regulation of the contact force in high-speed train pantographs. J Dyn Syst Meas Control 2004;126(1):82–7.
[8] Chater E, Ghani D, Giri F, Haloua M. Output feedback control of pantograph–catenary system with adaptive estimation of catenary parameters. J Modern Transp 2015;23(4):252–61.
[9] Taran M, Rodriguez-Ayerbe P, Olaru S, Ticlea A. Moving horizon control and estimation of a pantograph-Catenary system. In: Proceedings of the 17th international conference on system theory, control and computing. IEEE; 2013, p. 527–32.
[10] Balestrino A, Bruno O, Landi A, Sani L. Innovative solutions for overhead catenary-pantograph system: wire acturated control and observerd contact force. Veh Syst Dyn 2000;33(2):69–89.
[11] Pisano A, Usai E. Contact force estimation and regulation in active pantographs: an algebraic observability approach. Asian J Control 2011;13(6):761–72.
[12] Allotta B, Pugi L, Bartolini F. Design and experimental results of an active suspension system for a high-speed pantograph. IEEE/ASME Trans Mechatronics 2008;13(5):548–57.
[13] Rachid A. Pantograph catenary control and observation using the LMI approach. In: Proceedings of the 50th IEEE conference on decision and control and European control conference. IEEE; 2011, p. 2287–92.
[14] Chen M, Ward CP, Hubbard EM, Hubbard P. Modelling and active control design of trolleybus catenary-pantograph system. IFAC-PapersOnLine 2016;49–21:282–7.
[15] Song Y, Ouyang H, Liu Z, Mei G, Wang H, Lu X. Active control of contact force for high-speed railway pantograph–catenary based on multi-body pantograph model. Mech Mach Theory 2017;115:35–59.

[16] Mokrani N, Rachid A, Rami MA. A tracking control for pantograph–catenary system. In: Proceedings of 2015 IEEE 54th annual conference on decision and control (CDC). Osaka (Japan); 2015. p. 185–90.
[17] Allotta B, Papi M, Pugi L, Toni P, Violi AG. Experimental campaign on a servo-actuated pantograph. In: Proceedings of international conference on advanced intelligent mechatronics. IEEE/ASME; 2001, p. 237–42.
[18] Cho YH. Numerical simulation of the dynamic responses of railway overhead contact lines to a moving pantograph, considering a nonlinear dropper. J Sound Vib 2008;315:433–54.
[19] Bucca G, Collina A. A procedure for the wear prediction of collector strip and contact wire in pantograph–catenary system. Wear 2009;266:46–59.
[20] Lim SC, Ashby MF. Wear-mechanicsm maps. Acta Metall 1987;35(1):1–24.
[21] Holm R, Holm E. Electric contacts handbook. Berlin: Springer-verlag; 1958.
[22] Levant A, Pisano A, Usai E. Output-feedback control of the contact-force in high-speed-train pantograph. In: Proceedings of the 40th IEEE conference on decision and control. Orlando (USA); 2001. p. 1831–6.
[23] British Standards Institution. Railway applications - current collection systems - validation of simulation of the dynamic interaction between pantograph and overhead contact line. London, British Standards Institution; 2002, BS EN 50318:2002.
[24] Krstic M, Kanellakopoulos I, Kokotovic P. Nonlinear and adaptive control design. Inc.: John Wiley & Sons; 1995.
[25] Khalil H. Nonlinear systems. 3rd ed. New Jercey: Prentice Hall Inc.; 2002.
[26] Bu X, Wu X, Zhang R, Ma Z, Huang J. Tracking differentiator design for the robust backstepping control of a flexible air-breathing hypersonic vehicle. J Franklin Inst B 2015;352(4):1739–65.
[27] Wang X, Chen Z, Yuan Z. Design and analysis for new discrete tracking-differentiators. Appl Math 2003;18(2):214–22.
[28] Nunes EVL, Peixoto AJ, Oliveira TR, Hsu L. Global exact tracking for uncertain MIMO linear systems by output feedback sliding mode control. J Franklin Inst B 2014;351(4):2015–32.

Applied Energy 242 (2019) 393–402

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

A hierarchical predictive control for supercapacitor-retrofitted grid-connected hybrid renewable systems

Mukalu Sandro Masaki, Lijun Zhang*, Xiaohua Xia

Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa

HIGHLIGHTS

- A hierarchical model predictive controller is developed.
- Little to no modification is required on the architecture of the existing system.
- A stable power exchange between the renewable system and the grid is achieved.
- Fast variations are completely removed from the battery power.
- Increased utilization of intermittent renewable energy is achieved.

ARTICLE INFO

Keywords:
Power smoothing
Hybrid energy storage system
Supercapacitor
Model predictive control
Renewable energy

ABSTRACT

This paper presents a two-layer control strategy designed for easy integration of supercapacitors in a grid-integrated solar photovoltaic-battery hybrid renewable system, initially controlled by a typical model predictive control method. To operate the upgraded energy system, either without or with little modifications of the pre-existing architecture, an additional control layer is applied at the bottom of the original control system. Considering the complementary characteristics of batteries and supercapacitors, the design of the new model predictive control layer and its coordination with the original one help to deliver a stable power flow between the hybrid renewable system and the utility grid, and remove fast variations from the battery power. Actual measurements of solar radiation in South Africa are used to test the effectiveness of the proposed strategy. Simulations carried out on a 1-MW photovoltaic plant confirm the benefits in terms of adherence to power quality regulations, improved conditioning of the power generated by the intermittent renewable sources, and lifetime extension of the battery.

1. Introduction

For more than a decade, grid-integration of intermittent renewables such as wind turbines and solar photovoltaics (PVs) has proven to be an effective means to achieve progressive decarbonization of power systems [1–3]. However, the increasing penetration of weather-dependent generation poses risks to the reliability, stability, and economy of power supply [4,5]. Among the proven solutions to this concern is the inclusion of energy storage systems (ESSs) such as a battery, flywheel, supercapacitor, superconducting magnetic energy storage, fuel-cell and pumped hydro [6–8].

Until recently, batteries were one of the most popular ESS due to their high energy density, flexibility and scalability [8,9]. Accordingly, a multitude of control strategies for grid-integrated hybrid renewable systems (HRS) with a battery ESS were proposed in the literature. Besides the routine energy and power constraints applied to batteries, a few models were also presented to reduce wear of battery caused by the current flow. This includes minimizing a battery aging factor [10], penalizing the charge and discharge operations [11], keeping the state of health of batteries above a threshold [12], and implementing state of charge (SoC)-oriented control of the battery current [13]. The battery wear process can be hastened through fast/large variations of current flow that generate excessive heat and increase the internal resistance of the battery, causing further heating by the Joule's effect [14–16]. Presently, owners of grid-integrated renewable energy systems are being increasingly required by regulatory authorities to maintain a stable power profile at the point of common coupling (PCC) to the power grid [17,18]. Therefore, despite the existence of wear control schemes, must handle the fluctuations introduced by intermittent renewable sources such as PV panels.

* Corresponding author.
E-mail address: lijun.zhang@up.ac.za (L. Zhang).

https://doi.org/10.1016/j.apenergy.2019.03.049
Received 27 August 2018; Received in revised form 5 February 2019; Accepted 7 March 2019

A cost-effective method to address the need for an EES that features fast response and long-term energy support is to combine two or more energy storage technologies into a hybrid ESS. Among the various possible options, the battery-supercapacitor (SC) association is presently preferred for power supply due to feasibility and maturity reasons [19–22]. Batteries have a high energy density, but low power density and slow response speed. These characteristics are well balanced by the high power density, rapid response speed and low energy density of SCs. The power peak enhancement, internal losses reduction and lifetime extension achieved by a passive battery-SC hybrid ESS (energy storage devices directly mounted in parallel) were previously established under pulsed load [23] and pulsed charging sources [24]. However, in these works, the current sharing between the two energy storage devices is only determined by their internal voltages and internal resistances. The power flows of the battery and the SC cannot be controlled separately since their terminal voltages are forced to be equal at all times.

Various control strategies for power sharing between the energy storage components of a battery-SC ESS have been proposed in the literature. DC-bus voltage-based control schemes that use the battery to regulate the power balance in the DC grid and the SC to handle fast DC-bus voltage dynamics were presented in [25]. A model predictive controller (MPC) for battery-SC ESS that aims at supplying/absorbing the power allocated to the hybrid ESS, while directing the fast and slow current components to the SC and battery, respective, was proposed in [26]. Energy losses in the SC result in increased stress levels for the battery during power supply. An MPC scheme that minimizes the magnitude/fluctuation of the battery current and the energy loss seen in the SC was provided in [14]. A heuristic algorithm using a modified active parallel hybrid ESS, with SC-only connected through the DC/DC converter, was presented in [27]. In this strategy, the paths and directions of power flow within the battery-SC ESS are determined by a number of factors, including the power balance requirement, the terminal voltages of energy storage devices and the battery SoC. A variable two-stage rate-limit scheme for batteries was presented in [28]. Two different rate-limits are designed to optimize the charge/discharge rates and the amount of energy stored/released by the battery, taking the load requirement and the settling time into account. The SC is used to complement the battery during the transient period. Because only a few components of the energy system were involved in the design, these controllers achieve local energy management.

Control strategies that aim to coordinate the power flow across the entire renewable energy system have been proposed in the literature. A rule-based power management scheme for the dispatch of a PV power plant in compliance with the Australian grid regulation was proposed in [18]. An improved model, more robust against forecasting errors, was presented in [29]. A model predictive heuristic control that regulates the charge/discharge power of the battery was presented in [30]. In this study, wavelet theory is used to achieve a multi-layer decomposition of the power output of a wind generator. A control scheme with dynamic rate limiter designed for grid-connected wave energy park was provided in [31]. The dynamic rate limiter allows direct control over the magnitude of power variations of the battery. Ref. [32] proposed an energy management framework for a grid-integrated concentration photovoltaic plant. A second-order filter is developed for power allocation between the hybrid ESS components. A hierarchical dynamic optimal model for real-time tracking of the grid power reference was proposed in [33]. These strategies mainly focus on the control of utility-scale renewable plants, with little attention paid to the case of smaller systems, where the presence of local demand plays a significant role in the definition of operational objectives.

Among the previous research on this topic, a hierarchical energy management framework for multiple distributed PV-SC-load HRS with a centralized battery ESS was developed in [34]. First-order filters allow to allocate the high-frequency power components of the net power to the SC on site, while the low-frequency power components are direct to the centralized battery ESS. This ESS component also helps to maintain the power balance at the PCC. [35] proposed a heuristic designed to regulate the DC-bus voltage and smooth the power profile at the PCC. The power allocation between the ESS components is mainly determined by their SoC. A heuristic algorithm that regulates the DC-bus voltage, and the grid voltage and frequency, taking the grid availability and the electricity price into consideration, was presented in [36]. A heuristic algorithm that realizes the automatic selection of the operation mode of the battery among the pre-set modes was presented in [37]. While the SC is directly connected on the DC bus, the suitable operation mode, which depends on the direction and amplitude of the battery current, is chosen on the basis of the PV power output, load demand and the battery SoC. A dynamic power sharing of excess and deficit powers between the grid and the EES components of a PV-based HRS be means of a heuristic algorithm was proposed in [38]. The power allocation considers the sign and magnitude of the net power, the SoC of the battery and the SC.

The main drawback of the various methods presented so far is that, in the case of pre-existing HRS equipped with batteries, a complete restructuring of the control system is required in order to implement the new controller for battery-SC ESS. This may raise concerns from plant owners with respect to technical (shut-down, decommissioning process), financial (decommissioning and disposal costs) and environmental (disposal of material) implications. In the literature, available to the authors, only a filtration-based control strategy for hybrid ESS retrofit in autonomous PV/battery domestic HRS was proposed in [39]. In this control model, the high-frequency components filtered from the measured battery current are used as the set-point for the SC, which absorbs from/feeds into the common bus through a DC-DC converter. Immediate benefits, particularly, decreased fluctuations in the battery current and reduction in battery health cost, were reported in this study. While the proposed solution can be also applied to grid-integrated PV-battery HRS, significant fluctuations remained in the battery current profile despite the presence of SC. The failure to account for the predictions of the battery current during the control of the SC can be cited among the reasons for this situation. Moreover, no control was conducted over the current flow and the energy level of the SC. Finally, by focusing on the battery alone, the SC provides no substantial advantage at the system level.

This paper presents a hierarchical predictive control of a grid-integrated PV/battery HRS retrofitted with SC. Under the existing system considered in this study, the control strategy presented in this paper has the following advantages: (1) no re-programming is required on the existing controller, since the SC is supervised by a new controller added in the control system; (2) reduces sudden variations in the power flow of the battery; (3) increased utilization rate of the renewable energy; (4) Stable power flow at the PCC; and (5) better tracking of the grid power reference.

2. Modelling of the grid-integrated hybrid power system

The HRS evaluated in this paper is illustrated in Fig. 1, where the shaded area indicates the retrofitted equipment. It comprises solar PV panels, a battery bank, a supercapacitor bank and loads. Electric power can be absorbed from or fed into the utility grid at the PCC. Before the retrofit, the power management unit 1 (PMU-1) ensures the control of the solar PV and the battery via their respective DC-DC converters and the utility grid via an AC-DC inverter. A circuit breaker (CB) allows PMU-1 to connect and disconnect the power network. After the addition of the new equipment, the command signal of the PV and the grid is transferred to PMU-2, while the measurements are sent to both control units. Only the battery remains under control of PMU-1. Both PMUs are supplied with forecast data of the load demand and the PV generation. In practice, a single PMU, with sufficient computing power and memory resources, can play the role of PMU-1 and PMU-2. In that case, no modification is required to the architecture of the original control

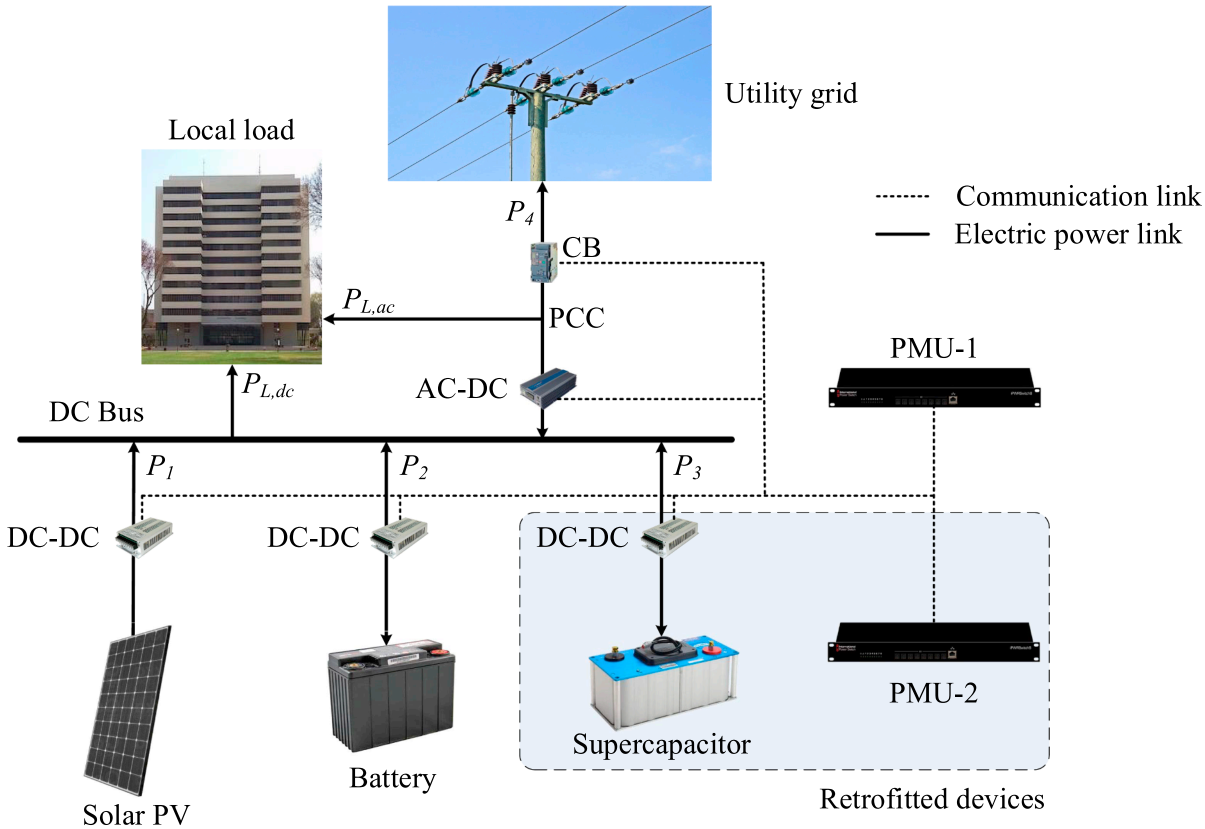


Fig. 1. Layout of the grid-integrated HRS.

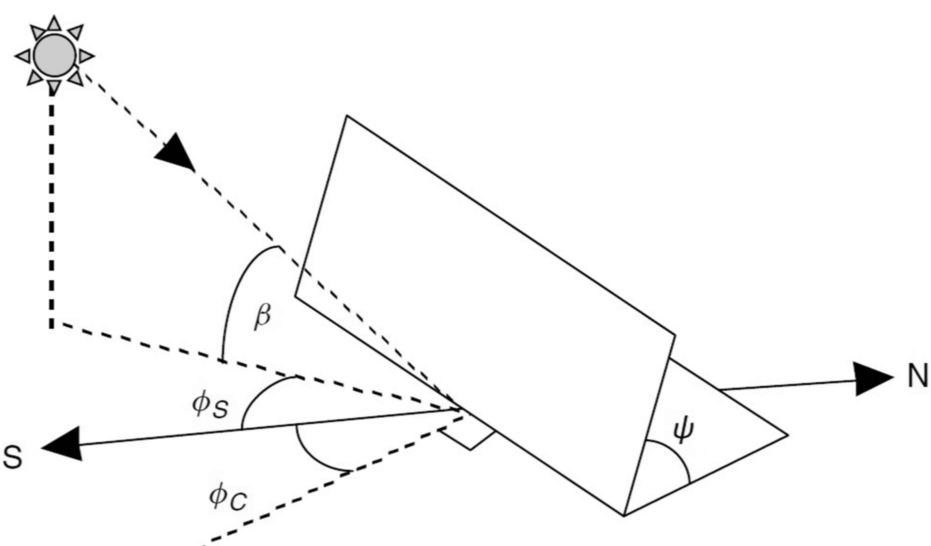


Fig. 2. Collector azimuth angle $\phi_S(k)$, tilt angle ψ, solar azimuth angle $\phi_S(k)$ and altitude angle $\beta(k)$.

system.

In this study, we assume that the original HRS is controlled by an MPC scheme, designed for the purpose of either the minimization or the maximization of an objective. An example might be to increase the self-sufficiency and to encourage the use of solar energy, which implies minimizing the power consumption from the grid and maximizing the power supply by the PV system. In this context, because of both the relatively "large" control step (affecting the sampling rate of forecast data), necessary to prevent rapid variations of the battery power, and the limited degree of freedom offered by the HRS at this stage, little attention can be paid to the actual power quality at the PCC and the battery. The proposed retrofit with the SC aims to address this problem and to further increase the use of solar energy. The mathematical models of the HRS components are presented below.

2.1. Solar photovoltaic system

The solar PV system consists of solar arrays that harvest solar radiation and convert it into DC power. The solar radiation that strikes the surface of collectors of a solar panel has three components: direct-beam radiation, diffuse radiation and reflected radiation.

The PV power output $P_{pv,BC}$ due to beam radiation $I_{BC}(k)$ that strikes the active surface of the PV panel at sample time k is given by [40]:

$$P_{pv,BC}(k) = \eta_{pv} I_{BC}(k) A_c, \tag{1}$$

where η_{pv} denotes the conversion efficiency of the solar panels, and A_c denotes the total active area of the panels. Beam radiation is translated from the direct-beam radiation $I_B(k)$ (normal to the rays) by

$$I_{BC}(k) = I_B(k)\cos\theta(k), \tag{2}$$

with the incident angle $\theta(k)$ given by

$$\cos\theta(k) = \cos\beta(k)\cos(\phi_S(k) - \phi_C(k))\sin\psi + \sin\beta(k)\cos\psi. \tag{3}$$

In (3), $\beta(k)$ denotes the altitude angle, $\phi_S(k)$ denotes the solar azimuth angle, $\phi_C(k)$ denotes the azimuth angle of the panels, and ψ denotes the tilt angle of the PV panels as shown in Fig. 2.

In the absence of actual measurements of diffused radiation, an estimation of the PV power output $P_{pv,DC}$ due to this component is given by [40]

$$P_{pv,DC}(k) = \eta_{pv} C I_B(k)\left(\frac{1 + \cos\psi}{2}\right)A_c, \tag{4}$$

with the sky diffuse coefficient C is approximated by [40]

$$C = 0.095 + 0.04\sin\left[\frac{360}{365}\left(n - 100\right)\right], \tag{5}$$

where n denotes the number of the day in the year.

In the absence of actual measurements of reflected radiation, an estimation of the PV power output $P_{pv,RC}(k)$ due to this component is given by [40]

$$P_{pv,RC}(k) = \eta_{pv}\rho I_B(k)(\sin\beta(k) + C)\left(\frac{1 - \cos\psi}{2}\right)A_c, \tag{6}$$

where ρ denotes the ground reflectance.

The total power generation of the PV panel at sample time k, denoted by $P_{pv}(k)$, can be obtained by the summation of the components given in (1), (4) and (6):

$$P_{pv}(k) = P_{pv,BC}(k) + P_{pv,DC}(k) + P_{pv,RC}(k). \tag{7}$$

Depending on the operating conditions, the net power supply of the solar panels to the power system, denoted by P_1 in Fig. 1, can vary between zero and the generated power

$$0 \leqslant P_1(k) \leqslant \eta_1 P_{pv}(k), \tag{8}$$

where η_1 denotes the efficiency of the DC-DC converter of the PV system. The excess portion of P_{pv} is dumped in a dissipative load (not shown in the Fig. 1) whenever the set-point sent to the converter of the PV system is such as $P_{pv}(k) > P_1(k)/\eta_1$.

2.2. Battery bank

The battery power P_2 can be decomposed into charging power P_2^+ and discharging power P_2^-. The variation of the battery SoC induced by charging and discharging operations can be approximated by

$$SoC_b\left(k + 1\right) = SoC_b(k) + \eta_2\eta_{b,c}P_2^+\left(k + 1\right)\Delta T - \frac{1}{\eta_2\eta_{b,d}}P_2^-\left(k + 1\right)\Delta T, \tag{9}$$

where $SoC_b(k + 1)$ and $SoC_b(k)$ denote the battery SoC at, respectively, sample times $k + 1$ and k, η_2 denotes the conversion efficiency of the DC-DC converter of the battery, $\eta_{b,c}$ and $\eta_{b,d}$ denote charging efficiency and discharging efficiency, respectively, and ΔT represents the sampling step. Based on (9), the battery SoC at sample time k is expressed as a function of the initial value $SoC_b(0)$ by

$$SoC_b(k) = SoC_b(0) + \eta_2\eta_{b,c}\sum_{\tau=0}^{k} P_2^+(\tau)\Delta T - \frac{1}{\eta_2\eta_{b,d}}\sum_{\tau=0}^{k} P_2^-(\tau)\Delta T, \tag{10}$$

At any sample time k, the battery SoC is subject to

$$\underline{SoC_b} \leqslant SoC_b(k) \leqslant \overline{SoC_b}, \tag{11}$$

where $\underline{SoC_b}$ and $\overline{SoC_b}$ are respectively the lower and upper bounds.

The battery must be operated so that the charging and discharging do not exceed their respective upper bound $\overline{P_{b,ch}}$ and $\overline{P_{b,disch}}$

$$0 \leqslant P_2^+(k) \leqslant \overline{P_{b,ch}}/\eta_2, \tag{12}$$

$$0 \leqslant P_2^-(k) \leqslant \eta_2\overline{P_{b,disch}}. \tag{13}$$

The following constraint prevents simultaneous charging and discharging of batteries

$$P_2^+(k)P_2^-(k) = 0. \tag{14}$$

The resulting battery power P_2 is given by (15)

$$P_2(k) = P_2^-(k) - P_2^+(k). \tag{15}$$

2.3. Supercapacitors

The SC power P_3 can be decomposed into charging power P_3^+ and discharging power P_3^-. Similarly to the battery, the SC SoC at sample time k is expressed of in terms of the initial value $SoC_{sc}(0)$ by

$$SoC_{sc}(k) = SoC_{sc}(0) + \eta_3\eta_{sc,c}\sum_{\tau=0}^{k} P_3^+(\tau)\Delta T - \frac{1}{\eta_3\eta_{sc,d}}\sum_{\tau=0}^{k} P_3^-(\tau)\Delta T, \tag{16}$$

where $SoC_{sc}(k + 1)$ and $SoC_{sc}(k)$ are the SC SoC at, respectively, sample times $k + 1$ and k, η_3 denotes the conversion efficiency of the DC-DC converter of the SC, and $\eta_{sc,c}$ and $\eta_{sc,d}$ are, respectively, charging efficiency and discharging efficiency of the SC.

At any sample time k, the SC SoC is subject to

$$\underline{SoC_{sc}} \leqslant SoC_{sc}(k) \leqslant \overline{SoC_{sc}}, \tag{17}$$

where $\underline{SoC_{sc}}$ and $\overline{SoC_{sc}}$ denote, respectively, the lower and upper bounds.

The charging and discharging powers of the SC cannot exceed their respective upper bounds $\overline{P_{sc,ch}}$ and $\overline{P_{sc,disch}}$, and neither take place simultaneously

$$0 \leqslant P_3^+(k) \leqslant \overline{P_{sc,ch}}/\eta_3, \tag{18}$$

$$0 \leqslant P_3^-(k) \leqslant \eta_3\overline{P_{sc,disch}}, \tag{19}$$

$$P_3^+(k)P_3^-(k) = 0. \tag{20}$$

The resulting SC power P_3 is given by (21)

$$P_3(k) = P_3^-(k) - P_3^+(k). \tag{21}$$

2.4. Utility grid

The power flow P_4 between the HRS and the utility grid is composed of the power absorbed from the grid P_4^- and the power fed into to it P_4^+. The thermal capacity of the power link between the HRS and the grid, denoted by $\overline{P_{tie}}$, should not be exceeded at all times:

$$P_4^+(k) \leqslant \overline{P_{tie}}, \tag{22}$$

$$P_4^-(k) \leqslant \overline{P_{tie}}. \tag{23}$$

Simultaneous power consumption from and supply to the utility grid is also prevented by

$$P_4^+(k)P_4^-(k) = 0. \tag{24}$$

The resulting power exchange between the utility grid and the HRS is given by

$$P_4(k) = P_4^-(k) - P_4^+(k). \tag{25}$$

3. Model predictive controllers

In order to present the proposed retrofit control approach, an MPC strategy is assumed to be implemented on the existing grid-integrated HRS, which consists of PV panels, batteries, AC and DC loads. This MPC model is provided first. The design of the new control system is conducted afterwards.

3.1. MPC of grid-integrated PV-battery HRS

In this paper, the existing HRS is managed by an MPC strategy implemented on the control unit PMU-1 as shown Fig. 1. In this context, the coordination of power flow across the HRS is performed for the purpose of either the maximization or the minimization of a performance index J that can be technical (e.g. energy autonomy), economic (e.g. operation cost), environmental (e.g. carbon footprint), social (e.g. comfort level), or a combination of these, and taking into consideration the various operation constraints (e.g. power balance, power and energy bounds). In this study, we assume that the control strategy aims to maximize both the self-sufficiency and the use of solar energy, which implies the minimization of the energy supplied by the utility grid and the minimization of the energy dumped in the dissipative load of the PV systems. To limit the thermal stress of the battery power, the sampling time ΔT is in the range of minutes.

In view of this and the system modeling detailed earlier, the discrete-time formulation of the MPC strategy for the existing HRS is as follows:

$$\min J(k) = \sum_{i=1}^{N_p} [P_4^-(k+i) + (\eta_1 P_{pv}(k+i) - P_1(k+i))], \tag{26}$$

subject to

$$P_1(k+i) + P_2^-(k+i) - P_2^+(k+i) + \eta_4 P_4^-(k+i) - P_4^+(k+i)/\eta_4 = P_L(k+i), \tag{27a}$$

$$P_2^-(k+i)P_2^+(k+i) = 0, \tag{27b}$$

$$P_4^+(k+i)P_4^-(k+i) = 0, \tag{27c}$$

$$\underline{SoC_b} \leqslant SoC_b(k) + \eta_2\eta_{b,c}\sum_{\tau=k}^{k+i} P_2^+(\tau)\Delta T - \frac{1}{\eta_2\eta_{b,d}}\sum_{\tau=k}^{k+i} P_2^-(\tau)\Delta T \leqslant \overline{SoC_b}, \tag{27d}$$

$$0 \leqslant P_1(k+i) \leqslant \eta_1 P_{pv}(k+i), \tag{27e}$$

$$0 \leqslant P_2^-(k+i) \leqslant \eta_2 \overline{P_{b,disch}}, \tag{27f}$$

$$0 \leqslant P_2^+(k+i) \leqslant \overline{P_{b,ch}}/\eta_2, \tag{27g}$$

$$0 \leqslant P_4^+(k+i) \leqslant \overline{P_{tie}}, \tag{27h}$$

$$0 \leqslant P_4^-(k+i) \leqslant \overline{P_{tie}}, \tag{27i}$$

with $i = 1, \cdots, N_p$. Here, N_p denotes the prediction horizon, η_4 denotes efficiency of the inverter situated at the PCC, and the power demand $P_L(k+i)$ at time sample $k+i$ is given by

$$P_L(k+i) = P_{L,dc}(k+i) + P_{L,dc}(k+i)/\eta_4 \tag{28}$$

Due to the nonlinearity of (27b) and (27c), the optimization problem (26) and (27) is categorized as a nonlinear programming (NLP) to be solved at each sample time k.

3.2. Unified MPC of grid-integrated PV-battery-SC HRS

In this paper, the goals pursued by the addition of SC to the HRS are threefold: (1) to reduce the impact of the short-term fluctuations of solar energy and load demand upon the attainment of the operational objective expressed by the performance index; (2) to deliver a stable power profile at the PCC; (3) to prevent frequent variations of the battery power. At the control level, one approach to operate the upgraded HRS considering these goals and the SC characteristics is to replace the previous MPC strategy presented earlier by a new MPC strategy purposefully designed. In that case, a higher sampling rate of forecast data and a shorter control step ΔT are necessary to allow the SC to play an effective role despite its limited energy capacity.

The proposed MPC strategy for grid-integrated PV-battery-SC HRS is indicated from (29) to (30p). Besides the update of the power balance in (30a) and the addition of constraints that control the power and energy flows of the SC ((30d), (30i), (30m) and (30n)), a few new constraints are also added to the previous MPC. Particularly, (30b) forces the grid power to remain constant over N_g consecutive control intervals. On the other hand, constraints (30f) and (30g) maintain the battery power between the ramp-rates limits $\underline{\Delta P_2}$ and $\overline{\Delta P_2}$.

$$\min J(k) = \sum_{i=1}^{N_p} [P_4^-(k+i) + (\eta_1 P_{pv}(k+i) - P_1(k+i))], \tag{29}$$

subject to

$$P_1(k+i) + P_2^-(k+i) - P_2^+(k+i) + P_3^-(k+i) - P_3^+(k+i) + \eta_4 P_4^-(k+i) - P_4^+(k+i)/\eta_4 = P_L(k+i), \tag{30a}$$

$$P_4^+(k+s) = P_4^+(k+r), \tag{30b}$$

$$P_2^-(k+i)P_2^+(k+i) = 0, \tag{30c}$$

$$P_3^+(k+i)P_3^-(k+i) = 0, \tag{30d}$$

$$P_4^+(k+i)P_4^-(k+i) = 0, \tag{30e}$$

$$\underline{\Delta P_2} \leqslant P_2^-(k+m) - P_2^-(k+n) \leqslant \overline{\Delta P_2}, \tag{30f}$$

$$\underline{\Delta P_2} \leqslant P_2^+(k+m) - P_2^+(k+n) \leqslant \overline{\Delta P_2}, \tag{30g}$$

$$\underline{SoC_b} \leqslant SoC_b(k) + \eta_2\eta_{b,c}\sum_{\tau=k}^{k+i} P_2^+(\tau)\Delta T - \frac{1}{\eta_2\eta_{b,d}}\sum_{\tau=k}^{k+i} P_2^-(\tau)\Delta T \leqslant \overline{SoC_b}, \tag{30h}$$

$$\underline{SoC_{sc}} \leqslant SoC_{sc}(k) + \eta_3\eta_{sc,c}\sum_{\tau=k}^{k+i} P_3^+(\tau)\Delta T - \frac{1}{\eta_3\eta_{sc,d}}\sum_{\tau=k}^{k+i} P_3^-(\tau)\Delta T \leqslant \overline{SoC_{sc}}, \tag{30i}$$

$$0 \leqslant P_1(k+i) \leqslant \eta_1 P_{pv}(k+i), \tag{30j}$$

$$0 \leqslant P_2^-(k+i) \leqslant \overline{P_{b,ch}}/\eta_2, \tag{30k}$$

$$0 \leqslant P_2^+(k+i) \leqslant \eta_2 \overline{P_{b,disch}}, \tag{30l}$$

$$0 \leqslant P_3^-(k+i) \leqslant \eta_3 \overline{P_{sc,disch}}, \tag{30m}$$

$$0 \leqslant P_3^+(k+i) \leqslant \overline{P_{sc,ch}}/\eta_3, \tag{30n}$$

$$0 \leqslant P_4^+(k+i) \leqslant \overline{P_{tie}}, \tag{30o}$$

$$0 \leqslant P_4^-(k+i) \leqslant \overline{P_{tie}}, \tag{30p}$$

with
$i = 1, \cdots, N_p$, $m = n+1, \cdots, n+N_b-1$, $n = 0, N_b, 2N_b, \cdots, N_p - N_b$, $s, = r+1, \cdots, r+N_g-1$
and $r = 0, N_g, 2N_g, \cdots, N_p - N_g$.

When compared with the MPC strategy implemented on the existing HRS in Section 3.1, the above MPC ensures a longer service life for the batteries and a better power profile at the PCC. However, because of the long prediction horizon N_p required for energy management reasons (in

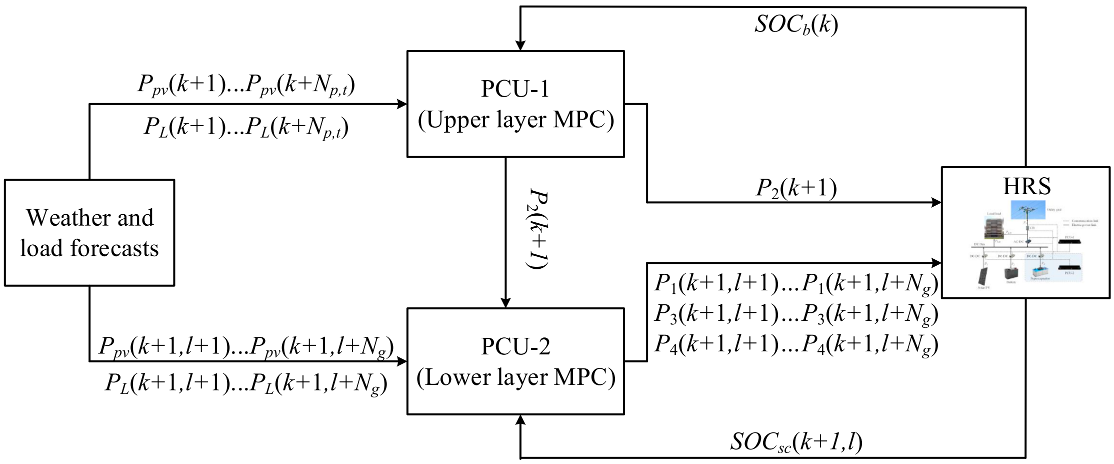


Fig. 3. Control architecture of the hierarchical MPC strategy.

the range of 24 h or more), the relatively short control step ΔT required for power quality reasons (in the range of tens of seconds), and the nature and size of the new MPC strategy, the indicated benefits can be achieved only at the expense of significant increase in computing power and memory resources to allow the implementation of such a controller. A less resource-intensive alternative to this control approach is provided in the next Section.

3.3. Hierarchical MPC of grid-integrated PV-battery-SC system

3.3.1. Architecture and design

Fig. 3 shows the general architecture of the proposed two-layer control framework. Using the MPC of the existing PV-battery HRS as the upper layer, a second MPC strategy implemented in the control unit PMU-2 (see Fig. 1) operates at the bottom layer to achieve a finer control. As mentioned earlier, a single PMU with sufficient resources can be used to execute the two controllers, thus avoiding a partial modification of the original control architecture.

The MPC at the upper layer is used the same way as before the addition of the SC (see Section 3.1), with the only difference being that the optimal control sequences obtained for the PV and the grid are discarded. As shown in Fig. 3, the first element $P_2(k+1)$ in the optimal sequence of the battery is used to control the device and is passed on to PMU-2 for further optimization at the bottom layer. As indicated earlier, the prediction horizon of the existing MPC is in the range of 24 h or more to account for the cycles of solar energy and local demand. Moreover, a sampling step in the range of minutes is applied to extend the service life of the battery. Accordingly, the MPC executed at the upper layer is provided in Section 3.3.2, and is derived from that of the existing HRS in Section 3.1.

At the bottom layer stage, the next upper layer control step $k+1$ is divided into even subintervals in the lower control layer, as shown in Fig. 3. At the time instant $(k+1, l)$, which marks the end of the current subinterval, the SC SoC, the power set-point $P_2(k+1)$ of the battery, and short-term forecast data of solar energy and load demand are used as input to the MPC at the bottom layer. The short-term forecast data supplied to PMU-2 are in the range of minutes, with a sampling rate in the order of seconds. While keeping P_2 as provided by the upper layer helps address the thermal stress of the battery, relevant constraints are included in the bottom layer MPC to deliver a smooth power profile at the PCC, as detailed in Section 3.3.3. Upon completion of the computation process, the first N_g elements of the control sequences of P_1, P_3 and P_4 computed at the bottom layer controller are sent to the power conditioning units for implementation.

3.3.2. Upper layer MPC model

Since this control strategy is directly inherited from the existing HRS, Eqs. (31)–(32i) is taken from Section 3.1. Let the t subscript denote the upper layer variables, the MPC implemented at the upper layer ca be formulated as follows

$$\min J_t(k) = \sum_{i=1}^{N_{p,t}} [P^{-}_{4,t}(k+i) + (\eta_1 P_{pv,t}(k+i) - P_{1,t}(k+i))], \tag{31}$$

s.t.

$$P_{1,t}(k+i) + P^{-}_{2,t}(k+i) - P^{+}_{2,t}(k+i) + \eta_4 P^{-}_{4,t}(k+i) - P^{+}_{4,t}(k+i)/\eta_4 = P_{L,t}(k+i), \tag{32a}$$

$$P^{+}_{2,t}(k+i) P^{-}_{2,t}(k+i) = 0, \tag{32b}$$

$$P^{+}_{4,t}(k+i) P^{-}_{4,t}(k+i) = 0, \tag{32c}$$

$$\underline{SoC_b} \leqslant SoC_b(k) + \eta_2 \eta_{b,c} \sum_{\tau=k}^{k+i} P^{+}_2(\tau)\Delta T_t - \frac{1}{\eta_2 \eta_{b,d}} \sum_{\tau=k}^{k+i} P^{-}_2(\tau)\Delta T_t \leqslant \overline{SoC_b}, \tag{32d}$$

$$0 \leqslant P_{1,t}(k+i) \leqslant \eta_1 P_{pv,t}(k+i), \tag{32e}$$

$$0 \leqslant P^{+}_{2,t}(k+i) \leqslant \overline{P_{b,ch}}/\eta_2, \tag{32f}$$

$$0 \leqslant P^{-}_{2,t}(k+i) \leqslant \eta_2 \overline{P_{b,disch}}, \tag{32g}$$

$$0 \leqslant P^{+}_{4,t}(k+i) \leqslant \overline{P_{tie}}, \tag{32h}$$

$$0 \leqslant P^{-}_{4,t}(k+i) \leqslant \overline{P_{tie}}, \tag{32i}$$

with $i = 1, \cdots, N_{p,t}$.

3.3.3. Lower layer MPC model

Let the b subscript denote the lower layer variables, the MPC implemented at the bottom layer can be formulated as follows

$$\min J_b\left(k, l\right) = \sum_{j=1}^{N_{p,b}} [P^{-}_{4,b}(k+i) + (\eta_1 P_{pv,b}(k+i) - P_{1,b}(k+i))], \tag{33}$$

s.t.

$$P_{1,b}(k, l+j) + P^{-}_{2,t}(k, l+j) - P^{+}_{2,t}(k, l+j) + P^{-}_{3,b}(k, l+j) - P^{+}_{3,b}(k, l+j) + \eta_4 P^{-}_{4,b}(k, l+j) - P^{+}_{4,b}(k, l+j)/\eta_4 = P_{L,b}(k, l+j), \tag{34a}$$

$$P^{+}_{4,b}(k, l+r) = P^{+}_{4,b}(k, l+s), \tag{34b}$$

$$P^+_{3,b}(k,\, l+j)P^-_{3,b}(k,\, l+j) = 0, \tag{34c}$$

$$P^+_{4,t}(k,\, l+j)P^-_{4,t}(k,\, l+j) = 0, \tag{34d}$$

$$\underline{SoC_{sc}} \leqslant SoC_{sc}\left(k,\, l\right) + \eta_3\eta_{sc,c}\sum_{\tau=l}^{l+i} P^+_3\left(k,\, \tau\right)\Delta T_b - \frac{1}{\eta_3\eta_{sc,d}}\sum_{\tau=l}^{l+i} P^-_3\left(k,\, \tau\right)\Delta T_b \leqslant \overline{SoC_{sc}}, \tag{34e}$$

$$0 \leqslant P_{1,b}(k,\, l+j) \leqslant \eta_1 P_{pv,b}(k,\, l+j), \tag{34f}$$

$$0 \leqslant P^+_{3,b}(k,\, l+j) \leqslant \overline{P_{sc,ch}}/\eta_3, \tag{34g}$$

$$0 \leqslant P^-_{3,b}(k,\, l+j) \leqslant \eta_3\overline{P_{sc,disch}}, \tag{34h}$$

$$0 \leqslant P^+_{4,b}(k,\, l+j) \leqslant \overline{P_{tie}}, \tag{34i}$$

$$0 \leqslant P^-_{4,b}(k,\, l+j) \leqslant \overline{P_{tie}}, \tag{34j}$$

with $l = 1,\, N_g+1,\, 2N_g+1,\, \cdots,(M-1)N_g+1$, $M = \Delta T_t/(\Delta T_b N_g)$, j, and $= 0,\, \cdots, N_{h,b}-1$, $r = 0,\, N_g,\, 2N_g,\, \cdots, N_{h,b}-N_g$ $s = r+1,\, r+2,\, \cdots, r+N_g-1$. As before, N_g denotes the number of control steps over which the grid power should be kept constant. The Algorithm 1 summarizes the operation of the hierarchical control strategy.

Table 1
Parameters of the hybrid renewable system.

Parameter	Value	Unit
Solar PV		
η_{pv}	0.124	
ρ	0.20	
Battery bank		
$\eta_{b,c}$	0.85	
$\eta_{b,d}$	1	
$\overline{P_{b,ch}}$ & $\overline{P_{b,disch}}$	250	kW
$\overline{SoC_b}$	0.9	p.u
$\underline{SoC_b}$	0.50	p.u
Supercapacitors		
$\eta_{sc,c}$	0.80	
$\eta_{sc,d}$	1	
$\overline{P_{sc,ch}}$ & $\overline{P_{sc,disch}}$	4 857	kW
$\overline{SoC_{sc}}$	1	p.u
$\underline{SoC_{sc}}$	0.25	p.u
Utility grid		
$\overline{P_{tie}}$	1 200	kW
DC-DC converters		
η_1	0.90	
η_2, η_3	0.85	
AC-DC inverter		
η_4	0.95	

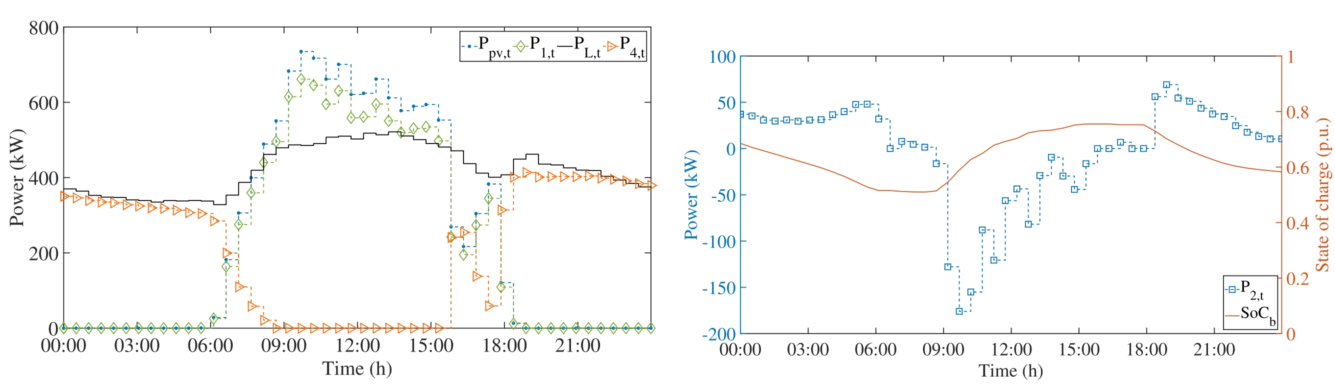
Algorithm 1. Hierarchical Model Predictive Control

1. For time k, minimize (31) subject to (32)
2. **for** $m = 0$ to $M-1$ **do**
3. Set $l = mN_g + 1$
4. Minimize (33) subject to (34)
5. **for** $n = 0$ to $N_g - 1$ **do**
6. Implement
   $P_{1,b}(k,\, l+n)$, $P^+_{2,t}(k)$, $P^+_{2,t}(k)$, $P^+_{3,b}(k,\, l+n)$, $P^-_{3,b}(k,\, l+n)$, $P^+_{4,b}(k,\, l+n)$, $P^-_{4,b}(k,\, l+n)$
7. **End for**
8. **End for**

4. Case study

Considering the load profile shown in Fig. 4(a) and the grid connection of the HRS [40], a 1-MW solar PV is considered to test and compare the control strategies presented in this study. Historical measurements of the direct and diffuse solar radiations are collected from the solar radiometric station at the University of Pretoria (25° 45′S and 28° 13.72′ E) [41,42]. The solar radiation data are applied to the PV arrays oriented South-North and tilted at 36°. The PV power output due to reflected radiation is estimated using Eq. (6), and the contributions due to direct-beam and diffused radiations are determined by Eqs. (1) and (4), respectively. The conversion efficiency of solar panels is provided in Table 1.

The prediction horizon and the control horizon are set at 24 h and 30 min, respectively, at the upper layer, and 30 min and five minutes, respectively, at the bottom layer. The sampling times are respectively set at 30 min and 10 s. The solar radiation during a sampling interval of the upper layer is set equal to the average of actual measurements



(a) Load demand and power profiles of the PV and the grid

(b) Power and SoC of the battery

Fig. 4. Demand profile and control inputs of original MPC strategy (upper layer).

sampled at a one-minute interval. The presumed impact of the forecast errors and the difference of sampling rates between the upper layer (one minute) and the lower layer (ten seconds) is mimicked by adding a white Gaussian noise to the actual solar radiation data, with a signal-to-noise ratio (SNR) of 25 dB. Similarly, the load profile at the bottom layer (sampled at 5-min intervals) is derived from that of the upper layer (sampled at 30-min intervals) by the addition of a white Gaussian noise, with a SNR of 35 dB. When compared to the solar radiation, a higher SNR is applied to the load since, at a facility level, the fluctuations of the load are usually less deep than those of the solar radiation, which is directly affected by the movement of clouds.

The energy storage system consists of a 250-kW/1464-kWh bank of lead-acid batteries [43] and a 13.1-MW/5.74-kWh bank of electrostatic double-layer capacitors (supercapacitors) [44]. The rest of the simulation parameters are provided in Table 1.

5. Simulation and discussion

A PC Core(TM) i5, 3.00 GHz, with 8 GB of RAM running Windows 10, was used to simulate the control strategies presented in this paper. Classified as nonlinear programmings (NLPs), the optimization problems were solved in MATLAB using the “fmincon” function. Because of the limited resources of the PC, the short sampling period (10 s) and the long prediction horizon (24 h), the unified MPC of grid-integrated HRS provided in Section 3.2 takes far too long to simulate and is therefore impractical for real-time applications. Consequently, the discussion is limited to comparing the MPC of PV-battery HRS and the hierarchical MPC of PV-battery-SC HRS. Moreover, because the former is identical in all respects to the upper layer of the latter, we discuss the performances of the upper layer against those of the full hierarchical MPC.

5.1. MPC of grid integrated PV-battery power system (Upper layer)

Fig. 4(a) shows the forecasted load profile $P_{L,t}$, the forecasted PV generation $P_{pv,t}$ and the resulting PV power supply $P_{1,t}$ and grid power $P_{4,t}$. The power and SoC of the battery are shown in Fig. 4(b). It is observed that during night hours and before sunrise, the utility grid covers most of the energy needs in the HRS, while the remaining part is locally supplied by the battery until the minimum SoC is reached. During daylight hours, the power generated by the solar PV is primarily used to supply the load and charge the battery. Only excess power is fed into the utility grid.

A superimposition of the PV generation based on forecast during the upper layer and the bottom layer is shown in Fig. 5(a), and that of the load demand is shown in Fig. 5(b). To maintain the power system balanced, the excess and deficit induced by the fluctuations of the PV generation and the load demand should be handled either by the grid or the battery. The resulting power profiles at the PCC and the battery are presented in Fig. 6(a) and (b), respectively. In Fig. 6(a), $P_{4,ac}$ denotes the actual power profile at the PCC when the fluctuations are handled by the utility grid alone. Similarly, $P_{2,ac}$ denotes the actual battery power when this device handles the fluctuations alone. It is worth reminding that any excess power from the PV is dumped to the dissipative load connected to the PV system. Fig. 6 shows that, unless appropriate actions are taken, the large peaks observed may pose a risk either to the safety of the network system (frequency stability) or the safety and lifetime of the battery (overheating).

5.2. Hierarchical MPC of grid integrated PV-battery-SC system

Fig. 7 presents the power profile at the PCC before and after the implementation of the bottom layer. The new layer proves to be effective in both regulating the power exchanged with the power network and conditioning the PV generation to maximize its supply to the HRS. The increase observed in the amount of renewable energy fed into the grid contributes to creating eco-friendly networks for little extra investment.

The comparison between the battery power and the SC power in Fig. 8(a) shows that the fluctuations are fully handled by the SC so that the battery power is perfectly stable. Fig. 8(b) shows that, like the battery, the SC get discharged after sunset as a result of an attempt to reduce the power consumption from the utility grid.

5.3. Comparison of power and energy performances

A quantitative comparison between the existing control strategy and the proposed one can be carried out using the performance indicators shown in Table 2. Here, the second and third columns correspond to the existing MPC before and after considering the impact of the fluctuations, respectively. The last column corresponds to the hierarchical MPC control strategy. It is worth mentioning that the losses in the DC-DC converter of the PV contribute, with the dumping load, to the difference between the PV energy generation and the PV energy supply. Comparing the first two columns Table 2, it shows that failing to handle the fluctuations of PV power and load demand prevents the HRS to benefit from the increase in PV generation over the implementation stage. With regard to the use of the energy generated by the PV system, a smaller part of it is effectively supplied to the HRS, while an increased quantity (+12.13%) is dumped in the dissipative load. Moreover, an increase in energy consumption from the grid without a significant counterpart is also noticed. Depending on which system component provides support for power balance in this condition, the frequency stability of the network system or the health of the battery may be jeopardized.

Comparing the last two columns in Table 2, a net increase in the use of solar energy is observed, with only 3.08% wasted in the dissipative

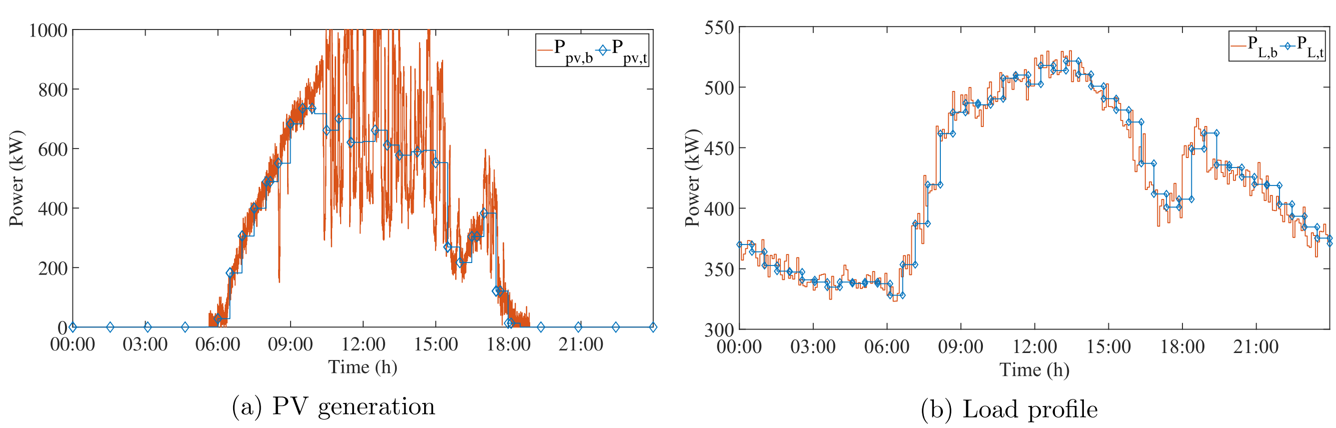
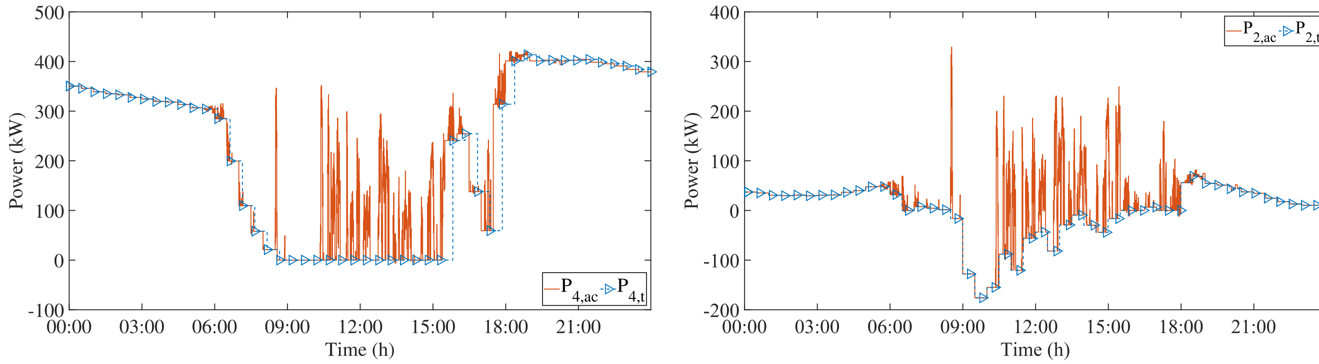


Fig. 5. Upper vs. bottom layer predictions of the PV generation and the load demand.



(a) Power flow at the PCC

(b) Power flow of the battery

Fig. 6. Optimized vs. actual power flows at the PCC and the battery.

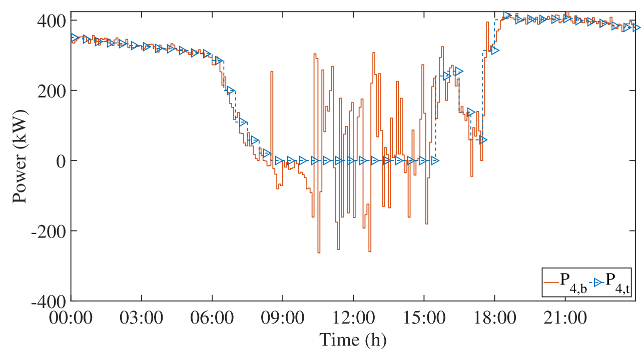


Fig. 7. Power flow at the PCC: MPC PV-battery vs. Hierarchical MPC PV-batt-SC

Table 2
Comparison between existing MPC and hierarchical MPC.

Performance indicator	MPC (optimized)	MPC (actual)	Hierarchical MPC
PV energy generation (kWh)	5796.93	6153.18	6153.18
PV energy supply (kWh)	5217.24	4866.26	5348.00
PV energy dissipated (%)	0.00	12.13	3.41
Total energy import (kWh)	5187.26	5540.97[a]	5594.46
Total energy export (kWh)	0.00	14.86[a]	313.47
5-min intervals with stable P_4 (%)	100	66.67[a]	100
$max(\lvert P_2 + P_3 \rvert)$ (kW)	176.00	329.08[a]	509.03

[a] Fluctuations handled by the utility grid.

load. Despite the operation of the SC leads to a direct increase in energy absorbed from the utility grid (+53.49 kWh), a net diminution (−245.12 kWh) is finally achieved thanks to the additional electricity fed into it. Moreover, the hierarchical control ensures the stability of the power profile at the PCC. The hybrid ESS also provides greater power support to the HRS without further involvement of the battery.

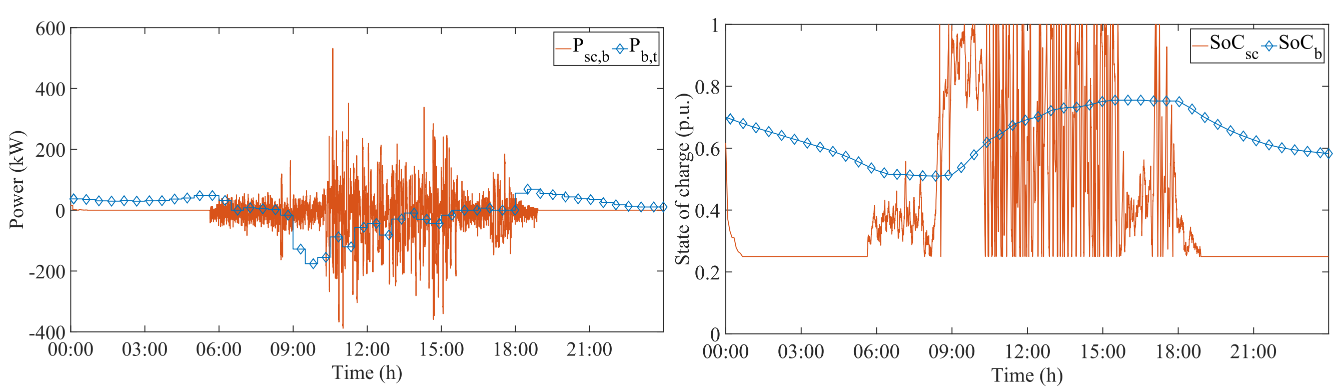
5.4. Setting of the duration of the prediction horizon at the bottom layer

Table 3 presents the average and maximum computation times and the selected energy indicators as functions of the length of the prediction horizon at the bottom layer. In general, long prediction horizons lead to better energy performances than short ones, i.e., increase in energy supply by the solar PV, decrease in consumption from the utility grid, and increase in energy fed into it can be observed. However, this is achieved at the cost of extra computation time, which negatively affects the implementability of the optimal control sequences. Accordingly, a trade-off is necessary between the energy benefit and the extra computation time incurred. In that regard, Table 3 shows that ten-minute prediction horizon leads to energy performances fairly close to those of

Table 3
Duration of the lower layer prediction horizon vs energy performances.

Predic. horiz.	Avg time[a]	Max. time[a]	PV supply[b]	Grid import[b]	Grid export[b]
5 min.	0.09	0.88	5301.84	5668.60	331.42
10 min.	0.42	9.39	5349.00	5594.46	313.47
15 min.	1.09	31.17	5380.64	5591.11	347.50
20 min.	3.45	125.07	5388.33	5591.42	352.72
30 min.	8.68	237.88	5387.80	5590.56	351.83

[a] In seconds.
[b] In kWh.



(a) Power flow comparison

(b) SoC comparison

Fig. 8. Comparison of power and SoC profiles of the ESS components.

longer duration, with an acceptable timing in average. However, five-minute prediction horizon might be preferred, since the computation time under the ten-minute prediction horizon can reach up to 9 s.

6. Conclusion

This paper has presented a hierarchical model predictive strategy designed to facilitate the addition of supercapacitors to a pre-existing grid-integrated hybrid renewable system equipped with batteries, initially controlled by a typical model predictive controller. By means of a second control layer, this control strategy uses the supercapacitor to deliver a stable power profile at the point of common coupling. Moreover, the variable components of the power requested from the hybrid energy storage system are fully handled by the supercapacitor, so that the battery power remains stable. Simulations carried out on a practical case study have shown the validity and effectiveness of the proposed control strategy. Opportunities in terms of adherence to power quality regulations, improved conditioning of the power generated by the intermittent renewable sources, and lifetime extension of the battery have been also established.

References

[1] Haller M, Ludig S, Bauer N. Decarbonization scenarios for the eu and mena power system: considering spatial distribution and short term dynamics of renewable generation. Energy Policy 2012;47:282–90.
[2] Fripp M. Switch: a planning tool for power systems with large shares of intermittent renewable energy. Environ Sci Technol 2012;46(11):6371–8.
[3] Winskel M, Markusson N, Moran B, Taylor G. Decarbonising the UK energy system: accelerated development of low carbon energy supply technologies, < http://www.ukerc.ac.uk/publications/decarbonising-the-uk-energy-system-accelerated-development-of-low-carbon-energy-supply-technologies.html > [accessed: Aug. 01, 2018].
[4] Hart EK, Stoutenburg ED, Jacobson MZ. The potential of intermittent renewables to meet electric power demand: current methods and emerging analytical techniques. Proc IEEE 2012;100(2):322–34.
[5] Bird L, Milligan M, Lew D. Integrating variable renewable energy: challenges and solutions Tech. rep. Golden, CO: National Renewable Energy Laboratory (NREL); 2013.
[6] Tazvinga H, Xia X, Zhang J. Minimum cost solution of photovoltaic-diesel-battery hybrid power systems for remote consumers. Sol Energy 2013;96:292–9.
[7] Wamalwa F, Sichilalu S, Xia X. Optimal control of conventional hydropower plant retrofitted with a cascaded pumpback system powered by an on-site hydrokinetic system. Energy Convers Manage 2017;132:438–51.
[8] Bocklisch T. Hybrid energy storage systems for renewable energy applications. Energy Proc 2015;73:103–11.
[9] Soloveichik GL. Battery technologies for large-scale stationary energy storage. Ann Rev Chem Biomol Eng 2011;2:503–27.
[10] Wu Z, Tazvinga H, Xia X. Demand side management of photovoltaic-battery hybrid system. Appl Energy 2015;148:294–304.
[11] Zhu B, Tazvinga H, Xia X. Switched model predictive control for energy dispatching of a photovoltaic-diesel-battery hybrid power system. IEEE Trans Control Syst Technol 2015;23(3):1229–36.
[12] Riffonneau Y, Bacha S, Barruel F, Ploix S, et al. Optimal power flow management for grid connected pv systems with batteries. IEEE Trans Sustain Energy 2011;2(3):309–20.
[13] Hu J, Xu Y, Cheng KW, Guerrero JM. A model predictive control strategy of pv-battery microgrid under variable power generations and load conditions. Appl Energy 2018;221:195–203.
[14] Choi M-E, Kim S-W, Seo S-W. Energy management optimization in a battery/supercapacitor hybrid energy storage system. IEEE Trans Smart Grid 2012;3(1):463–72.
[15] Linden D. Handbok of batteries. McGraw-Hill; 1994.
[16] Pay S, Baghzouz Y. Effectiveness of battery-supercapacitor combination in electric vehicles. Power Tech Conference Proceedings, 2003 IEEE Bologna, Vol. 3. 2003. pp. 6.
[17] National Energy Regulator of South Africa, Grid connection code for renewable power plants (RPPs) connected to the electricity transmission system (TS) or the distribution system (DS) in South Africa (July 2016) < http://www.nersa.org.za/Admin/Document/Editor/file/Electricity/TechnicalStandards/Newable%20Energy/SAGC%20Requirements%20for%20Renewable%20Power%20Plants%20Rev%202%209.pdf > [accessed: Aug. 01, 2018].
[18] Wang G, Ciobotaru M, Agelidis VG. Power smoothing of large solar pv plant using hybrid energy storage. IEEE Trans Sustain Energy 2014;5(3):834–42.
[19] Feng L, Zhang J, Li G, Zhang B. Cost reduction of a hybrid energy storage system considering correlation between wind and pv power. Protection Control Modern Power Syst 2016;1(1):11.
[20] Li J, Xue Y, Tian L, Yuan X. Research on optimal configuration strategy of energy storage capacity in grid-connected microgrid. Protection Control Modern Power Syst 2017;2(1):35.
[21] Ding M, Wang B, Chen Z, Chen Z, Luo Y, Zheng G. Stabilizing control strategy of complementary energy storage in renewable energy system. Innovative Smart Grid Technologies-Asia (ISGT Asia), 2012 IEEE. IEEE; 2012. p. 1–5.
[22] Zuo W, Li R, Zhou C, Li Y, Xia J, Liu J. Battery-supercapacitor hybrid devices: recent progress and future prospects. Adv Sci 2017;4(7):1600539.
[23] Dougal RA, Liu S, White RE. Power and life extension of battery-ultracapacitor hybrids. IEEE Trans Compon Packag Technol 2002;25(1):120–31.
[24] Ma T, Yang H, Lu L. Development of hybrid battery–supercapacitor energy storage for remote area renewable energy systems. Appl Energy 2015;153:56–62.
[25] Tummuru NR, Mishra MK, Srinivas S. Dynamic energy management of hybrid energy storage system with high-gain pv converter. IEEE Trans Energy Convers 2015;30(1):150–60.
[26] Hredzak B, Agelidis VG, Jang M. A model predictive control system for a hybrid battery-ultracapacitor power source. IEEE Trans Power Electron 2014;29(3):1469–79.
[27] Xiang C, Wang Y, Hu S, Wang W. A new topology and control strategy for a hybrid battery-ultracapacitor energy storage system. Energies 2014;7(5):2874–96.
[28] Kollimalla SK, Ukil A, Gooi H, Manandhar U, Tummuru NR. Optimization of charge/discharge rates of a battery using a two-stage rate-limit control. IEEE Trans Sustain Energy 2017;8(2):516–29.
[29] Wang G, Ciobotaru M, Agelidis VG. Power management for improved dispatch of utility-scale pv plants. IEEE Trans Power Syst 2016;31(3):2297–306.
[30] Han X, Chen F, Cui X, Li Y, Li X. A power smoothing control strategy and optimized allocation of battery capacity based on hybrid storage energy technology. Energies 2012;5(5):1593–612.
[31] Parwal A, Fregelius M, Temiz I, Göteman M, de Oliveira JG, Boström C, et al. Energy management for a grid-connected wave energy park through a hybrid energy storage system. Appl Energy 2018;231:399–411.
[32] Roncero-Sánchez P, Parreño Torres A, Vázquez J. Control scheme of a concentration photovoltaic plant with a hybrid energy storage system connected to the grid. Energies 2018;11(2):301.
[33] Masaki MS, Zhang L, Xia X. Hierarchical power flow control of a grid-tied photovoltaic plant using a battery-supercapacitor energy storage system. Energy Proc 2018;145:32–7.
[34] Ye C, Miao S, Lei Q, Li Y. Dynamic energy management of hybrid energy storage systems with a hierarchical structure. Energies 2016;9(6):395.
[35] Tummuru NR, Mishra MK, Srinivas S. Dynamic energy management of renewable grid integrated hybrid energy storage system. IEEE Trans Ind Electron 2015;62(12):7728–37.
[36] Kotra S, Mishra MK. A supervisory power management system for a hybrid microgrid with hess. IEEE Trans Industr Electron 2017;64(5):3640–9.
[37] Aktas A, Erhan K, Özdemir S, Özdemir E. Dynamic energy management for photovoltaic power system including hybrid energy storage in smart grid applications. Energy 2018;162:72–82.
[38] Manandhar U, Ukil A, Beng GH, Tummuru NR, Kollimalla SK, Wang B, et al., Energy management and control for grid connected hybrid energy storage system under different operating modes. IEEE Trans Smart Grid. https://doi.org/10.1109/TSG.2017.2773643.
[39] Jing W, Ling DK, Lai CH, Wong WS, Wong MD. Hybrid energy storage retrofit for standalone photovoltaic-battery residential energy system. Innovative Smart Grid Technologies-Asia (ISGT Asia), 2017 IEEE. IEEE; 2012. p. 1–6.
[40] Masters GM. Renewable and efficient electric power systems. John Wiley & Sons; 2013.
[41] Brooks M, du Clou S, van Niekerk J, Gauche P, Leonard C, Mouzouris M, et al. Sauran: a new resource for solar radiometric data in southern africa. J Energy Southern Afr 2015;26:2–10.
[42] Southern African Universities Radiometric Network, University of Pretoria - Hatfield Campus metering station [Online] < http://sauran.net/ShowStation.aspx?station=5 > (Feb. 06, 2017).
[43] Current Automation, Sealed and vented lead-acid batteries [Online] < http://www.rectifier.co.za/Catalogue/hoppecke.pdf > (Accessed: Apr. 01, 2017).
[44] Maxwell ® Technologies, Supercapacitor datasheet [Online], < http://www.maxwell.com/products/ultracapacitors/downloads > (Accessed: Apr. 01, 2017).

energies

MDPI

Article

A Hierarchical Optimisation of a Compressed Natural Gas Station for Energy and Fuelling Efficiency under a Demand Response Program

Charles Kagiri, Lijun Zhang and Xiaohua Xia *

Centre of New Energy Systems, Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa; kagiricharles@gmail.com (C.K.); lijun.lzhang@gmail.com (L.Z.)

* Correspondence: xxia@up.ac.za; Tel.: +27-815-476-363

Received: 17 April 2019; Accepted: 15 May 2019; Published: 6 June 2019

check for updates

Abstract: Compressed natural gas stations serve customers who have chosen compressed natural gas powered vehicles as an alternative to diesel and petrol based ones, for cost or environmental reasons. The interaction between the compressed natural gas station and electricity grid requires an energy management strategy to minimise a significant component of the operating costs of the station where demand response programs exist. Such a strategy when enhanced through integration with a control strategy for optimising gas delivery can raise the appeal of the compressed natural gas, which is associated with reduced criteria air pollutants. A hierarchical operation optimisation approach adopted in this study seeks to achieve energy cost reduction for a compressed natural gas station in a time-of-use electricity tariff environment as well as increase the vehicle fuelling efficiency. This is achieved by optimally controlling the gas dispenser and priority panel valve function under an optimised schedule of compressor operation. The results show that electricity cost savings of up to 60.08% are achieved in the upper layer optimisation while meeting vehicle gas demand over the control horizon. Further, a reduction in filling times by an average of 16.92 s is achieved through a lower layer model predictive control of the pressure-ratio-dependent fuelling process.

Keywords: optimal scheduling; demand response; model predictive control; hierarchical control; compressed natural gas

1. Introduction

1.1. Background

Global efforts to minimise environmental pollution have become a priority of many governments, with the transport industry targeted to replace diesel and petrol fuels with less polluting alternatives such as compressed natural gas (CNG) [1]. The use of CNG correlates with the lowest emissions of particulate matter, non-methane organic gases (NMOG), nitrogen oxides (NO_x), carbon monoxide (CO) and other air toxics, among hydrocarbon fuels [2] as well as lower carbon dioxide emission for the same quantity of energy delivered [3]. The availability of the infrastructure to deliver CNG to vehicular customers is a major success factor in the growth of CNG as an alternative fuel for the transportation sector [4] because of fuelling convenience considerations [5]. There has been steady growth in the number of commercial fuelling stations in both developing [6] and developed [7] countries, which has corresponded to the increase in number of CNG vehicles on roads. For commercial fuelling stations, vehicles needing refuelling arrive randomly and are required to be filled quickly, hence the fast-fill CNG fuelling configuration has been the prominent design of choice [8]. In fast-fill stations, gas from the utility line is compressed into a pressurised cascade storage consisting of gas tanks in three pressure levels, from which arriving vehicles are filled [9]. In this type of operation, the compressor is cycled between the upper and lower limits of the cascade storage capacity [10]. Given that the compressor is the main electrical load in a CNG fast-fill station, the cycling of the compressor and its potential for being scheduled present opportunities for the improvement of operation efficiency.

1.2. Improving the CNG Station Operation Efficiency

The improvement of operation efficiency encompasses both energy cost reduction and ensuring performance levels in product delivery are sustained or improved, under the optimised energy cost operation. Operation optimisation for energy cost efficiency through equipment scheduling is a major area of consideration in demand response research [11]. Given the significant consequences of compressor energy consumption on the operating costs of the CNG station [12], it is necessary to study how proposed interventions for energy cost reduction, interact with other operation efficiency improvements at the gas dispensing level. Vehicle fuelling time has been studied as one of the major factors customers consider when deciding whether or not to transition to alternative fuels [13].

Kountz et al. [14] initiated the evaluation of the fast-fill CNG station with a study which involved the development of a model for the flow of gas from one of the cascade storage tanks into the target vehicle tank. Kountz [15] further developed an approach to the design of dispenser algorithm, to ensure correct quantities of gas are dispensed into the target vehicle tank with compensation for temperature effect [16]. Studies of the effects of other components of the CNG station on gas flow such as the hoses [17] and dispensers [18], have aided in developing a basis for their standardisation. Farzaneh et al. [19] developed a numerical method of analysing thermodynamic characteristics of gas flow in the reservoir filling process. The ratio of target vehicle tank pressure to the pressure of the storage tank and the evolution of this ratio as the vehicle tank gets filled are shown to have an effect on the vehicle filling time and profile [20]. Further, studies to determine the optimal location of CNG stations in a network that also includes petrol and diesel fuelling stations [21] have been carried out. Kuby [5] took a deeper look at evaluating the location problem for stations serving alternative fuel vehicles (AFVs) by reviewing the state of relevant research work, and thereby concluded that drivers of AFVs exhibit deliberate behaviour in choosing where to refuel within sparse refuelling networks, with convenience weighing more significantly than price.

Bang et al. [22] modelled the CNG residential refuelling system, and demonstrated the potential effects of an increase in the number of such systems on the existing electricity grid. The study of these effects is especially important, given the significant size of the compressor motor as an electric load in comparison with regular loads of petrol and diesel fuelling stations [23]. Cycling of the compressor in a fast-fill station to replenish the cascade storage may present an opportunity to minimise the energy cost of the CNG station, if the CNG station is located in an area where demand response programs have been implemented through time differentiated pricing [24]. Demand response programs are implemented with an overall goal of achieving lower fluctuations in electricity demand which has been shown to lead to more efficient operation of the grid [25] and to increase the reliability and stability of the grid network [11]. Electricity consumption patterns are modified by raising the price charge per unit of electricity at times when the system reliability is compromised by high demand [26]. This encourages consumers to shift their flexible loads to times when the rates charged are favourable, achieving for them lower overall energy costs [27].

In [12,28], a strategy to minimise electricity cost for a CNG fast fill station was undertaken for a station operating under a time-of-use (TOU) electricity tariff. The station was modelled as a mass balance system where the storage was modelled as a single reservoir with an outflow from a known demand profile and inflow from an optimally scheduled compressor. Further, in [29], an optimal control to determine the operation of the priority panel valves under a known demand profile for each of the three reservoirs of the cascade storage was carried out. These studies considered only the flow of quantity of gas in mass from the compressor to satisfy mass of gas demand at the dispenser. Without evaluating and optimising the pressure conditions during the flow of gas from the cascade storage to the vehicle tanks, it is impossible to guarantee that the level of fuelling time performance is maintained after energy cost saving operation interventions. Disruption of fuelling time performance threatens convenience and could sour consumer sentiment on use of CNG, even when costs are lowered [5].

In the present work, a novel study for the efficient operation of a CNG fast-fill station is presented. The hierarchical model includes an upper layer, which is an optimisation of compressor scheduling to minimise energy cost, and a lower layer to control the valves of the priority panel and the gas dispenser so as to achieve desirable conditions of pressure for minimum vehicle filling time. On the upper layer, the scheduling of the compressor operation to minimise electricity cost incurred under a TOU tariff is realised while minimising compressor switching frequency and meeting the gas demand in the control horizon. The compressor operation schedule obtained is implemented on the lower layer as an input for the optimal control of vehicle fuelling to achieve minimum filling time using a model predictive control strategy (MPC). MPC strategies are popular in modern control applications with demonstrated benefits of their closed loop robustness and stability [30,31], and the ability handle constraints in complex applications [32]. This study presents the first attempt to combine the optimal minimisation of CNG station energy cost through compressor scheduling, with the optimal control of the vehicle filling pressure conditions from the cascade storage to achieve minimum filling times. This proposed approach will safeguard the gains from energy cost savings, by ensuring a simultaneous improvement in gas transfer performance which is of great importance to fuelling convenience. The current work and case study highlight how adoption of alternative fuels intersects with electricity demand response programs, and how the operation optimisation for demand response must be enhanced with performance optimisation to secure the resulting complementary benefits.

This article is laid out as follows: In Section 2, the models for the upper and lower layers are presented. The case study considered for the proposed strategy is described in Section 3. Results and discussions for the outcomes of the study are reported in Section 4. Section 5 concludes the study.

2. System Modelling and Formulation

2.1. The Energy Cost Minimisation Layer

Figure 1 shows the configuration of the CNG fast-fill station. Under normal operation, the compressor receives natural gas from the utility's distribution pipeline at low to medium pressure, approximately 4–15 bar [20], and compresses it into a three level cascade storage system. The gas being compressed passes through a priority panel valve system that alternates the flow of CNG between the three levels of the cascade storage usually called the high pressure, medium pressure and low pressure levels according to their minimum allowed operating pressures [18]. The series of valves v_{hp}, v_{mp} and v_{lp} in the priority panel represent the inlet valves to the high pressure, medium pressure and low pressure tanks of the cascade storage respectively. When the upper pressure limit for all the cascade storage level is achieved, the compressor switch u is turned off so that no more gas flows into the cascade storage. Vehicles arriving at the dispenser have their tanks filled through the dispenser valves v_{ohp}, v_{omp} and v_{olp} for the high pressure, medium pressure and low pressure cascade storage tanks, respectively. The gas flow is alternated so that a lower limit of flow rate determines the tank from which the vehicle is filled, starting with the lowest pressure tank to the medium pressure tank as the vehicle tank fills up and topping off with the high pressure tank [33]. As CNG leaves the cascade storage, the pressure in storage drops and when the minimum pressure limits are reached, the compressor switch u comes on to replenish the storage [34] and the cycle is repeated. The gas demand at the dispenser, m_o, determines the cycling of the compressor and thus the total cost of electricity incurred in a TOU electricity tariff [24]. The energy cost minimisation layer is formulated as a mass flow problem, as we proposed in our previous study [29]. This means that the scheduling of the compressor operation is optimised around mass inflow to the cascade storage from the municipal supply line and mass outflow as determined by mass of gas demand at the dispenser over the control horizon.

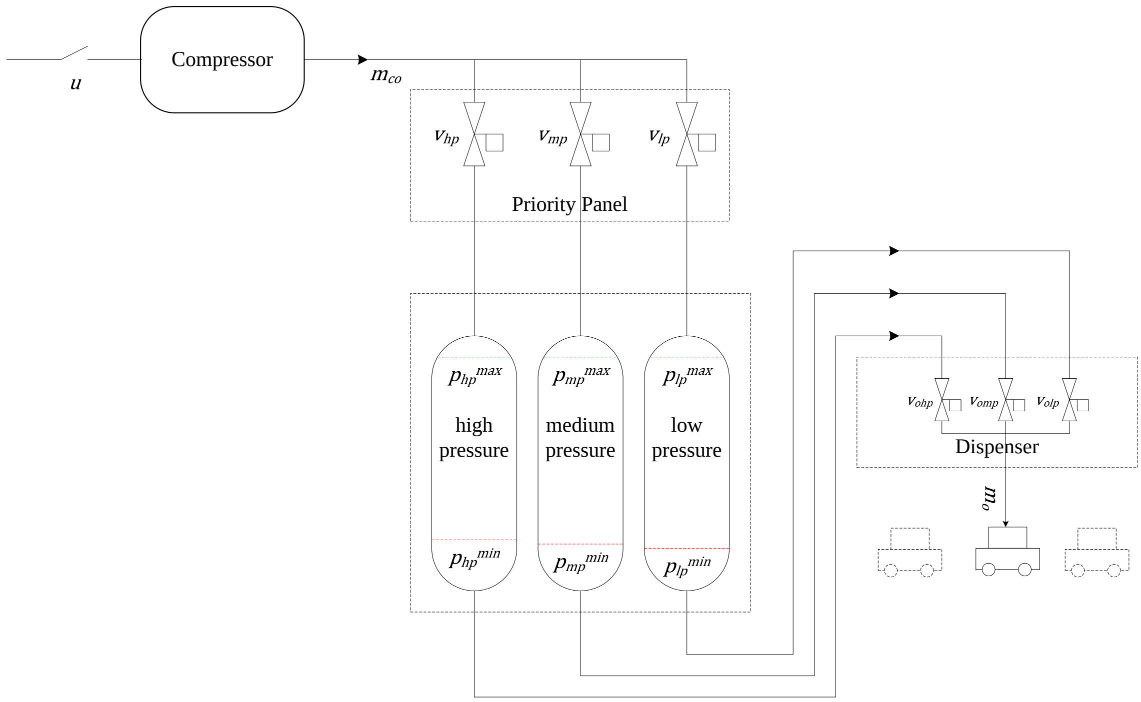


Figure 1. Layout of the fast-fill CNG station.

2.1.1. Objective Function

The objective of this layer is to minimise the cost of electricity incurred by the compressor operation over the control horizon so that the following objective function is used:

$$J = \sum_{t=1}^{N} P_{co} P_e(t) t_s u(t), \tag{1}$$

where t is the counter for the sampling instants, N is the total number of sampling instants over the control horizon, P_{co} is the compressor power rating, $P_e(t)$ is a vector of the price of electricity per kWh in a TOU tariff, t_s is the sampling period and $u(t)$ is the status of the compressor switch which is the control variable such that

$$u(t) \in \{0, 1\} \quad \text{for} \quad 1 \leq t \leq N. \tag{2}$$

It is important to modify the objective function so that the switching frequency of the compressor is minimised as well. This is because the frequency of on/off instances positively correlates to increased wear and tear of moving components of the compressor [35,36]. For the optimised minimum cost of electricity incurred over the control horizon, this study seeks to achieve the lowest number of switching instances for the compressor. One of the methods considered is the one used in [12], where the approach involves the reduction of the ramp rate between successive instances of the switch so that the element of the objective function dealing with minimising compressor frequency is

$$J_q = \sum_{t=1}^{N-1} \Big(u(t+1) - u(t) \Big)^2. \tag{3}$$

Elsewhere, in [29], the approach is based on the introduction of an auxiliary variable $s(t)$ [37,38] that assumes a value of 1 when a switch-on occurs and tries to minimise the summation of the auxiliary variable over the control horizon such that

$$J_{pr} = \sum_{t=1}^{N} s(t), \tag{4}$$

and

$$u(1) - s(1) \leq 0, \tag{5}$$

$$u(t) - u(t-i) - s(t) \leq 0. \tag{6}$$

Although both methods have been found to be effective, in the present study, we propose to introduce a new method where the operation is optimised to prefer the occurrence of on-instances in succession of each other by minimising the summation of the negative product of successive instances of the solution to the control variable u, so that the objective function becomes

$$J_U = \varrho \sum_{t=1}^{N} P_{co} P_e t_s u(t) + (1-\varrho) \sum_{t=1}^{N-1} -\Big(u(t)u(t+1)\Big), \tag{7}$$

where ϱ is a weighting factor. The weighting factor can be set to reduce the number of switching instances so that the minimum number possible is attained for the same energy cost incurred such as was the case in [29]. The method proposed in the current study for minimising the frequency of compressor switching involves a single mathematical operation and no additional constraints which reduces the computational complexity of the problem when compared with Equations (3) and (4).

2.1.2. Constraints

The constraints for this upper layer minimising energy cost are based on the total mass storage capacity of the cascade storage as well as the terminal conditions so that

$$m_{min} \leq m(t) \leq m_{max}, \tag{8}$$

where m_{max} is the maximum mass limit of gas for the cascade storage corresponding to the maximum pressure limits, m_{min} is the minimum mass limit of gas for the cascade storage at the minimum pressure limits and the mass of gas in the cascade storage $m(t)$ is

$$m(t) = m(0) + t_s \sum_{i=0}^{t-1} \dot{m}_{co} u(i) - \sum_{i=0}^{t-1} m_o(i), \tag{9}$$

where $m_o(i)$ is the gas flowing out of the cascade storage into a vehicle in a sampling instant and $\dot{m}_{co}$ is the mass flow rate of the compressor which is obtained as [39]

$$\dot{m}_{co} = \rho_{std} \times Q_{std} = (\frac{Mw_g}{Mw_a}) \times \rho_{a,std} \times Q_{std}, \tag{10}$$

where ρ_{std} is the density of CNG under standard conditions (0 °C temperature and 10^5 pascals pressure) [40], Mw_g is the molecular weight of the CNG, Mw_a is the molecular weight of air, $\rho_{a,std}$ is the air density under standard conditions and Q_{std} is the capacity of the compressor under standard conditions.

The mass limits of gas for the cascade storage capacity constraints m_{min} and m_{max} are derived from working pressure limits of the cascade storage and the physical properties of the gas

$$PV = znRT, \tag{11}$$

where P is the value of the pressure rating, V is the total volume of the cascade storage, z is the compressibility factor, R is the ideal gas constant and n the quantity of gas in moles which is correlated with the mass as

$$n = \frac{m}{M}, \tag{12}$$

where M is the molar mass. The working mass limits for the cascade storage therefore become

$$m_{max} = \frac{MVP_{max}}{zRT} \quad m_{min} = \frac{MVP_{min}}{zRT}. \tag{13}$$

2.1.3. Algorithm

To solve the problem using OPTI toolbox SCIP solver interfaced in Matlab, the upper layer energy cost minimisation layer problem is formulated in the form

$$\text{minimise}_x \quad f(x), \tag{14}$$
$$\text{subject to} \quad Ax \leq b, \tag{15}$$
$$l_b \leq x \leq u_b, \tag{16}$$
$$x \in \{0,1\}. \tag{17}$$

The objective function in Equation (1) is expressed as

$$f(x) = \Big(\varrho P_{co} P_e t_s \times \big(u(1) + u(2) + \cdots + u(N)\big)\Big) - \Big((1-\varrho) \times \big(u(1) \times u(2) + u(2) \times u(3) + \cdots + u(N-1) \times u(N)\big)\Big). \tag{18}$$

From the constraint in Equation (8) and the dynamic equation of mass in Equation (9), these linear inequalities can be expressed as

$$Ax \leq b_1, \tag{19}$$
$$-Ax \leq b_2, \tag{20}$$

where

$$A = \begin{bmatrix} -t_s\dot{m}_{co} & 0 & \cdots & 0 \\ -t_s\dot{m}_{co} & -t_s\dot{m}_{co} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -t_s\dot{m}_{co} & -t_s\dot{m}_{co} & \cdots & -t_s\dot{m}_{co} \end{bmatrix}_{N\times N}, \tag{21}$$

$$b_1 = \begin{bmatrix} m(0) - m_{min} - m_o(1) \\ m(0) - m_{min} - \big(m_o(1) + m_o(2)\big) \\ \vdots \\ m(0) - m_{min} - \big(m_o(1) + m_o(2) + \cdots + m_o(N)\big) \end{bmatrix}_{N\times 1}, \tag{22}$$

$$b_2 = \begin{bmatrix} m_{max} - m(0) + m_o(1) \\ m_{max} - m(0) + \big(m_o(1) + m_o(2)\big) \\ \vdots \\ m_{max} - m(0) + \big(m_o(1) + m_o(2) + \cdots + m_o(N)\big) \end{bmatrix}_{N\times 1}. \tag{23}$$

The linear inequality constraints in the form of $Ax \leq b$ become

$$A = \begin{bmatrix} A \\ -A \end{bmatrix}_{2N\times N}, b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}_{2N\times 1}. \tag{24}$$

The control vector for the problem, x, can be written in the standard form

$$x = [u(1), u(2) \cdots u(N)]^T_{N\times 1}. \tag{25}$$

2.2. Gas Flow Optimisation Layer

A model predictive control (MPC) strategy is implemented on the lower layer with a prediction horizon N_p and the sampling time t_{ss}. The status of the compressor switch u is obtained from the solution of optimisation of the upper layer. Whenever switch u is on, gas flows into the three tank storage via valves v_{hp}, v_{mp} and v_{lp} of the priority panel. The gas flows in from the compressor at a constant mass flow rate $\dot{m}_{co}$. Each of the three tanks has maximum and minimum pressures, p^{max}_{hp}, p^{max}_{mp}, p^{max}_{lp} and p^{min}_{hp}, and p^{min}_{mp} and p^{min}_{lp}, respectively. Gas flows into the vehicle from the storage tanks via the dispenser valves v_{ohp}, v_{omp} and v_{olp}. The initial pressure for each vehicle tank p_{veh} is a known quantity from the demand data while the initial pressure for the high pressure tank p_{hp}, medium pressure tank p_{mp} and low pressure tank p_{lp} are measured from the final conditions after the previous control action.

2.2.1. Objective Function

The objective of this layer is to minimise the difference between the vehicle tank pressure $p_{veh}(k+j)$ and the target pressure $p_T(k+j)$ which corresponds to the quantity of gas ordered by the customer for the vehicle at step j based on the current sampling instant k. This ensures continuous flow of gas from the cascade storage tanks to the vehicle tank. Additionally, we minimise the summation of dispenser valve action instances, which ensures minimisation of filling time. This is because lowering the total number of instances required for the dispenser valves to be on in order to fill the vehicle tank, corresponds to a shorter filling time of the vehicle tank. Therefore, the controller prefers the cascade filling profile with the least number of total dispenser valve open instances. The objective function based on the current sampling instant k is therefore to minimise

$$J_L(k) = (\varsigma)\sum_{j=0}^{N_p-1}\Bigg(p_T(k+j) - p_{veh}(k+j)\Bigg) + (1-\varsigma)\sum_{j=0}^{N_p-1}\Bigg(v_{ohp}(k+j) + v_{omp}(k+j) + v_{olp}(k+j)\Bigg), \tag{26}$$

where ς is a weighting factor and $v_{ohp}(k+j)$, $v_{omp}(k+j)$ and $v_{olp}(k+j)$ are the dispenser statuses for the high pressure, medium pressure and low pressure cascade storage tanks, respectively. Gas flow from the cascade storage tanks to the vehicle tank, $\dot{m}_{veh}(k+j)$ ensures that the vehicle pressure approaches the target pressure value and is the sum of flow rates from the three tanks, so that based on the current sampling instant k

$$\dot{m}_{veh}(k+j) = \dot{m}_{hp}(k+j)v_{ohp}(k+j) + \dot{m}_{mp}(k+j)v_{omp}(k+j) + \dot{m}_{lp}(k+j)v_{olp}(k+j). \tag{27}$$

The equations for the instantaneous flow rates $\dot{m}_{hp}(k+j)$, $\dot{m}_{mp}(k+j)$ and $\dot{m}_{lp}(k+j)$ between the high, medium and low pressure tanks of the cascade storage, respectively, and the vehicle tank, are based on the ideal gas model for an adiabatic system [41] and are governed by the pressure ratios between the storage tanks and the vehicle tank. i.e.,

$$\begin{aligned}\dot{m}_{hp}(k+j) =& C_d\rho_{hp}(k+j)A_{orifice}\left(\frac{p_{veh}(k+j)}{p_{hp}(k+j)}\right)^{\frac{1}{\gamma}}\left\{(\frac{2\gamma}{\gamma-1})(\frac{p_{hp}(k+j)}{\rho_{hp}(k+j)})\left(1-(\frac{p_{veh}(k+j)}{p_{hp}(k+j)})^{\frac{\gamma-1}{\gamma}}\right)\right\}^{\frac{1}{2}}\\ &\text{for}\quad \frac{p_{veh}(k+j)}{p_{hp}(k+j)} \leq \left(\frac{2}{\gamma+1}\right)^{\frac{\gamma}{\gamma-1}},\end{aligned} \tag{28}$$

and

$$\dot{m}_{hp}(k+j) = C_d\sqrt{\gamma p_{hp}(k+j)\rho_{hp}(k+j)A_{orifice}}\left(\frac{2}{\gamma+1}\right)^{\frac{\gamma+1}{2(\gamma-1)}} \text{for}\quad \frac{p_{veh}(k+j)}{p_{hp}(k+j)} \geq \left(\frac{2}{\gamma+1}\right)^{\frac{\gamma}{\gamma-1}}, \tag{29}$$

and similarly for the $\dot{m}_{mp}(k+j)$

$$\begin{aligned}\dot{m}_{mp}(k+j) =& C_d \rho_{mp}(k+j) A_{orifice}\left(\frac{p_{veh}(k+j)}{p_{mp}(k+j)}\right)^{\frac{1}{\gamma}}\left\{(\frac{2\gamma}{\gamma-1})(\frac{p_{mp}(k+j)}{\rho_{mp}(k+j)})\left(1-(\frac{p_{veh}(k+j)}{p_{mp}(k+j)})^{\frac{\gamma-1}{\gamma}}\right)\right\}^{\frac{1}{2}} \\ &\text{for} \quad \frac{p_{veh}(k+j)}{p_{mp}(k+j)} \leq \left(\frac{2}{\gamma+1}\right)^{\frac{\gamma}{\gamma-1}},\end{aligned} \tag{30}$$

and

$$\dot{m}_{mp}(k+j) = C_d\sqrt{\gamma p_{mp}(k+j)\rho_{mp}(k+j)A_{orifice}}\left(\frac{2}{\gamma+1}\right)^{\frac{\gamma+1}{2(\gamma-1)}} \text{for} \quad \frac{p_{veh}(k+j)}{p_{mp}(k+j)} \geq \left(\frac{2}{\gamma+1}\right)^{\frac{\gamma}{\gamma-1}}, \tag{31}$$

and for $\dot{m}_{lp}(k+j)$

$$\begin{aligned}\dot{m}_{lp}(k+j) =& C_d \rho_{lp}(k+j) A_{orifice}\left(\frac{p_{veh}(k+j)}{p_{lp}(k+j)}\right)^{\frac{1}{\gamma}}\left\{(\frac{2\gamma}{\gamma-1})(\frac{p_{lp}(k+j)}{\rho_{lp}(k+j)})\left(1-(\frac{p_{veh}(k+j)}{p_{lp}(k+j)})^{\frac{\gamma-1}{\gamma}}\right)\right\}^{\frac{1}{2}} \\ &\text{for} \quad \frac{p_{veh}(k+j)}{p_{lp}(k+j)} \leq \left(\frac{2}{\gamma+1}\right)^{\frac{\gamma}{\gamma-1}},\end{aligned} \tag{32}$$

and

$$\dot{m}_{lp}(k+j) = C_d\sqrt{\gamma p_{lp}(k+j)\rho_{lp}(k+j)A_{orifice}}\left(\frac{2}{\gamma+1}\right)^{\frac{\gamma+1}{2(\gamma-1)}} \text{for} \quad \frac{p_{veh}(k+j)}{p_{lp}(k+j)} \geq \left(\frac{2}{\gamma+1}\right)^{\frac{\gamma}{\gamma-1}}, \tag{33}$$

where γ is the ratio of specific heats

$$\gamma = \frac{c_p}{c_v}, \tag{34}$$

and c_p is the specific heat capacity of the gas at constant pressure while c_v is specific heat capacity of the gas at constant volume. C_d is the coefficient of discharge of the dispenser valve orifice, $A_{orifice}$ is the area of the dispenser valve orifice and ρ_{hp}, ρ_{mp} and ρ_{lp} are the densities of the gas in the high pressure ,medium pressure and low pressure reservoirs, respectively.

2.2.2. Constraints

The valves at the dispenser and the priority panel, as the control variables, are subject to operational constraints. The valves of the priority panel open one at a time when the compressor is filling the cascade storage reservoirs which gives the constraint in Equation (35). The valves of the dispenser also open one at a time during the filling of the vehicle from the cascade storage as represented by the constraint in Equation (36).

$$v_{hp}(k+j) + v_{mp}(k+j) + v_{lp}(k+j) - u(k+j) = 0, \tag{35}$$

$$v_{ohp}(k+j) + v_{omp}(k+j) + v_{olp}(k+j) \leq 1, \tag{36}$$

$$v_{ohp}(k+j), v_{omp}(k+j), v_{olp}(k+j), v_{hp}(k+j), v_{mp}(k+j), v_{lp}(k+j), u(k+j) \in \{0,1\}.$$

Further, the vehicle tank pressure p_{veh} and the pressure in the three cascade reservoirs p_{hp}, p_{mp} and p_{lp}, as the states of the gas flow optimisation layer, are also subject to operational constraints. The limits of pressure for the vehicle tank and each of the reservoirs of the cascade storage are such that

$$p_{hp}^{min} \leq p_{hp}(k+j) \leq p_{hp}^{max}, \tag{37}$$

$$p_{mp}^{min} \leq p_{mp}(k+j) \leq p_{mp}^{max}, \tag{38}$$

$$p_{lp}^{min} \leq p_{lp}(k+j) \leq p_{lp}^{max}, \tag{39}$$

$$p_{veh}(k+N_p+1-j) \geq p_T(k), \tag{40}$$

Equations (37)–(39) ensure that the maximum and minimum working pressures of the cascade storage tanks are not exceeded, while Equation (40) ensures that, at the end of the control horizon, the vehicle tank is filled to the target pressure corresponding to the requested quantity of gas by the customer.

Based on the described flow of gas for the proposed approach, the general differential equations for pressure change in the vehicle and cascade storage reservoirs are

$$\frac{d}{dt}p_{veh}(t) = \dot{m}_{veh}(t)K_1, \tag{41}$$

$$\frac{d}{dt}p_{hp}(t) = -\dot{m}_{hp}(t)K_{hp}v_{ohp}(t) + \dot{m}_{co}v_{hp}(t), \tag{42}$$

$$\frac{d}{dt}p_{mp}(t) = -\dot{m}_{mp}(t)K_{mp}v_{omp}(t) + \dot{m}_{co}v_{mp}(t), \tag{43}$$

$$\frac{d}{dt}p_{lp}(t) = -\dot{m}_{lp}(t)K_{lp}v_{olp}(t) + \dot{m}_{co}v_{lp}(t), \tag{44}$$

where the constants K_1, K_{hp}, K_{mp} and K_{lp} are

$$K_1 = T\left(\frac{c_p}{c_v}\frac{R}{V_{veh}}\right), \quad K_{hp} = T\left(\frac{c_p}{c_v}\frac{R}{V_{hp}}\right), \quad K_{mp} = T\left(\frac{c_p}{c_v}\frac{R}{V_{mp}}\right) \quad \text{and} \quad K_{lp} = T\left(\frac{c_p}{c_v}\frac{R}{V_{lp}}\right), \tag{45}$$

where V_{veh}, V_{hp}, V_{mp} and V_{lp} are the volumes of the vehicle tank, high pressure reservoir, medium pressure reservoir and low pressure reservoir, respectively. This yields the following discrete equations of pressure, for the current sampling instant k

$$p_{veh}(k+j) = p_{veh}(k) + t_{ss}K_1 \sum_{\tau=k}^{k+j} \dot{m}_{veh}(\tau), \tag{46}$$

$$p_{hp}(k+j) = p_{hp}(k) - t_{ss}K_{hp} \sum_{\tau=k}^{k+j} \dot{m}_{veh}(\tau)v_{ohp}(\tau) + t_{ss}\dot{m}_{co} \sum_{\tau=k}^{k+j} v_{hp}(\tau), \tag{47}$$

$$p_{mp}(k+j) = p_{mp}(k) - t_{ss}K_{mp} \sum_{\tau=k}^{k+j} \dot{m}_{veh}(\tau)v_{omp}(\tau) + t_{ss}\dot{m}_{co} \sum_{\tau=k}^{k+j} v_{mp}(\tau), \tag{48}$$

$$p_{lp}(k+j) = p_{lp}(k) - t_{ss}K_{lp} \sum_{\tau=k}^{k+j} \dot{m}_{veh}(\tau)v_{olp}(\tau) + t_{ss}\dot{m}_{co} \sum_{\tau=k}^{k+j} v_{lp}(\tau). \tag{49}$$

2.2.3. Algorithm

To solve the gas flow optimisation layer problem using the Mixed Integer Distributed Ant Colony Optimisation (MIDACO) solver, the components of the problem have to be formulated as

$$\begin{aligned} &\text{minimise} && f(x) && \text{(objective function)} && (50)\\ &\text{subject to} && g(x)=0 && \text{(equality constraints)} && (51)\\ & && h(x)\geq 0 && \text{(inequality constraints)} && (52)\end{aligned}$$

The control vector consists of the conditions of the three priority panel valves and the three dispenser valves, and for each sampling instant k, x can be written in the standard form

$$\begin{aligned} x = [&v_{hp}(k+1), v_{hp}(k+2), \cdots v_{hp}(k+N_p), v_{mp}(k+1), v_{mp}(k+2), \cdots v_{mp}(k+N_p), v_{lp}(k+1),\\ &v_{lp}(k+2), \cdots v_{lp}(k+N_p), v_{ohp}(k+1), v_{ohp}(k+2), \cdots v_{ohp}(k+N_p), v_{omp}(k+1),\\ &v_{omp}(k+2), \cdots v_{omp}(k+N_p), v_{olp}(k+1), v_{olp}(k+2), \cdots v_{olp}(k+N_p),]^T_{6N_p\times 1}. \end{aligned} \tag{53}$$

The objective function in Equation (26),

$$\begin{aligned} f = [&\varsigma \times \Big(P_T - P_{veh}(k+1) + P_T - P_{veh}(k+2) \cdots P_T - P_{veh}(k+N_p)\Big) + (1-\varsigma) \times \Big(v_{ohp}(k+1) +\\ &v_{omp}(k+1) + v_{olp}(k+1) + v_{ohp}(k+2) + v_{omp}(k+2) + v_{olp}(k+2) \cdots\\ &v_{ohp}(k+N_p) + v_{omp}(k+N_p) + v_{olp}(k+N_p)\Big)]. \end{aligned} \tag{54}$$

The equality constraint in Equation (35) yields the $g(x)=0$ set for the algorithm so that

$$g(x) = \begin{bmatrix} v_{hp}(k+1)+v_{mp}(k+1)+v_{lp}(k+1)-u(k+1)\\ v_{hp}(k+2)+v_{mp}(k+2)+v_{lp}(k+2)-u(k+2)\\ \vdots\\ v_{hp}(k+N_p)+v_{mp}(k+N_p)+v_{lp}(k+N_p)-u(k+N_p) \end{bmatrix}_{N_p\times 1}, \tag{55}$$

while the inequality in Equation (36) yields the first set of $h(x)\geq 0$ such that

$$h_1(x) = \begin{bmatrix} 1-\Big(v_{ohp}(k+1)+v_{omp}(k+1)+v_{olp}(k+1)\Big)\\ 1-\Big(v_{ohp}(k+2)+v_{omp}(k+2)+v_{olp}(k+2)\Big)\\ \vdots\\ 1-\Big(v_{ohp}(k+N_p)+v_{omp}(k+N_p)+v_{olp}(k+N_p)\Big) \end{bmatrix}_{N_p\times 1}. \tag{56}$$

The next set of inequality constraints is derived from Equations (37)–(39) such that

$$h_2(x) = \begin{bmatrix} p_{hp}^{max} - p_{hp}(k+1) \\ \vdots \\ p_{hp}^{max} - p_{hp}(k+N_p) \\ p_{mp}^{max} - p_{mp}(k+1) \\ \vdots \\ p_{mp}^{max} - p_{mp}(k+N_p) \\ p_{lp}^{max} - p_{lp}(k+1) \\ \vdots \\ p_{lp}^{max} - p_{lp}(k+N_p) \\ p_{hp}(k+1) - p_{hp}^{min} \\ \vdots \\ p_{hp}(k+N_p) - p_{hp}^{min} \\ p_{mp}(k+1) - p_{mp}^{min} \\ \vdots \\ p_{mp}(k+N_p) - p_{mp}^{min} \\ p_{lp}(k+1) - p_{lp}^{min} \\ \vdots \\ p_{lp}(k+N_p) - p_{lp}^{min} \end{bmatrix}_{6N_p \times 1}. \tag{57}$$

The final element of the inequality constraints is derived from Equation (40), yielding

$$h_3(x) = \begin{bmatrix} p_{veh}(k+N+1-j) - p_T(k) \end{bmatrix}_{1\times 1}. \tag{58}$$

The combined set of inequality constraints therefore becomes

$$h(x) = \begin{bmatrix} h_1(x) \\ h_2(x) \\ h_3(x) \end{bmatrix}_{(7N_p+1)\times 1}. \tag{59}$$

At the current sampling instant k, an open loop optimisation problem is solved by the controller for the prediction horizon N_p. Only the first elements of the control variables v_{hp}, v_{mp}, v_{lp}, v_{ohp}, v_{omp} and v_{olp} are implemented on the CNG filling station plant. The vehicle pressure p_{veh} and the pressure in the cascade storage tanks p_{hp}, p_{mp} and p_{lp}, which are the system states, are measured and the values fed back to the MPC controller, forming the initial states for the following sampling instant $k+1$. The input variables are then updated and the cycle repeated until all control actions for the intended period are implemented.

The MPC controller workflow is such that:

1. For the current sampling instant k, the controller minimises the objective function in Equation (26) and finds an optimum solution for the control variables v_{hp}, v_{mp}, v_{lp}, v_{ohp}, v_{omp} and v_{olp}, subject to the constraints set out in Section 2.2.2.
2. From the solution, only the first elements of the solution $v_{hp}(k|k)$, $v_{mp}(k|k)$, $v_{lp}(k|k)$, $v_{ohp}(k|k)$, $v_{omp}(k|k)$ and $v_{olp}(k|k)$ are implemented.
3. The states $p_{veh}(k+1)$, $p_{hp}(k+1)$, $p_{mp}(k+1)$ and $p_{lp}(k+1)$ are measured to be fed back.
4. The value of k is set to $k = k+1$ and system states, inputs and outputs are updated.
5. Steps 1–4 are repeated until k reaches a value predetermined to mark the end of the control period.

3. Case Study

The case study involves a roadside vehicle fuelling station based in Johannesburg South Africa, that is currently in operation, located in an industrial zone, as shown on the map in Figure 2.

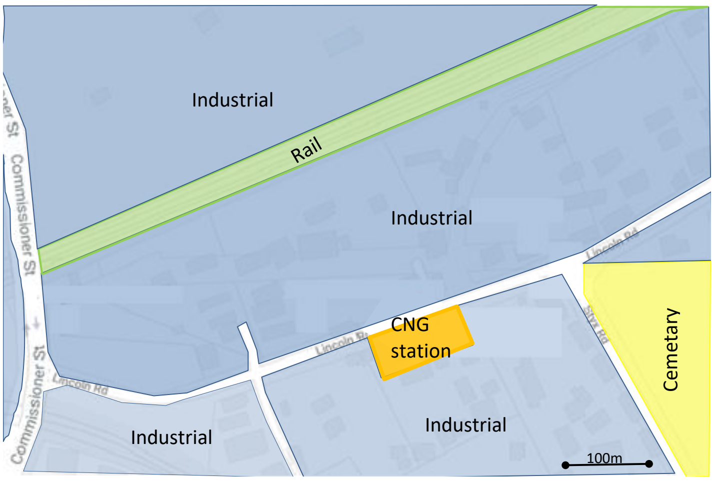


Figure 2. Station location and land use map.

The average hourly demand profile for a 24-h period, which is the upper layer control horizon N, is shown in Figure 3. The station serves vehicles mainly in the public transportation sector and fleets of courier and security firms. Both individual and fleet customers arrive one by one on their individual need basis, and there is currently no scheduled fleet refuelling at this station. Vehicles serviced by the fuelling station are hybrid fuelled, with combined CNG and diesel/petrol powered engines. The vehicles are run on CNG and resort to diesel and petrol power when the CNG in their tanks runs out. The station itself obtains gas from a municipal line, which is compressed by a 132 kW motor powered compressor, into three levels of the cascade storage, which are 2000 L each. The three level tanks have a maximum operating pressure of 250 bar and are in the baseline operated at minimum pressures of 75 bar, 150 bar and 210 bar for the low pressure, medium pressure and high pressure reservoirs, respectively. The compressor pumps gas into the storage at a rate of 900 m^3/h. Although the station has two installed compressors and three dispensers, the station only operates one compressor and fills vehicles from one dispenser, since the current number of customers visiting the station is modest and no congestion or queuing problem has arisen. The station compressor operates between the limits of the quantity of gas in storage with the compressor being switched on at the lower limit to fill the cascade storage, and once the compressor is on, stays on to fill the cascade storage to the upper limit. The compressed natural gas station purchases electricity from South Africa's national utility firm Eskom based on a time-of-use tariff known as the Miniflex tariff (http://eskom.co.za/tariffs) which is priced in South African Rands as

$$p_e(t) = \begin{cases} p_{offpeak} = 0.5157\text{R/kWh} & \text{if } t \in [0,6] \cup [22,24] \\ p_{standard} = 0.9446\text{R/kWh} & \text{if } t \in [9,17] \cup [19,22] \\ p_{peak} = 3.1047\text{R/kWh} & \text{if } t \in [6,9] \cup [17,19] \end{cases} \tag{60}$$

The tariff is divided into peak, offpeak and standard times during the day, reflecting the times during the day when electricity demand is high, low and intermediate, respectively.

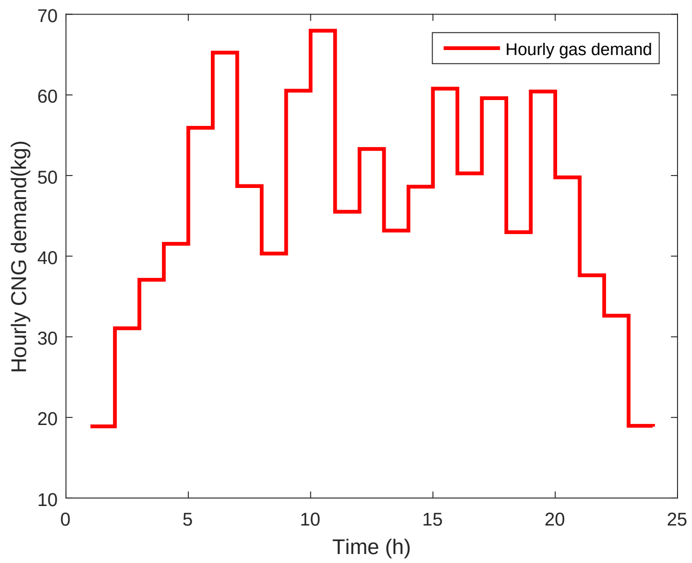


Figure 3. Average hourly gas demand profile for the Johannesburg CNG fuelling station.

A sampling time of 15 min is used for the upper layer of this study with a control horizon of 24 h. For the lower layer MPC problem, a sampling time of 20 s for a receding prediction horizon N_p of five minutes is used. During existing baseline operation, vehicle tank filling starts at the low pressure reservoir and the transfer to medium and high pressure reservoir occurs when the flow rate between the reservoir and the vehicle tank falls to a set point. This study seeks to allow the flexibility of the transfer of vehicle filling between the reservoirs through optimised control of the dispenser valves in the lower layer. The priority panel valves and the dispenser valves are modelled as binary valves with orifice diameters of 5 mm each. There are two sizes of vehicle cylinders for the vehicles fuelled in the 24 h control horizon at 80 and 100 L respectively. The initial vehicle tank pressure is assumed to be 1 bar since the vehicles are hybrid CNG and petrol/diesel powered and typically refill CNG tanks on empty.

The solution of the upper layer compressor schedule obtained from optimisation for the average 24-h demand is applied to the MPC receding horizon control of vehicle filling, for a day in which 143 vehicles fuel at the CNG station with gas demand as shown in Figure 4. Table 1 shows additional parameters and constants

Table 1. Additional parameters and constants.

Parameter	Value
$\rho_{a,std}$	1.225 kg/m^3
C_d	0.61
γ	1.304
Mw_a	0.028966 g
Mw_g	0.0164 g
R	0.083145 LbarK^{-1}mol^{-1}
T	294.15 K

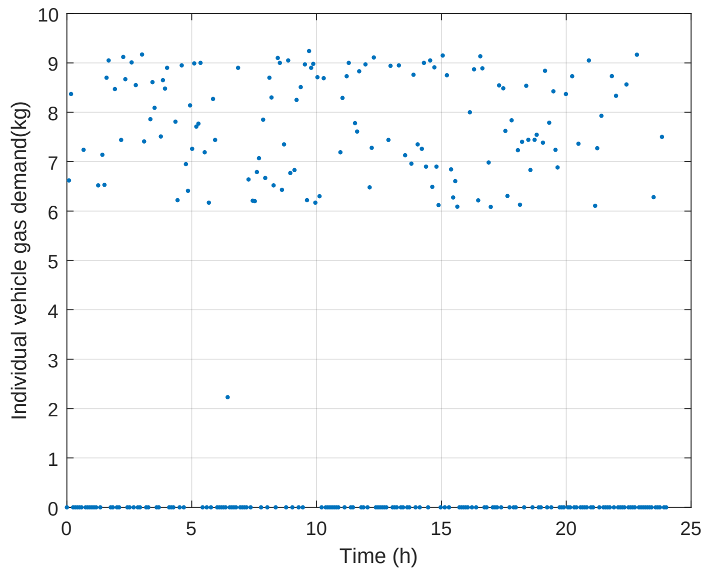
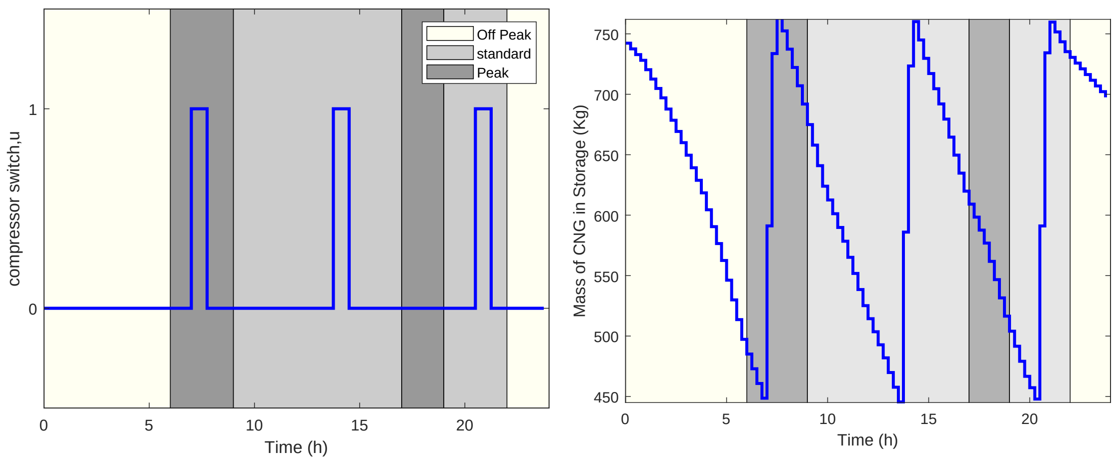


Figure 4. Individual vehicle gas demand over 24 h.

Figure 5 demonstrates the functioning of the compressor under the baseline cycling operation. The compressor cycles between the minimum and maximum limits of the storage to maintain the level of gas within designed operation limits. Indeed, in Figure 5a, it is clear that the compressor operates during the peak electricity pricing period in the morning, as well as during the standard electricity pricing period during the rest of the day. This means that, under baseline operation, the station does not take advantage of the low electricity prices of the offpeak electricity pricing periods. The total cost of electricity incurred as a result of this baseline operation profile is R440.23. The main component of this cost is the cost of electricity consumed as a result of the compressor-on status during the significantly expensive peak electricity pricing period and an optimal avoidance of this occurrence should lower the cost considerably.



(**a**) Compressor switch action

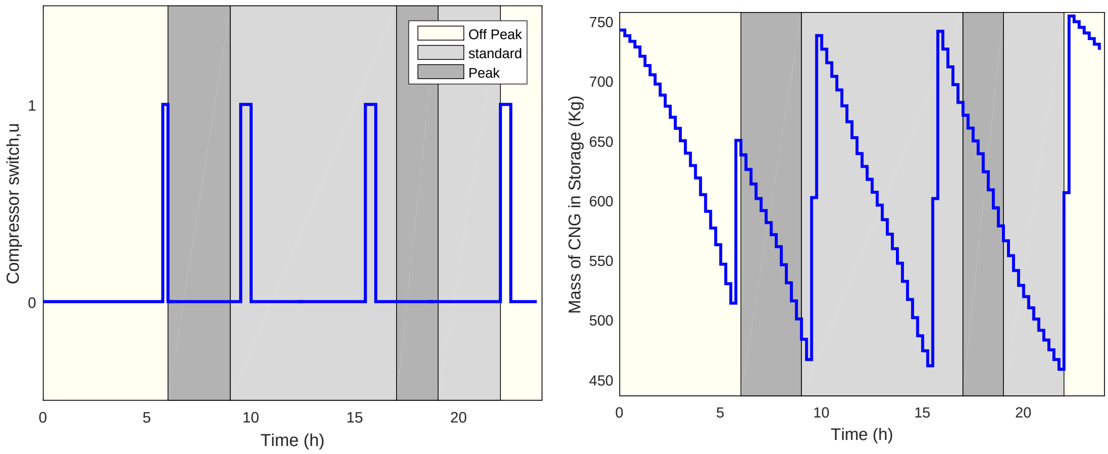
(**b**) Total mass of gas in storage

Figure 5. Baseline Operation profile.

4. Results

4.1. Energy Cost Minimisation Layer Results

The optimised results for the operation schedule of the compressor are shown in Figure 6. The proposed approach is successful in preventing compressor-on instances during both peak electricity demand periods for the average 24-h gas demand of the control horizon. Preceding the first peak demand period between 06:00 and 09:00, the compressor switches on to replenish the gas in storage so as to sustain demand during the peak electricity pricing period during which the compressor stays off. There are two instances when the compressor is turned on during the standard electricity pricing period to meet the gas demand as well as to refill the cascade storage before the second peak electricity demand period of the 24 h. Prior to the second peak, the cascade storage is refilled. The level of gas is thereafter enough to sustain the demand until the end of the second standard electricity pricing time at 22:00.



(**a**) Compressor switch action (**b**) Total mass of gas in storage

Figure 6. Optimised compressor operation result for the average 24-h horizon.

The delay in switching on the compressor after the second peak demand period is a desirable outcome of the optimisation strategy. This is in response to the pricing of the standard electricity demand period that is higher than the pricing during the offpeak electricity demand period, which causes the controller to prefer delaying compressor-on status beyond the peak electricity demand period [42]. The delay is an important quality of the realised schedule as it reduces the contribution of the CNG station to the grid comeback load associated with the surge in electricity demand immediately following the end of the peak pricing period [43,44]. A total of four switching instances occur over the control horizon, which is comparable with observed results for alternative strategies observed in [12,29]. This implies the superiority of the current method as it matches the performance of previously proposed techniques with the added benefit of fewer mathematical operations, thereby achieving the goal of reducing computational complexity of the problem. The computing time achieved for the upper layer using the proposed approach was 15.15 s, compared to 20.22 s for the method using Equation (3) and 35.58 s for the method in Equation (4). The strategy reduces the cost of electricity over the 24-h horizon from the baseline R440.23 to R175.74, which represents a 60.08% reduction in energy cost. This is a significant reduction in cost of operation through a strategy that involves only a change in the operation schedule, without additional investment in new hardware.

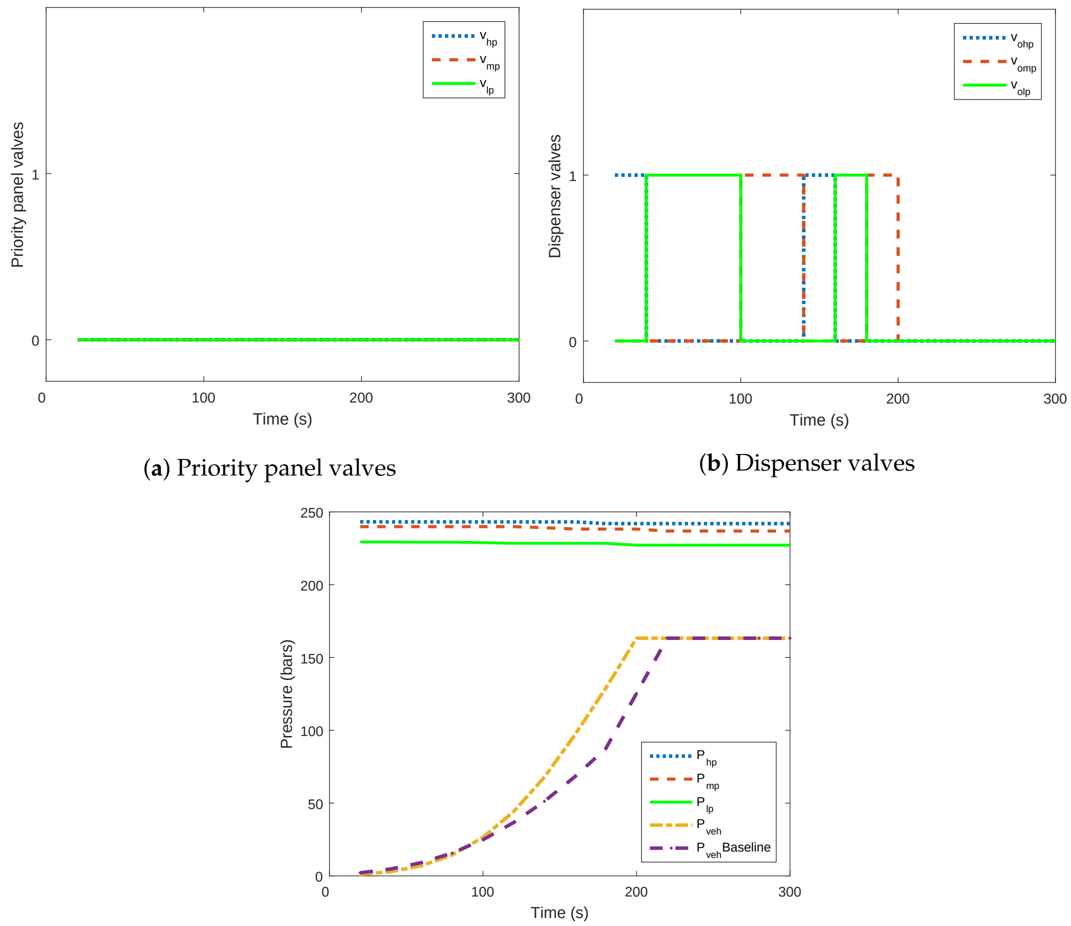
4.2. Gas Flow Optimisation Layer Results

The compressor scheduling results from the energy cost minimisation layer are passed onto the lower layer, for the implementation of the MPC strategy in the vehicle tank filling for each of 143 vehicles fuelled over 24 h. These upper layer results determine the status of the compressor switch for a particular sampling instant in the lower layer. In the lower layer problem, four scenarios emerge, with each having a different priority panel and dispenser status combination. To ascertain the validity of the proposed approach, the system operation must remain valid and consistent with the system constraints under the four scenarios. These four scenarios are, vehicle tank filling with compressor off, vehicle tank filling with compressor on, compressor on without a vehicle filling at the dispenser, and compressor off without a vehicle filling at the dispenser.

4.2.1. Vehicle Filling with the Compressor Off

The results of optimised MPC filling process when a vehicle is at the dispenser and the compressor is off is shown in Figure 7. The results show the filling profile of the fourth vehicle of the 143 filled over the 24-h control horizon of the upper layer. The priority panel remains inactive since no gas flows into the cascade storage from the compressor given the off-status of the compressor switch, which is scheduled from the optimisation of the upper layer. All levels of cascade storage are utilised in the implemented control actions of the filling process to attain the pressure corresponding to the target level of gas to be filled in the vehicle.

The dispenser valves from the three cascade storage tanks switch between each other to fill the vehicle's tank, as shown in Figure 7b, producing the optimal pressure profile of the filling shown in Figure 7c. In Figure 7b, the MPC controller shuffles the operation of the dispenser valves between the three levels of the cascade storage in the filling process which is dependent on the pressure ratio between the vehicle tank and the reservoirs. This is different from the baseline operation where filling is sequentially scheduled and switching occurs at the set point of the dropping flow rate. A comparison of the pressure increase in the vehicle tank under the optimal control strategy and the baseline can be seen in Figure 7c. A filling time of 200 s is achieved, which is shorter than the 220 s achieved under the baseline operation.



(**a**) Priority panel valves (**b**) Dispenser valves

(**c**) Pressure

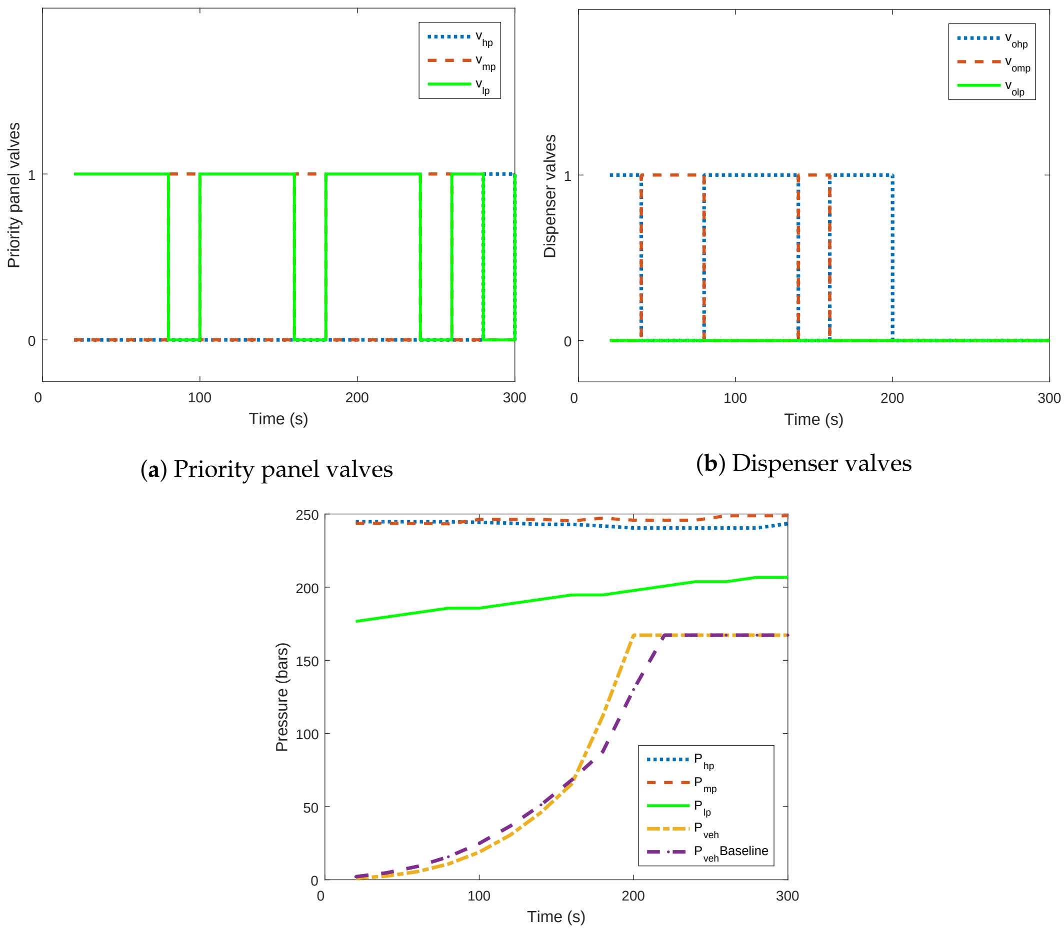
Figure 7. Vehicle filling without the compressor off (fourth vehicle of 143).

4.2.2. Vehicle Filling with the Compressor On

When the vehicle is filled as the compressor is filling the cascade storage, the profile of operation is shown in Figure 8. The results show the profile of the filling process for 38th vehicle of the 143 filled over the 24-h control horizon. A smooth filling profile is obtained with a filling time of 200 s, which is shorter by 20 s from the baseline filling profile.

Similar to the filling of the vehicle while the compressor is off, the dispenser valves switches between levels of cascade storage to produce the optimal filling profile of the vehicle tank so that the targeted quantity of gas is transferred. The priority panel valves alternate the filling of the gas from the compressor into the cascade storage between the three levels.

For both cases in Sections 4.2.1 and 4.2.2, the proposed control strategy produces an efficient accelerated filling of the vehicle tank by switching optimally between the dispenser valves to achieve the minimum possible number of total dispenser valve-on instances to reach the targeted transfer of gas to the vehicle, which corresponds to the shortest possible filling time under the given constraints. This filling profile represents the optimal increase in pressure in the vehicle tank as achieved through the MPC strategy of the lower layer.



(**a**) Priority panel valves (**b**) Dispenser valves

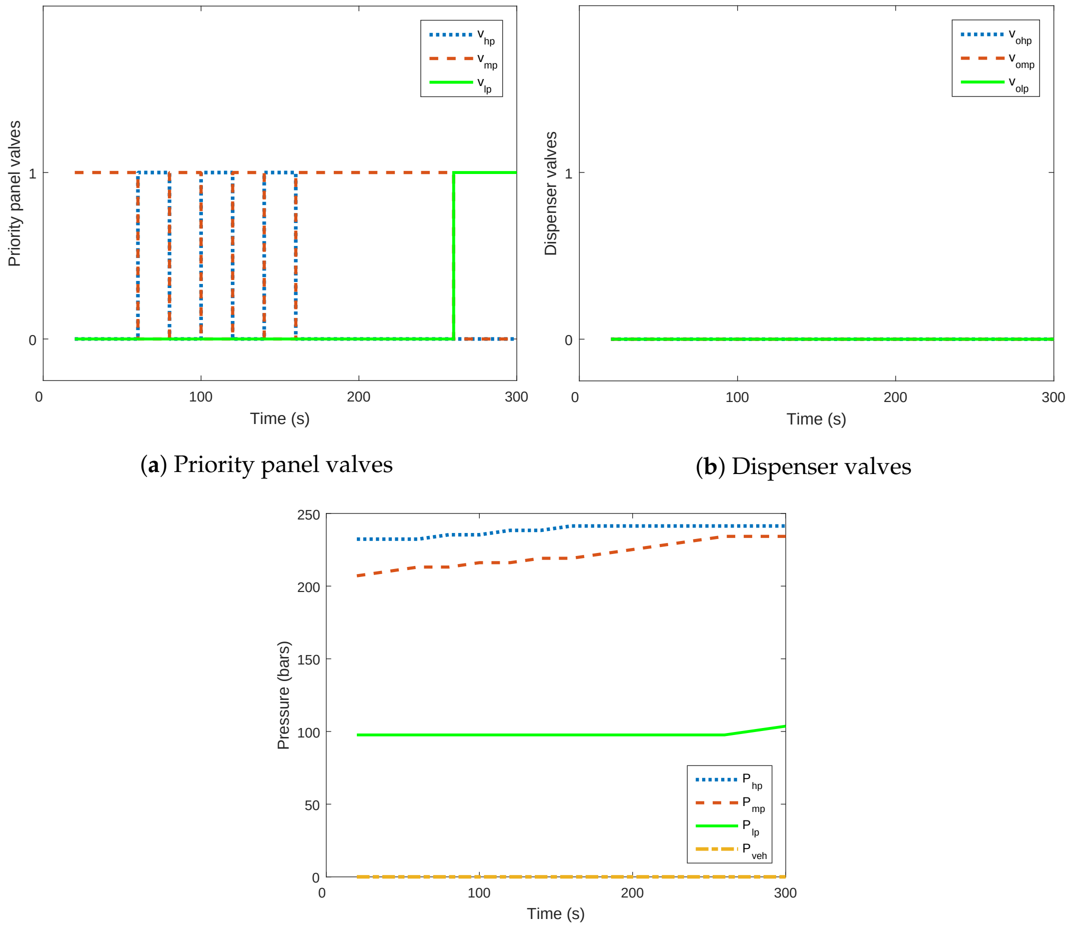
(**c**) Pressure

Figure 8. Vehicle filling with the compressor on (38th vehicle of 143).

Under this novel optimised MPC filling approach, the improvement in filling time is achieved with a median of 20 s reduction in filling time for an average saving of 16.92 s for the 143 vehicles. This outcome confirms that, by altering the operation of the dispenser valves through optimisation, better CNG fuelling performance can be achieved, which would further justify the optimisation of the CNG station operations, beyond the minimising of electricity costs.

4.2.3. Cascade Reservoir Filling without Vehicle Fuelling

As dictated by the energy cost optimisation layer schedule, during the interval when there is no vehicle fuelling at the dispenser but the compressor is on, the profile for the filling of the cascade storage from the compressor is shown in Figure 9.



(**a**) Priority panel valves (**b**) Dispenser valves

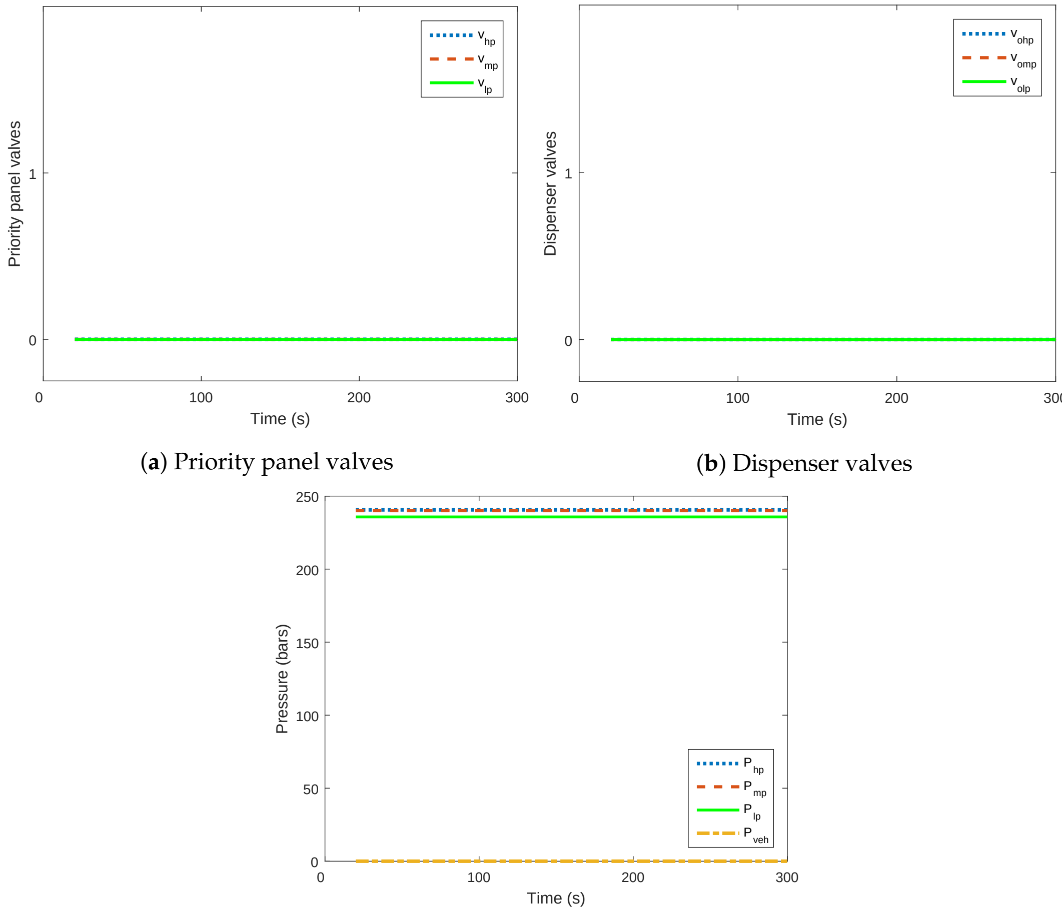
(**c**) Pressure

Figure 9. Compressor operation without vehicle being filled.

A preference to keep the high pressure reservoir of the cascade storage at high pressure is observed under the MPC strategy for the lower layer. The pressure level is flexibly controlled to fulfil the optimal control goals of vehicle filling and meet the conditions of the operation constraints. By successfully keeping the dispenser valves in the off position, the strategy demonstrates feasibility under this condition and achieves the expected performance profile for the given operational constraints.

4.2.4. Control Action during Idle Time

It is necessary to report on the system performance during idle time when the compressor is off and there is no vehicle fuelling at the dispenser. The state of the control variables and the pressure in the cascade storage for the lower layer is shown in Figure 10. The results show that the dispenser valves and priority panel valves remain in the off position over the entire control period demonstrating that the MPC strategy for the lower layer remains feasible when there is no inflow or outflow of gas from the cascade storage of the filling station. This indicates that the strategy is sufficiently constrained for all operation scenarios of the CNG filling station.



(**a**) Priority panel valves (**b**) Dispenser valves

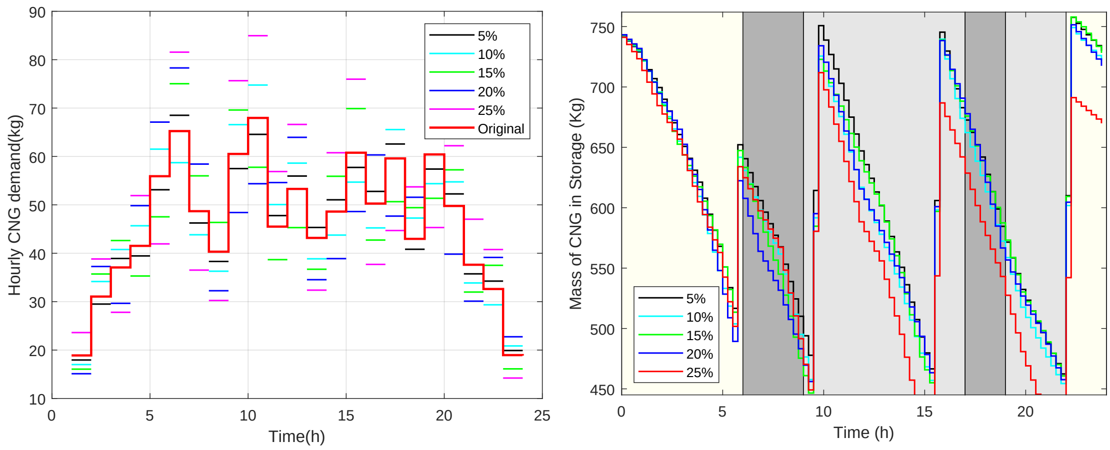
(**c**) Pressure

Figure 10. Profile with compressor off and no vehicle at the dispenser.

The proposed two-layer optimisation for the reduction of energy cost and improvement filling efficiency shows a feasibility demonstrated in the results with a significant reduction in energy cost and vehicle filling times. The achieved reduction in the cost of electricity for the CNG fast-fill station by 60.08% can be passed on to consumers on cost per unit because of the lowered cost of gas delivery. Further, shorter filling times are achieved for vehicles fuelling at the station with a median reduction in filling time of 20 s and an average reduction by 16.92 s for all the vehicles through the MPC strategy for the lower layer. Reduced fuelling time is a welcome improvement in fuelling convenience that could serve to maintain existing CNG vehicle customers and lower the level of concern for new CNG vehicle users. The benefits from the optimisation of the two layers can be viewed as complementary to one another, given the cascaded improvement in financial and technical performance of the CNG station that has been realised.

4.3. Sensitivity Analysis

An analysis was carried out to scrutinise the validity of the model's output under disturbances, which originate from scenarios that alter the input parameters. A random change in the gas demand, which is an important parameter for compressor scheduling, is the most probable source of disturbance for the proposed optimal operation approach and could affect the feasibility of the compressor schedule solution. Consequently, an inspection of the possible effects of random disturbances, implemented as random percentage increase or decrease in hourly gas demand, on the quantity of gas in storage was done. The disturbance demand profiles are shown in Figure 11a.



(**a**) Percentage disturbances to demand (**b**) Mass of gas in storage with disturbances

Figure 11. System performance under random gas demand disturbances.

It is evident in Figure 11b that the solution of the compressor schedule obtained through the optimal scheduling remains valid through the series of variations in hourly gas demand of up to 20%. This means that the limits of the quantity of gas stored in the cascade storage are not violated if the schedule obtained is implemented, even if some variation in demand occur. However, when effecting the optimal schedule solution to the existing station controller, it is necessary to include safety interrupts to tackle circumstances where large disturbances cause a violation of operating limits, as shown in Figure 11b for the 25% variation in demand. These interrupts would correct the violation by either shutting off the compressor outside of schedule when the maximum quantity of gas limit in the cascade storage is reached, or by turning on the compressor outside of schedule when the minimum limit is reached as a result of unexpected gas demand circumstances. The feedback characteristic of the MPC strategy in the lower layer allows for the controller to adapt to disturbances in the system inputs [32], which means that when disturbances in the quantity of gas condition in the cascade storage occur, the controller updates operation control to attain the lower layer objectives.

Table 2 shows the fuelling time for each of the two vehicles in Sections 4.2.1 and 4.2.2 when gas level disturbances in one of the three cascade storage cylinders are implemented. By feeding back the conditions of the system states to the controller input, the approach allows for new calculation of future control action so as to meet the set optimisation goals. The solutions from the proposed approach to optimal CNG station operation have thus been shown to be valid for the predicted operating conditions as well as under some variations to these conditions.

Table 2. Vehicle filling time.

Gas Level Disturbance	4th Vehicle(s)	38th Vehicle(s)
5%	200	200
10%	200	200
15%	200	200
20%	200	200

5. Conclusions

Participation of CNG delivery infrastructure in demand response programs not only saves cost for the station operators and CNG users, but it is also a participation in contributing to the wider goals of such programs in increasing grid reliability and efficiency for all electricity users in society. Cities seeking to expand the use of alternative fuels as cleaner means of transportation also need the associated infrastructure to develop responsibly with regard to use of electricity, which is an indispensable resource.

This study provides an expansive perspective of the operation profile of an optimised CNG fast-fill station, which is the major component of gas delivery infrastructure, incorporating both energy savings and pressure conditions management. The proposed approach achieves a huge reduction in the cost of electricity for the CNG fast-fill station, and delivers on shorter filling times for vehicles fuelling at the station. The results demonstrate savings of up to 60.08% in electricity cost for the upper layer as well as average savings of 16.92 s in vehicle fuelling times for the lower layer. Further, the sensitivity analysis shows an ability of the solutions obtained to withstand some disturbances in the inputs, which is important for the station operation reliability.

Implementation of energy cost reduction strategies by energy users should remain sensitive to other performance considerations that may affect the business under consideration. For compressed natural gas vehicle users, vehicle fuelling time cannot be jeopardised as it is one of the main consideration consumers make when deciding on adoption of cleaner gaseous alternative fuels. The study demonstrates that benefits associated with adoption of CNG can be amplified by optimally operating delivery infrastructure with respect to existing demand response programs while simultaneously improving customer convenience. As an introductory study on the implementation of a combined energy cost and filling time optimisation, this study is a timely highlight to an important intersection between different approaches to better use of energy and system performance.

Author Contributions: The research work and manuscript drafting was done by C.K. under the guidance and supervision of L.Z. and X.X., who also provided critical review of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

$A_{orifice}$	Area of dispenser valve orifice (m^2)
C_d	Co-efficient of discharge of dispenser valve orifice
c_p	Specific heat capacity of CNG at constant pressure (J/KgK)
c_v	Specific heat capacity of CNG at constant volume (J/KgK)
J_U	Objective function of the upper layer
J_L	Objective function of the lower layer
m	Mass of gas (kg)
m_{max}	Maximum mass of gas for the cascade storage (kg)
m_{min}	Minimum mass of gas for the cascade storage (kg)
m_o	Gas demand (kg)
$\dot{m}_{hp}^{max}, \dot{m}_{mp}^{max}, \dot{m}_{lp}^{max}$	Instantaneous mass flow rate from high pressure, medium pressure and low pressure reservoirs to vehicle tank (kg/h)
$\dot{m}_{veh}$	Instantaneous total mass flow rate from cascade storage to vehicle tank (kg/h)
$\dot{m}_{co}$	Compressor outlet mass flow rate (kg/h)
M	Molar mass (kg)
Mw_a	Molecular weight of the air (g)
Mw_g	Molecular weight of the gas (g)
N	Upper layer control horizon
N_p	Lower layer model predictive control prediction horizon
n	Gas quantity (moles)

P	Pressure (bars)
p_{co}	Compressor motor power rating (kW)
p_e	Price of electricity under TOU tariff (currency/kW h)
P_{hp}, P_{mp}, P_{lp}	Pressure in high, medium and low pressure reservoirs (bars)
$P_{hp}^{max}, P_{mp}^{max}, P_{lp}^{max}$	Maximum pressure for high pressure, medium pressure and low pressure reservoirs (bars)
$P_{hp}^{min}, P_{mp}^{min}, P_{lp}^{min}$	Minimum pressure for high pressure, medium pressure and low pressure reservoirs (bars)
P_T	Target vehicle pressure (bars)
P_{veh}	Vehicle pressure (bars)
Q_{std}	Capacity of the compressor under standard conditions (Nm^3/h)
R	Universal gas constant (L bar/K mol)
t_s	Sampling period (s)
T	Absolute temperature (K)
u	State of compressor switch
v_{hp}, v_{mp}, v_{lp}	State of priority panel valves for high pressure, medium pressure and low pressure reservoirs
$v_{ohp}, v_{omp}, v_{olp}$	State of dispenser valves for high pressure, medium pressure and low pressure reservoirs
V	Volume of cascade reservoir tanks (L)
V_{hp}, V_{mp}, V_{lp}	Volume of high, medium and low pressure reservoirs (L)
z	Compressibility factor of CNG
ϱ	Weighting factor for the upper layer
ς	Weighting factor for the lower layer
γ	ratio of specific heats
$\rho_{a,std}$	Density of air under standard conditions (kg/m^3)
$\rho_{hp}, \rho_{mp}, \rho_{lp}$	Density of gas in high pressure, medium pressure and low pressure reservoirs (kg/m^3)

References

1. Davies, J.; Grant, M.; Venezia, J.; Aamidor, J. US Transportation Sector Greenhouse Gas Emissions: Trends, Uncertainties and Methodological Improvements. In *TRB 2007 Annual Meeting*; Transport Research Board: Washington, DC, USA, 2007.
2. Yeh, S. An empirical analysis on the adoption of alternative fuel vehicles: The case of natural gas vehicles. *Energy Policy* **2007**, *35*, 5865–5875. [CrossRef]
3. Ou, X.; Zhang, X.; Zhang, X.; Zhang, Q. Life cycle GHG of NG-based fuel and electric vehicle in China. *Energies* **2013**, *6*, 2644–2662. [CrossRef]
4. Mikolajková-Alifov, M.; Pettersson, F.; Björklund-Sänkiaho, M.; Saxén, H. A Model of Optimal Gas Supply to a Set of Distributed Consumers. *Energies* **2019**, *12*, 351. [CrossRef]
5. Kuby, M. The opposite of ubiquitous: How early adopters of fast-filling alt-fuel vehicles adapt to the sparsity of stations. *J. Transp. Geogr.* **2019**, *75*, 46–57. [CrossRef]
6. Chala, G.; Abd Aziz, A.; Hagos, F. Natural Gas Engine Technologies: Challenges and Energy Sustainability Issue. *Energies* **2018**, *11*, 2934. [CrossRef]
7. Greene, D.L. Survey evidence on the importance of fuel availability to the choice of alternative fuels and vehicles. *Energy Stud. Rev.* **1998**, *8*, 215–231. [CrossRef]
8. Newhouse, N.L.; Liss, W.E. *Fast Silling of NGV Fuel Containers*; Technical Report; SAE Technical Paper; SAE: Warrendale, PA, USA, 1999.
9. Chen, S.; Xie, G.; LI, Q.; Chang, K. Comparison among LNG, CNG and L-CNG Filling Stations. *Gas Heat* **2007**, *7*, 006.
10. Khadem, J.; Saadat-Targhi, M.; Farzaneh-Gord, M. Mathematical modeling of fast filling process at CNG refueling stations considering connecting pipes. *J. Nat. Gas Sci. Eng.* **2015**, *26*, 176–184. [CrossRef]
11. Albadi, M.H.; El-Saadany, E. A summary of demand response in electricity markets. *Electr. Power Syst. Res.* **2008**, *78*, 1989–1996. [CrossRef]
12. Kagiri, C.; Zhang, L.; Xia, X. Optimal energy cost management of a CNG fuelling station. In Proceedings of the Control Conference Africa, Johannesburg, South Africa, 7–8 December 2017; Volume 50, pp. 94–97.
13. Khamforoush, M.; Moosavi, R.; Hatami, T. Compressed natural gas behavior in a natural gas vehicle fuel tank during fast filling process: Mathematical modeling, thermodynamic analysis, and optimization. *J. Nat. Gas Sci. Eng.* **2014**, *20*, 121–131. [CrossRef]

14. Kountz, K.; Blazek, C.; Liss, W. A new natural gas dispenser control system. In *International Gas Research Conference*; Government Institutes Inc.: Rockville, MD, USA, 1998; Volume 4, pp. 135–145.
15. Kountz, K.J.; Blazek, C.F.; Christopher, F. *NGV Fuelling Station and Dispenser Control Systems*; Technical Report; Gas Research Institute: Chicago, IL, USA, 1997.
16. Kountz, K. Modeling the fast fill process in natural gas vehicle storage cylinders. In *207th ACS National Meeting-Division of Fuel Chemistry;* American Chemical Society: San Diego, CA, USA, 1994.
17. Farzaneh-Gord, M. Compressed natural gas-Single reservoir filling process. *Int. Gas Eng. Manag.* **2008**, *48*, 16–18.
18. Thomas, G.; Goulding, J.; Munteam, C. *Measurement, Approval and Verification of CNG Dispensers*; Technical Report; National Weights and Measures Laboratory: London, UK, 1999.
19. Farzaneh-gord, M.; Hashemi, S.; Farzaneh-kord, A. Thermodynamics Analysis of Cascade Reserviors Filling Process of Natural Gas Vehicle Cylinders. *World Appl. Sci.* **2008**, *5*, 143–149.
20. Deymi-Dashtebayaz, M.; Gord, M.F.; Rahbari, H.R. Studying transmission of fuel storage bank to NGV cylinder in CNG fast filling station. *J. Braz. Soc. Mech. Sci. Eng.* **2012**, *34*, 429–435. [CrossRef]
21. Frick, M.; Axhausen, K.W.; Carle, G.; Wokaun, A. Optimization of the distribution of compressed natural gas (CNG) refueling stations: Swiss case studies. *Transp. Res. Part D Transp. Environ.* **2007**, *12*, 10–22. [CrossRef]
22. Bang, H.J.; Stockar, S.; Muratori, M.; Rizzoni, G. Modeling and analysis of a CNG residential refueling system. In *ASME 2014 Dynamic Systems and Control Conference*; American Society of Mechanical Engineers: New York, NY, USA, 2014.
23. Smith, M.; Gonzales, J. *Costs Associated with Compressed Natural Gas Vehicle Fueling Infrastructure*; Technical Report; National Renewable Energy Laboratory (NREL): Golden, CO, USA, 2014.
24. Kagiri, C.; Zhang, L.; Xia, X. Compressor and priority panel optimization for an energy efficient CNG fuelling station. In Proceedings of the IEEE 11th Asian Control Conference (ASCC), Gold Coast, QLD, Australia, 17–20 December 2017; pp. 2200–2203.
25. Xia, X.; Zhang, J.; Cass, W. Energy management of commercial buildings-a case study from a POET perspective of energy efficiency. *J. Energy S. Afr.* **2012**, *23*, 23–31.
26. Spees, K.; Lave, L.B. Demand response and electricity market efficiency. *Electr. J.* **2007**, *20*, 69–85. [CrossRef]
27. Xia, X.; Zhang, L. Industrial energy systems in view of energy efficiency and operation control. *Annu. Rev. Control* **2016**, *42*, 299–308. [CrossRef]
28. Kagiri, C.; Zhang, L.; Xia, X. Optimization of a compressed natural gas station operation to minimize energy cost. In Proceedings of the 9th International Conference on Applied Energy, Cardiff, UK, 21–24 August 2017; pp. 2003–2008.
29. Kagiri, C.; Wanjiru, E.M.; Zhang, L.; Xia, X. Optimized response to electricity time-of-use tariff of a compressed natural gas fuelling station. *Appl. Energy* **2018**, *222*, 244–256. [CrossRef]
30. Allgöwer, F.; Zheng, A. *Nonlinear Model Predictive Control*; Birkhäuser: Basel, Switzerland, 2012; Volume 26.
31. Qin, S.J.; Badgwell, T.A. A survey of industrial model predictive control technology. *Control Eng. Pract.* **2003**, *11*, 733–764. [CrossRef]
32. Xia, X.; Zhang, J. Operation efficiency optimisation modelling and application of model predictive control. *IEEE/CAA J. Autom. Sin.* **2015**, *2*, 166–172.
33. Kountz, K.J.; Liss, W.E.; Blazek, C.F. Automated Process and System for Dispensing Compressed Natural Gas. U.S. Patent 5,810,058, 22 September 1998.
34. Shipley, E. Study of Natural Gas Vehicles (NGV) during the Fast Fill Process. Master's Thesis, West Virginia University, Morgantown, WV, USA, 2002.
35. Nguyen, H.H.; Uraikul, V.; Chan, C.W.; Tontiwachwuthikul, P. A comparison of automation techniques for optimization of compressor scheduling. *Adv. Eng. Softw.* **2008**, *39*, 178–188. [CrossRef]
36. Nguyen, H.H.; Chan, C.W. Applications of artificial intelligence for optimization of compressor scheduling. *Eng. Appl. Artif. Intell.* **2006**, *19*, 113–126. [CrossRef]
37. Mathaba, T.; Xia, X.; Zhang, J. Analysing the economic benefit of electricity price forecast in industrial load scheduling. *Electr. Power Syst. Res.* **2014**, *116*, 158–165. [CrossRef]
38. Wanjiru, E.M.; Xia, X. Energy-water optimization model incorporating rooftop water harvesting for lawn irrigation. *Appl. Energy* **2015**, *160*, 521–531. [CrossRef]
39. Saadat-Targhi, M.; Khadem, J.; Farzaneh-Gord, M. Thermodynamic analysis of a CNG refueling station considering the reciprocating compressor. *J. Nat. Gas Sci. Eng.* **2016**, *29*, 453–461. [CrossRef]

40. Calvert, J.G. Glossary of atmospheric chemistry terms (Recommendations 1990). *Pure Appl. Chem.* **1990**, *62*, 2167–2219. [CrossRef]
41. Farzaneh-Gord, M.; Deymi-Dashtebayaz, M. Optimizing natural gas fueling station reservoirs pressure based on ideal gas model. *Pol. J. Chem. Technol.* **2013**, *15*, 88–96. [CrossRef]
42. Cilliers, C.; van der Zee, L.; Kleingeld, M. Cost savings on mine dewatering pumps by reducing preparation-and comeback loads. In Proceedings of the IEEE International Conference on the Industrial and Commercial Use of Energy (ICUE), Cape Town, South Africa, 19–20 August 2014; pp. 1–8.
43. Cilliers, C. Cost Savings on Mine Dewatering Pumps by Reducing Preparation- and Comeback Loads. Master's Thesis, North West University, Potchefstroom, South Africa, 2014.
44. van Tonder, A.; Kleingeld, M.; Marais, J. Investigating demand response potential in a mining group. In Proceedings of the IEEE Industrial and Commercial Use of Energy Conference (ICUE), Cape Town, South Africa, 20–21 August 2013; pp. 1–5.

Hindawi
Complexity
Volume 2019, Article ID 4824837, 15 pages
https://doi.org/10.1155/2019/4824837

Research Article

A Competitive Swarm Optimizer-Based Technoeconomic Optimization with Appliance Scheduling in Domestic PV-Battery Hybrid Systems

Bo Wang,[1] **Yanjing Li**,[1] **Fei Yang**,[2] **and Xiaohua Xia**[3]

[1]*Key Laboratory of Ministry of Education for Image Processing and Intelligent Control, Artificial Intelligence and Automation School, Huazhong University of Science and Technology, Wuhan 430074, China*
[2]*Department of Artificial Intelligence and Automation, School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China*
[3]*Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0028, South Africa*

Correspondence should be addressed to Bo Wang; wb8517@hust.edu.cn

Received 19 July 2019; Accepted 28 August 2019; Published 13 October 2019

Guest Editor: Chun Wei

A technoeconomic optimization problem for a domestic grid-connected PV-battery hybrid energy system is investigated. It incorporates the appliance time scheduling with appliance-specific power dispatch. The optimization is aimed at minimizing energy cost, maximizing renewable energy penetration, and increasing user satisfaction over a finite horizon. Nonlinear objective functions and constraints, as well as discrete and continuous decision variables, are involved. To solve the proposed mixed-integer nonlinear programming problem at a large scale, a competitive swarm optimizer-based numerical solver is designed and employed. The effectiveness of the proposed approach is verified by simulation results.

1. Introduction

Making best use of renewable energies has been a topic that receives continuous attention [1]. The photovoltaic (PV) energy is one of the most concerned renewable energy because of the ubiquity of solar irradiation and very low carbon emission [2]. The PV energy generation is therefore integrated into power grids in many countries [3]. However, because of PV energy's intermittent nature, it is difficult to use PV energy alone to support sustained power demands in the complicated context, such as domestic electrical loads. A popular paradigm to utilize the PV power source is to integrate the PV energy into hybrid energy systems, where multiple power sources are adopted and dispatched cooperatively [4]. There were an enormous number of studies on hybrid system optimization over the past twenty years. Advanced technologies have been applied to the economic power dispatch problem in hybrid energy systems [5–7]. Most of such studies focused on power flow control strategies where the demand side was rather considered constraints in the system. There were also studies that introduced demand-side management into hybrid energy system management [8, 9], whereas these studies hardly explored the potentials of controlling both power flows and load behaviors. In general, the potentials of incorporating the power dispatch with demand-side management remain to be explored at the current stage. Indeed, such a problem is difficult because of the complex correlations between power sources and loads and its large-scale nature.

In this paper, a technoeconomic optimization problem is extended and improved to investigate further energy efficiency and economic potentials in such a domestic grid-connected PV-battery hybrid energy system, based on a series of previous studies by Tazvinga et al. [10–15]. There are two parts of the interventions: the power dispatch that decides which power is to be supplied to which load and appliance time scheduling that decides when to activate a specific domestic appliance to fulfill the user requirements.

The optimization is implemented over a finite time horizon. There are three objectives involved in the optimization. Firstly, the overall energy cost over the horizon has to be minimized. The energy cost mainly comes from the consumed grid power, based on the time-of-use (TOU) tariff [16]. The battery wear cost is also integrated into the overall cost. Secondly, the usage of renewable energy has to be maximized. Given that the overall power demand from the loads is considered constant, this objective is transformed into the grid power consumption minimization. Thirdly, the user satisfaction has to be maximized. An inconvenience indicator has been proposed by Setlhaolo and Xia [12] so that the overall difference between the scheduled appliance operation and the baseline appliance operation is calculated. The objective is thus introduced by minimizing such an inconvenience indicator. A weighted sum approach is employed to simultaneously optimize the three objectives, subject to a series of system constraints.

The main contributions of this study are listed as follows: Firstly, the hybrid system design is improved. The major improvement from the design perspective is that separate power dispatch is introduced to each connected appliance, instead of considering the electrical load (consisting of a variety of appliances) as a whole. This is realized by introducing additional power lines between each pair of the appliances and power sources. The power dispatch is thus managed in a more flexible way. In the previous model, all appliances are compelled to choose the same power source at one time. In the improved system design, additional power lines and switches are deployed such that appliances can choose the power sources themselves. The supply that combines various power sources provides the system more flexible power dispatch choices. For example, at peak hours, the battery bank cannot support heavy loads independently because of its capacity limitation. According to the previous design, the battery bank can only work for a short time in the late evening; otherwise, the power demands cannot be matched. In the new design, the battery bank has a much longer possible working time by only supporting a part of the loads, which leaves more space for the load scheduling. Improved flexibility brings more energy efficiency potentials, and thus, more economic benefits can be achieved.

Secondly, the mathematical formulation of the technoeconomic optimization is improved, as the dimension of decision variables is minimized. The aforementioned flexibility improves at the cost of additional decision variables, i.e., the ON/OFF states of the additional switches. The number of decision variables grows largely as the problem scale grows spatially with the number of switches and longitudinally with the number of control intervals, in comparison with the previous system design. To reduce the resulting computational burden, the interplay among the switch behaviors is identified. Some switch behavior constraints are involved; for example, the battery can only charge or discharge at one time, and the appliance can only have one active power line connection. The correlated switches manifest finite states; as a result, a set of discrete state variables are introduced. The values of the discrete variables indicate the combination of ON/OFF states of the correlated switches. Given the constraints, the state variables choose values from a limited range. In this way, the number of variables to describe the complex and interacting switch behaviors is minimized. As the ON/OFF states of the switches are a major part of the decision variables, the dimension of the decision variables is largely reduced; therefore, the computational burden is smaller such that the problem is more promising to be solved within limited time.

Thirdly, an advanced numerical solver is designed for the proposed problem. A major difficulty to implement the technoeconomic optimization is the solver. The investigated problem involves nonlinear objective functions and constraints, as well as continuous and discrete decision variables. It is thus a mixed-integer nonlinear programming (MINLP) problem. Furthermore, as mentioned above, the decision variable can be a large number, e.g., over 700. A proper solver to such a complicated and large-scale optimization problem is thus required. An intelligent optimization algorithm, namely, the competitive swarm optimizer (CSO), is employed to design the numerical solver. Cheng and Jin firstly proposed the CSO algorithm to solve large-scale optimization problems [17]. The CSO algorithm is designed on the basis of the particle swarm optimization (PSO) algorithm with a very different searching mechanism. In the PSO algorithm, the term “particle” is employed to refer to the individual solutions. The particles are characterized with two vectors, namely, the position and velocity vectors. The position vector describes the value of a solution and the velocity vector the incremental of the value. The PSO updates the position vector with the velocity vector via interacting with the global best position in the swarm (the population of individuals) and the personal best position in history [18]. The CSO algorithm adopts the position and velocity vector modelling from PSO but employs a random pairwise competition mechanism such that the loser particle can learn from the winner particle to update its position and velocity. In this way, the CSO algorithm can reduce the opportunity of convergence to local optimum, thereby manifesting better and satisfying overall performances than large-scale PSO algorithms. The numerical solver is designed on the pairwise competition concept basis and modified to better match the investigated scenarios.

A case study is employed to test and verify the effectiveness of the proposed approach, where the power dispatch and appliance time scheduling on a daily basis are applied to a typical South African household hybrid system. To thoroughly investigate the effectiveness, results from three cases, where different objective functions are applied with the new flexible power dispatch and the previous dispatch method, are illustrated and analyzed. For all cases, the CSO-based numerical solver is employed, and thus, the power dispatch methods are focused and compared.

The remainder of this paper is structured as follows: Section 2 introduces the hybrid system component modelling. Section 3 takes advantage of the component modelling to formulate the technoeconomic optimization problem. Section 4 describes the CSO-based numerical solver. Section 5 shows the case study with simulation results and analysis. Section 6 draws the conclusion.

2. Domestic PV-Battery Hybrid System

A domestic grid-connected PV-battery hybrid system is hereby employed as the investigated hybrid system. The general layout of the PV-battery hybrid system is illustrated in Figure 1. The main purpose of such a system is to supply the daily activities of a number of domestic appliances, e.g., electrical water heater (EWH), stove, television set, and washing machine. The involved appliances are connected to both the power grid and the PV system. A battery bank is also introduced to facilitate the power dispatch. The battery bank is able to charge from the power sources, which in this case are the power grid and PV system, and discharge to supply the appliances. In order to make use of all possible power sources in the system, a power management unit (PMU) is thereby introduced to implement the energy conversion and the power dispatch. In this way, the PMU manages the operation of the system, including (1) the selection of energy flow to support the active appliances, (2) the time scheduling of the appliances, and (3) the energy conversion and voltage/current matching. It is clear that the PMU is the central piece of the hybrid system that manages from power quality to energy balance. An assumption is employed that the voltage/current matching is well maintained by the PMU. Our investigation focuses on the power scheduling.

The power management diagram is depicted in Figure 2. There are several components in the PMU. From the PV side, there are a solar charge controller and inverter. The charge controller integrates a DC/DC converter to tune the PV output to match the DC loads, including the battery. The inverter receives input from the charge controller and battery and converts the DC power inputs into the AC loads, which in this case refer to the appliances. From the grid side, there is an AC charger integrating an AC/DC converter that allows the battery be charged by the grid. There are also a number of controllable switches to implement switching control strategies. As Figure 2 depicts, there is one switch for each power line. The ON/OFF states of the switches control the power flows in the system, i.e., the pair of the appliance and its supplier. For each connected appliance, there are a set of switches that control whether the appliance is supplied by the PV, the battery, or the power grid. The arrows on each power line indicate the direction of the power flow. As mentioned above, the additional switches are deployed such that any connected appliance can choose among multiple power sources by adjusting the ON/OFF status of switches. Taking advantage of the developing smart grid and smart building technologies, appliances are equipped with open communication interfaces, which allow the PMU to schedule the activities of the appliances and switches simultaneously, in both wired and wireless manners. As a result, the system design in Figure 2 becomes feasible in practice.

Remark 1. There are more and more domestic loads that can be made DC in the modern daily life; for example, an AC light bulb can be replaced with a DC light-emitting diode (LED) bulb. In the future, it is possible to connect the DC loads directly to the DC power sources, e.g., the PV and battery bank. Such a system may reduce the operational cost of renewable energy resources. However, currently, most domestic appliances are made AC for the sake of standardization. Consequently, the system design mainly focuses on the AC loads at the current stage; therefore, all involved power sources are tuned to be AC suppliers. Direct connections between the DC suppliers and the DC loads can be involved in the future design.

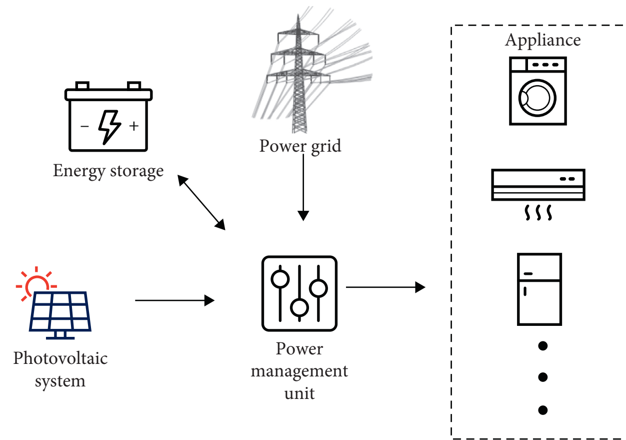


Figure 1: General layout of the hybrid system.

Assume that there are n appliances connected to the system. A set of binary variables are employed to denote the ON/OFF state of the switches. Let k denote the time instant during operation, $g_1(k)$ denotes the ON/OFF state of the switch between the PV charge controller and the battery, $g_2(k)$ the charge controller and the inverter, $g_3(k)$ the AC charger and the battery, and $g_4(k)$ the battery and the inverter. $g_{21}(k), g_{22}(k), \ldots, g_{2n}(k)$ denote the switches between the PV and the appliances, $g_{41}(k), g_{42}(k), \ldots, g_{4n}(k)$ the battery and the appliances, and $g_{51}(k), g_{52}(k), \ldots, g_{5n}(k)$ the grid and the appliances. t denotes the time instant over the operation. $P_1(k)$ and $P_4(k)$ denote the charging power from the PV and the grid, respectively. $P_2(k)$, $P_4(k)$, and $P_5(k)$ denote the power outputs of the respective power sources.

In this system, the operation of appliances is managed together with the power flows. In this study, the operation management is actually time scheduling, as the powers of appliances are considered known a priori and invariable. The time scheduling is implemented simultaneously with the power flow management such that the supply can match the demand and achieve higher energy efficiency potentials.

The mathematical formulations of the behavior for each component in the system are introduced as follows.

2.1. PV Systems. The PV consists of arrays of solar cells such that the solar energy is converted into electrical power. The converted power is proportional to the solar irradiation and the size of PV panels. As an alternative power source to the power grid, the power output is a major concern of the PV system. It is formulated as follows:

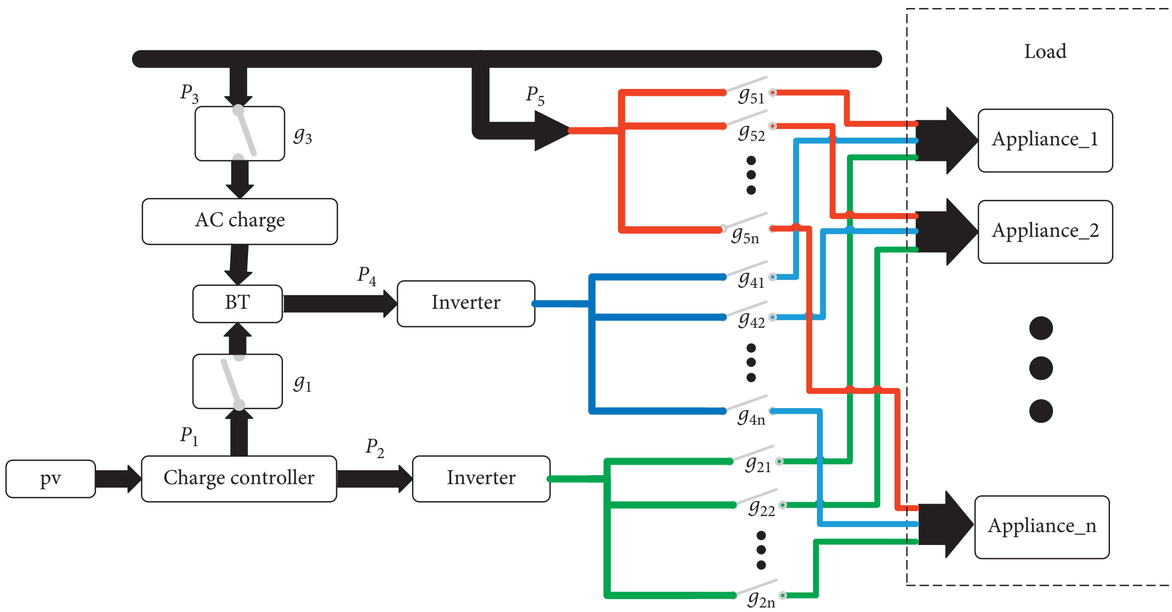


FIGURE 2: Schematic of the investigated power management system.

$$P_{pv} = \eta_{pv} I_{pv} A_c, \tag{1}$$

where P_{pv} denotes the hourly PV power output (kW), η_{pv} denotes the efficiency of the solar cells, I_{pv} denotes the hourly solar irradiation per unit area (kW/m^2), and A_c indicates the area of the PV panels that receive solar irradiation (m^2). There is an intermittent nature of the PV system; that is, I_{pv} can be absent at some sampling instants. An output profile of the PV system is therefore required. Usually, the PV outputs over succeeding 24 hours are predictable [19]. Q_{pv} denotes a time period where $P_{pv} > 0$. For $k \notin Q_{pv}$, P_{pv} is considered to be zero given a very small I_{pv}. A PV output profile is usually employed to indicate the hourly PV power outputs in a day . Q_{pv} can be identified via the PV output profile.

2.2. Battery Bank. The battery bank charges from other power sources and discharges to support electrical load activities. The battery bank behaviors are dynamic owing to the complicated scheduling of both power sources and appliances. The state of charge (SOC) is employed to characterize the battery bank status. The dynamics of the SOC can be formulated as follows:

$$\begin{aligned} \mathrm{Soc}(k+1) = {} & \mathrm{Soc}(k) + \eta_B g_1(k) P_1(k) + \eta_B \eta_c g_3(k) P_3(k) \\ & - g_4(k) P_4(k), \end{aligned} \tag{2}$$

where $\mathrm{Soc}(k)$ denotes the SOC at sampling instant k; $g_1(k)$, $g_3(k)$, and $g_4(k)$ are binary variables that denote the ON/OFF state of the respective switches at instant k, as depicted in Figure 2; $P_1(k)$, $P_3(k)$, and $P_4(k)$ are the aforementioned power outputs; η_c denotes the energy conversion efficiencies of the AC charger; and η_B denotes the battery charging efficiency during operation. Given the current-stage battery system limitations, the simultaneous charging from two different sources or simultaneous charging and discharging are considered unpermitted. A constraint of the switches must be taken into account:

$$g_1(k) + g_3(k) + g_4(k) \le 1 \tag{3}$$

such that only one of the switches g_1, g_3, and g_4 can be turned on at the same time. Following (2), the SOC at a given time τ can be formulated as follows:

$$\begin{aligned} \mathrm{Soc}(\tau) = {} & \mathrm{Soc}(0) + \eta_B \sum_{k=0}^{\tau} g_1(k) P_1(k) + \eta_B \eta_c \sum_{k=0}^{\tau} g_3(k) P_3(k) \\ & - \sum_{k=0}^{\tau} g_4(k) P_4(k), \end{aligned} \tag{4}$$

where $\mathrm{Soc}(0)$ denotes the initial state of the battery bank. $\mathrm{Soc}(k)$ is subject to the following constraint:

$$C^{\min} \le \mathrm{Soc}(k) \le C^{\max}, \tag{5}$$

where $C^{\min}$ and $C^{\max}$ denote the minimum and maximum available capacity (kWh), respectively.

The battery wear level is also evaluated. The wear cost is formulated as follows:

$$J_B(\tau) = \varphi_b C_D(\tau) \frac{B_C}{\mathrm{TH}}, \tag{6}$$

where $C_D(\tau) = \sum_{k=0}^{\tau} P_4(\tau)$ denotes the overall throughput of the battery bank until instant τ and B_C/TH denotes the battery wear cost per 1 kWh from throughput energy, in which B_C denotes the battery cost and TH denotes the overall throughput energy. The calculation of B_C/TH can be found from previous studies [15, 20, 21].

2.3. Power Grid. From the hybrid system viewpoint, the grid supplies infinite and stable electricity at an alternative voltage level of 220 V. The grid power comes with a price, which results in the major operational costs. As mentioned above, a TOU tariff is introduced such that the demand response can be implemented. Let $\rho(k)$ denote the TOU electricity price at time k of a day. $\rho(k)$ changes according to which period the time k lies within. The overall operational cost at a given time τ can be formulated as follows:

$$\mathrm{Cost}(\tau) = \sum_{k=0}^{\tau} \rho(k)\left[P_3(k) + P_5(k)\right]. \tag{7}$$

2.4. Appliances. The electrical loads consist of the load profiles of all appliances connected to the hybrid system. The boundary identification must be conducted before system design. Given the domestic scenario of the investigated system, several reasonable simplifications are made to characterize the demand-side activities. Firstly, all appliances within the system require standard AC power. Secondly, each appliance is subject to a respective constant operation duration. Control strategies can be involved to determine the operation of such appliances [22, 23].

According to previous studies [12–14], from the time-scheduling perspective, domestic appliances can be categorized into three types: the flexible loads, the shiftable loads, and the fixed loads. The flexible loads' working times can be scheduled freely, at any favorable working time. The shiftable loads' working times can be scheduled within a preferable but limited time period. The fixed loads' working times are fixed, unchangeable in any case. For an arbitrary appliance from any type, the operation duration is constant such that the time scheduling can be characterized as the selection of the starting instant.

In the investigated system, an appliance is connected to the PV, battery bank, and grid. An appliance i from the n appliances is given. As depicted in Figure 2, there is one switch for each power source connection. The ON/OFF states of the three switches at instant k are denoted by $g_{2i}(k)$, $g_{4i}(k)$, and $g_{5i}(k)$. The simultaneous supply from multiple power sources is unpermitted; therefore, a switch constraint is introduced as follows:

$$g_{2i}(k) + g_{4i}(k) + g_{5i}(k) \le 1, \quad i = 1, 2, \ldots, n. \tag{8}$$

Let St_i denote the starting instant of the appliance i and D_i the operation duration. The appliance continuously operates until the end instant, denoted by $En_i = St_i + D_i - 1$. A continuous operation constraint is introduced as Xia and Zhang proposed [24]:

$$\sum_{k=0}^{N-D_i+1} u_i(k)u_i(k+1)u_i(k+2)\ldots u_i\left(k+D_i-1\right) = 1, \quad i = 1, 2, \ldots, n, \tag{9}$$

where $u_i(k)$ is a binary variable that indicates whether the appliance i is active at time k, i.e., the time schedule, and N denotes the length of a finite scheduling period. In this case, $u_i(k) = 1$ if $St_i \le k \le En_i$, and $u_i(k) = 0$ if otherwise. It is notable that the continuous operation constraint is associated with the actual switch behavior in the following way:

$$\begin{cases} g_{2i}(k) + g_{4i}(k) + g_{5i}(k) = 1, & \text{if } u_i(k) = 1, \\ g_{2i}(k) + g_{4i}(k) + g_{5i}(k) = 0, & \text{if } u_i(k) = 0. \end{cases} \tag{10}$$

Let $\overline{p}_i(t)$ with $t = 0, 1, 2, \ldots, D_i - 1$ denote the load profile over the operation period $(0, D_i - 1)$ of the appliance i. Assuming that $\overline{p}_i(t)$ are known a priori, the power demand $p_i(k)$ can thus be determined: if $k = Si + t$, $p_i(k) = \overline{p}_i(t)$; otherwise, $p_i(k) = 0$.

2.5. System Constraints. A series of constraints must be introduced to the system such that the operation requirements are all satisfied and none of the physical laws is violated.

(a) The energy balance must be fulfilled anytime during operation, as indicated by the following equation:

$$\begin{bmatrix} \eta_{I2}P_2(k) \\ \eta_{I4}P_4(k) \\ P_5(k) \end{bmatrix} = \sum_{i=1}^{n} \begin{bmatrix} g_{2i}(k) \\ g_{4i}(k) \\ g_{5i}(k) \end{bmatrix} p_i(k), \tag{11}$$

where $P_2(k)$, $P_4(k)$, and $P_5(k)$ are the power flows from the PV output, battery bank discharge, and power grid supply and η_{I2} and η_{I4} denote the inverter efficiency of the PV and battery bank, respectively. For the convenience of calculation, a logical state variable $sw_i(k)$ is employed to describe the combination of $g_{2i}(k)$, $g_{4i}(k)$, and $g_{5i}(k)$. $sw_i(k)$ chooses a value from $\{0, 1, 2, 3\}$. $sw_i(k) = 0$ indicates the state that all switches are turned off and the switch states equal to 0. $sw_i(k) = 1$ indicates that only $g_{2i}(k) = 1$, $sw_i(k) = 2$ that only $g_{4i}(k) = 1$, and $sw_i(k) = 3$ that only $g_{5i}(k) = 1$.

(b) The capacity constraints must be followed such that the power flow is kept within the range of the component capacity. The power flows in the system are thus limited. The power flows $P_1(k)$ and $P_2(k)$ are subject to the following constraint:

$$0 \le P_1(k) + P_2(k) \le \eta_s P_{pv}(k), \tag{12}$$

where $P_{pv}(k)$, as mentioned above, is the PV output at time k and η_s denotes the charge controller efficiency. The power flows $P_2(k)$ and $P_4(k)$ as the supplier are subject to the following constraints:

$$\begin{cases} 0 \le \eta_{I2}P_2(k) \le P_{I2}(k), \\ 0 \le \eta_{I4}P_2(k) \le P_{I4}(k), \end{cases} \tag{13}$$

where $P_{I2}(k)$ and $P_{I4}(k)$ denote the inverter capacity of the PV and battery bank, respectively. The power flows $P_3(k)$ and $P_5(k)$ are subject to the following constraint:

$$0 \le g_3(k)P_3(k) + P_5(k) \le P_G^{\max}, \tag{14}$$

where $P_G^{\max}$ denotes the allocated grid maximum power for this grid-connected system. The power flows $P_2(k)$, $P_4(k)$, and $P_5(k)$ are obtained from the energy balance equation (11). $P_1(k)$ and $P_3(k)$ are decided by the scheduling algorithm, and they are part of the decision variables.

(c) The switch control strategies are also subject to constraints that prevent infeasible switch behaviors. These constraints are formulated in preceding sections along with the system component modelling, i.e., constraints (3) and (8)–(10). Some further associations of the switch behaviors are identified as follows:

$$\begin{cases} g_1(k) = 1, & \text{if } P_1(k) > 0, \\ g_1(k) = 0, & \text{if } P_1(k) = 0, \\ g_3(k) = 1, & \text{if } P_3(k) > 0, \\ g_3(k) = 0, & \text{if } P_3(k) > 0, \end{cases} \tag{15}$$

$$\begin{cases} g_2(k) = 1, & \text{if } \prod_{i=1}^{n}\left(\mathrm{sw}_i(k) - 1\right) = 0, \\ g_2(k) = 0, & \text{if otherwise}, \\ g_4(k) = 1, & \text{if } \prod_{i=1}^{n}\left(\mathrm{sw}_i(k) - 2\right) = 0, \\ g_4(k) = 0, & \text{if otherwise}. \end{cases} \tag{16}$$

In this way, the states of $g_1(k)$–$g_4(k)$ at time k can be identified from the values of $P_1(k)$, $P_3(k)$, and $\mathrm{sw}_i(k)$ with $i = 1, 2, \ldots, n$.

3. Problem Statement

The primary management objective of the investigated optimization problem is cost minimization. Secondly, the renewable energy penetration is involved; that is, the usage of grid power should be minimized as well. Owing to the TOU tariff, the two objectives manifest certain differences and must be equally considered in scheduling. Furthermore, the user satisfaction in time scheduling is taken into account. All these objectives are evaluated on a finite horizon basis. Let T denote the number of sampling instants. The optimization problem is formulated as follows.

3.1. Decision Variables. The involved decision variables consist of three parts: (1) the switch control strategy decision variables, (2) the charging power control variables, and (3) the appliance time-scheduling decision variables. It is given that $i = 1, 2, \ldots, n$ in the following discussion.

The switch control variables are $\mathrm{sw}_i(k)$, which can characterize the status of most switches in the system as constraints (15) and (16) imply. Given constraints (8)–(10), it is unnecessary to cover the whole finite horizon. For the appliance i, $\mathrm{sw}_i(k) > 0$ when $St_i \le k \le En_i$, and $\mathrm{sw}_i(k) = 0$ if k lies outside the working period. Therefore, the minimum required control variables are $\mathrm{sw}_i(k)$ with $k \in [St_i, En_i]$. The dimension of this part is $\sum_{i=1}^{n} D_i$.

The charging power control variables are $P_1(k)$ and $P_3(k)$. Given constraints (3) and (15), $P_1(k) * P_3(k) = 0$ at any given instant k; therefore, the dimension of this part is T.

As mentioned above, the appliance time scheduling is simplified with the known a priori and constant load profile $\overline{p}_i(t)$ and operation duration D_i. Taking advantage of the knowledge, the scheduling is implemented with decision variables St_i. The dimension of this part is n.

The dimension of the optimization is therefore $\sum_{i=1}^{n} D_i + T + n$. Comparing with the previous study [15], the problem is extended to a higher dimension but simplified by taking advantage of $\mathrm{sw}_i(k)$ and the constraints such that the dimension of the optimization grows slower than the problem scale.

3.2. Objectives. The cost minimization objective is formulated as follows:

$$J_c = \sum_{k=0}^{T} \rho(k)\left[P_3(k) + P_5(k)\right] + \varphi_b J_B(T), \tag{17}$$

where φ_b denotes a weight that indicates the preferable importance of the battery wearout to the decision-maker.

The renewable energy penetration objective is formulated as follows:

$$J_e = \sum_{k=0}^{T}\left[P_3(k) + P_5(k)\right]. \tag{18}$$

The user satisfaction is evaluated via an inconvenience indicator, which is adopted from the study of Setlhaolo and Xia [12]:

$$\beta = \sum_{i=1}^{n} \gamma_i \sum_{k=0}^{T}\left[u_i^{bl}(k) - u_i(k)\right]^2, \tag{19}$$

where γ_i denotes an importance factor of the appliance i and $u_i^{bl}(k)$ denotes the baseline time schedule of the appliance i. (19) quantifies the difference between the baseline time schedule and the adopted time schedule. Given the system modelling and constraints in this study, the user satisfaction evaluation is simplified as follows:

$$\beta = \sqrt{\sum_{i=1}^{n} \gamma_i \left[St_i^{bl} - St_i\right]^2}, \tag{20}$$

where St_i^{bl} denotes the baseline starting instant. Such a difference has to be minimized such that the user remains happy with the optimized appliance time schedule.

3.3. Technoeconomic Optimization. Taking advantage of the preceding objective functions and constraint formulations, the technoeconomic optimization problem is obtained by minimizing the following objective function:

$$J\left(\mathrm{sw}_i(k), P_1(k), P_3(k), St_i\right) = \lambda_c J_c + \lambda_e J_e + \lambda_b \beta, \tag{21}$$

subject to the battery dynamics (2) and constraints (3), (8)–(10), and (11)–(16).

According to the above formulations, there are nonlinear objective functions and constraints, as well as continuous and discrete decision variables in the problem. They result in a mixed-integer nonlinear programming (MINLP) problem, at a relatively large scale. The general theoretic approach of solving an MINLP problem remains an open question; as a result, numerical solvers are widely employed. In the previous study [15], an OTPI toolbox https://www.inverseproblem.co.nz/OPTI/index.php/DL/DownloadOPTI/ in MATLAB was adopted as the numerical solver. The former solver took quite a large amount of time for calculation. In this study, the implementation of intelligent optimization algorithms on such a problem is investigated. A newly proposed algorithm named the competitive swarm optimizer (CSO) is adopted as the numerical solver. The introduction to the CSO-based solver comes in the following section.

4. Numerical Solver Design

4.1. Competitive Swarm Optimizer. In the CSO algorithm, let x denote a particle and w and l the indices of the winner and loser particles in a pair. Assume that it is the G-th iteration, and there have been $k-1$ competitions. After the k-th competition, the next-generation winner particle $x_{w,k}(G+1)$ remains the same as $x_{w,k}(G)$. The loser particle $x_{l,k}(G+1)$, namely, the position vector, is thereby updated as follows:

$$x_{l,k}(G+1) = x_{l,k}(G) + V_{l,k}(G+1), \tag{22}$$

where $V_{l,k}(G+1)$ is the next-generation velocity vector, updated as follows:

$$\begin{aligned} V_{l,k}(G+1) = {} & rn_1(k,G) V_{l,k}(G) + rn_2(k,G)\left[x_{w,k}(G) - x_{l,k}(G)\right] \\ & + \varphi rn_3\left[\overline{x}_k(G) - x_{l,k}(G)\right], \end{aligned} \tag{23}$$

where rn_1, rn_2, and rn_3 are random vectors; $\overline{x}_k(G)$ is the center of neighborhood filed particles, i.e., a set of particles that are close enough to $x_{l,k}(G)$; and φ is the weighting coefficient of $\overline{x}_k(G)$. Such a neighborhood field is predefined. There is a special case that the neighborhood covers the whole swarm, where $\overline{x}_k(G)$ indicates the global mean position of the particles at iteration G. The velocity and position vectors are employed for the continuous cases. In a discrete case, e.g., the decision variables $\mathrm{sw}_i(k)$ in this study, other update mechanisms must be employed. A crossover mechanism is hereby employed as follows:

$$\begin{cases} x_{l,k}(rn, G+1) = x_{w,k}(rn, G), \\ x_{l,k}(\overline{rn}, G+1) = x_{l,k}(\overline{rn}, G), \end{cases} \tag{24}$$

where rn denotes randomly generated indices of the components in a particle x and $\overline{rn}$ the unselected indices. (24) suggests that the winner particle $x_{w,k}(G)$ selects and copies a part of its components into the next-generation loser particle $x_{l,k}(G+1)$. The unselected components of $x_{l,k}(G+1)$ remain the same as $x_{l,k}(G)$. In this way, the loser particle can learn from the winner particle.

For an MINLP problem, there are simultaneously continuous and discrete components in a particle. In this case, the continuous part and discrete part are separated, (23) and (22) are implemented on the continuous part, and (24) is implemented on the discrete part. After the learning process, the two updated parts are combined again to obtain the particle of next generation. The theoretical proof of the convergence of the CSO algorithm can be referred to [17].

The pseudocode of the CSO according to the preceding introduction is illustrated in Algorithm 1 [25].

Remark 2. The introduced CSO algorithm is mainly designed for continuous problems. For discrete decision variables, e.g., binary variables or integers, the discrete PSO algorithm [26, 27] can facilitate the algorithm design. The pairwise competition can be further introduced to other evolutionary algorithms, such as differential evolution (DE).

Remark 3. Given the investigated MINLP problem (21), the original CSO algorithm cannot be applied in a straightforward way. Modifications that match the decision variables are to be introduced such that satisfying performances can be achieved.

4.2. Modified CSO-Based Solver. In order to implement the CSO algorithm on a constrained problem, a penalty function is introduced to the original objective function (17). Given that there are N_C constraints to a problem, the penalty function is formulated as follows:

$$P_{\mathrm{en}} = \sum_{k=0}^{T}\sum_{i=0}^{N_C} \omega_{\mathrm{Pen},i} P_{\mathrm{en},i}(k), \tag{25}$$

where

$$P_{\mathrm{en},i}(k) = \begin{cases} 0, & \text{if constraint } i \text{ is obeyed}, \\ M, & \text{if constraint } i \text{ is violated}, \end{cases} \tag{26}$$

where M is a large positive number and $\omega_{\mathrm{Pen},i}$ is the weighting factor for constraint i. A fitness function is thereby formulated with (17), (25), and (26):

$$f(x) = J_c + P_{\mathrm{en}}. \tag{27}$$

In this way, for a minimization problem, a particle becomes much less competitive when any of the constraints

Definition:
x: the particle;
P: the swarm;
np: the swarm size, i.e., the number of particles;
G: number of iterations;
w and l: the indices of winner and loser particles;
$f(\cdot)$: the fitness function, assuming that this is a minimization problem;
Terminal condition: the maximum number of iteration Mg is reached;

Begin
Initialize population $P(1)$ with np particles;
while $G = 1$ to Mg **do**
  $P(G+1) = \varnothing$;
  while $P(G) \neq \varnothing$ **do**
    Generate two random indices r_1 and r_2 from np;
    if $f(x_{r_1}) \leq f(x_{r_2})$ **then**
      $w = r_1, l = r_2$;
    else
      $w = r_2, l = r_1$;
    end if
    put $x_w(G)$ into $P(G+1)$;
    If x is coded as continuous variables, update $x_l(G)$ with (23) and (22);
    If x is coded as discrete variables, update $x_l(G)$ with (24);
    If x contains both continuous and discrete parts, update the two parts separately;
    put the updated loser particle $x_l(G)$ into $P(G+1)$;
    remove particles x_{r_1} and x_{r_2} from $P(G)$;
  end while
  $G = G + 1$;
end while
choose x_{best} the particle with the best fitness $\hat{f}(\cdot)$ from P_{Mg};
Return x_{best};
End

Algorithm 1: Pseudo code of the CSO algorithm.

is violated, given that a large positive will be added to the objective function.

After competition, the loser particle must learn from the winner particle. However, the MINLP search space and constraints in the investigated problem are quite complicated. The dynamics of the battery bank charging and discharging invoke further difficulty to search for the optimum. Consequently, the performance of Algorithm 1 is not satisfying. The following modifications are thereby employed to improve the performances on this specific problem.

Firstly, the learning strategies are modified to improve the searching efficiency. Let $\overline{f}(G)$ denote the mean fitness of the current swarm (at the G-th iteration). After competition, the fitness of both winner and loser particles is evaluated:

(i) If $f(x_{w,k}(G)) \geq \overline{f}(G)$, both winner and loser particles are considered inferior; therefore, the winner particle must learn from the global best particle $x_{\text{gbest}}(G)$, while the loser particle must learn from the winner.

(ii) If $f(x_{l,k}(G)) \leq \overline{f}(G)$, both winner and loser particles are considered superior; therefore, the winner particle is moved into $P(G+1)$, and the loser particle implements mutation subject to the genetic algorithm style.

(iii) If $f(x_{w,k}(G)) \leq \overline{f}(G) < f(x_{l,k}(G))$, the original CSO learning strategies are applied.

Secondly, constraints (3) and (15) are employed to generate battery bank charging and discharging states, namely, the "knowledge-guided solution filter". $\forall k$, $g_4(k)$ is firstly identified subject to (16). If $g_4(k) = 1$, then $P_1(k) = P_3(k) = 0$. If $g_4(k) = 0$, then $P_1(k)$ is randomly generated within $[0, P_{pv}(k) - P_2(k))$. If $g_4(k) = 0$ and $P_1(k) = 0$, then $P_3(k)$ is randomly generated. If the $\mathrm{Soc}(k)$ reaches its upper bound at time k, then $P_1(k) = P_3(k) = 0$. In this way, the charging and discharging decision variables are guaranteed to be feasible. The knowledge-guided solution filter reduces cost of trial and error during optimization such that the algorithm can persist searching within a feasible space.

Remark 4. According to simulations, the modified solver constantly outperforms the original CSO algorithm on the investigated problem. There lacks a theoretical analysis on the performances, whereas a hypothesis is made that the superior performances are resulted from the knowledge-guided solution filter. Wang and Zheng [28] reported that exploitation of the algorithm is enhanced by knowledge-based local search. Further details and investigations are expected in future works.

Table 1: Typical usage profiles and baseline time schedules.

Appliances Index i	Power (kW)	Duration (min)	Baseline St_i	Baseline En_i	Preferable range of St_i
Shiftable					
(1) EWH	3.0	120	31	42	[19, 31]
		120	104	115	[91, 121]
(2) Stove	2.5	30	32	34	[25, 55]
		50	113	117	[97, 127]
(3) Washing machine	0.5	60	109	114	[43, 133]
(4) Electric dryer	2.0	30	116	118	[49, 139]
Fixed					
(5) Refrigerator	0.1	1440	1	144	N/A
(6) Television set	0.2	180	104	121	N/A
Flexible					
(7) Dishwasher	1.8	150	116	130	[1, 130]
(8) Bread maker	1.5	150	118	132	[1, 130]

5. Simulation Results and Analysis

5.1. Case Study. The case study investigates the operation of a household, grid-connected, PV-battery hybrid energy system. The data are retrieved from the South African domestic appliance operation studies [13–15]. There are eight appliances connected in the system. The usage profile of the appliances is shown in Table 1 on a daily basis, where there are 144 time slots, a.k.a., sampling instants. Each time slot lasts 10 minutes. The scheduling horizon is 24 hours, i.e., a whole day. The adopted usage profile and baseline time schedule are for the working day scheduling for a typical South African home. To emphasize, the power of the appliances is measured average power. The baseline time schedule reflects the preferable time according to the inhabitant's habits. For example, the inhabitant turns on the electrical water heater (EWH) twice a day for the hot water demand. In the morning, the EWH is turned on at 5 : 00 am (the 31-st time slot), operates for two hours, and is turned off at 7 : 00 am (the 42-nd time slot) such that the user can use heated water after breakfast. In the afternoon, the EWH is turned on again at 5 : 10 pm (the 104-th time slot) and turned off at 7 : 10 pm (the 115-th time slot) such that the hot water for the evening can be ready. This is the most convenient EWH operation plan for the user. Similarly, the stove must be turn on twice for the cooking demands. The other appliances have to be turned on and off only once daily.

There are several further constraints with the given scenario. Firstly, for the shiftable appliances, a preferable range of starting time slots are given in Table 1. The flexible appliance can start anytime in a day, and the only requirement is that the operation must be finished before the end of the horizon. The fixed appliance cannot be scheduled; therefore, the preferable range is not applicable (N/A). Secondly, the washing machine and electrical dryer work in a sequence; that is, the dryer must start after the washing machine job is finished. In this case, it results in an additional constraint to the preceding ones:

$$St_4 \geq St_3 + D_3. \quad (28)$$

Table 2: PV system and battery bank parameters.

PV capacity P_{pv}	3.5 kW$_p$
Battery bank maximum capacity $C^{\max}$	5.04 kWh
Battery bank minimum capacity $C^{\min}$	2.52 kWh
Battery bank cost (ZAR)	R5826
Initial state of the battery bank	60%$C^{\max}$
AC charger efficiency η_c	85%
PV charge controller efficiency η_s	90%
PV inverter efficiency η_{I2}	95%
Battery bank inverter efficiency η_{I4}	95%
Battery bank charging efficiency η_B	80%

Table 3: Hourly PV output.

Time slot k	[0, 39)	[39, 45)	[45, 51)	[51, 57)
$P_{pv}(k)$ (kWh)	0	0.15	0.85	1.65
[57, 63)	[63, 69)	[69, 75)	[75, 81)	[81, 87)
2.35	2.9	3	2.95	2.55
[87, 93)	[93, 99)	[99, 105)	[105, 111)	[111, 144)
2	1.45	0.75	0.1	0

The adopted PV system and battery bank have their own limitations, as shown in Table 2. The PV system integrates 14 solar panels with the rated power of 0.25 kW. Therefore, the overall capacity, i.e., the rated output, of the PV system is 3.5 kW. Actually, the PV system output at any given time slot depends on the solar irradiation profile. Such a profile is possible to forecast 24 hours ahead of the scheduling [19]. In the case study, the timely PV output is identified based on an hourly profile in [29], as shown in Table 3, where Q_{pv} is $k \in [39, 144)$. The battery bank consists of 4 lead-acid batteries, each with 12 V rated voltage and 105 Ah rated capacity; that is, the overall capacity is 5.04 kWh. The battery cost is calculated in South African rand (ZAR), which is R5826. The lifespan of the battery bank is 1000 cycles at 50% depth of discharge (http://www.trojanbattery.com/markets/renewable-energy-re/); therefore, the wear cost per 1 kWh throughput energy is $5826/(1000 * 0.5 * 5.04) = 2.312$ R/kWh. $\varphi_b = 0.1$

Table 4: TOU tariff.

Time periods	Electricity price	Hours
Peak hours	R2.2225/kWh	[08 : 00, 11 : 00) ∪ [19 : 00, 21 : 00)
Standard hours	R0.6773/kWh	[07 : 00, 08 : 00) ∪ [11 : 00, 19 : 00) ∪ [21 : 00, 23 : 00)
Off-peak hours	R0.3656/kWh	[00 : 00, 07 : 00) ∪ [23 : 00, 24 : 00)

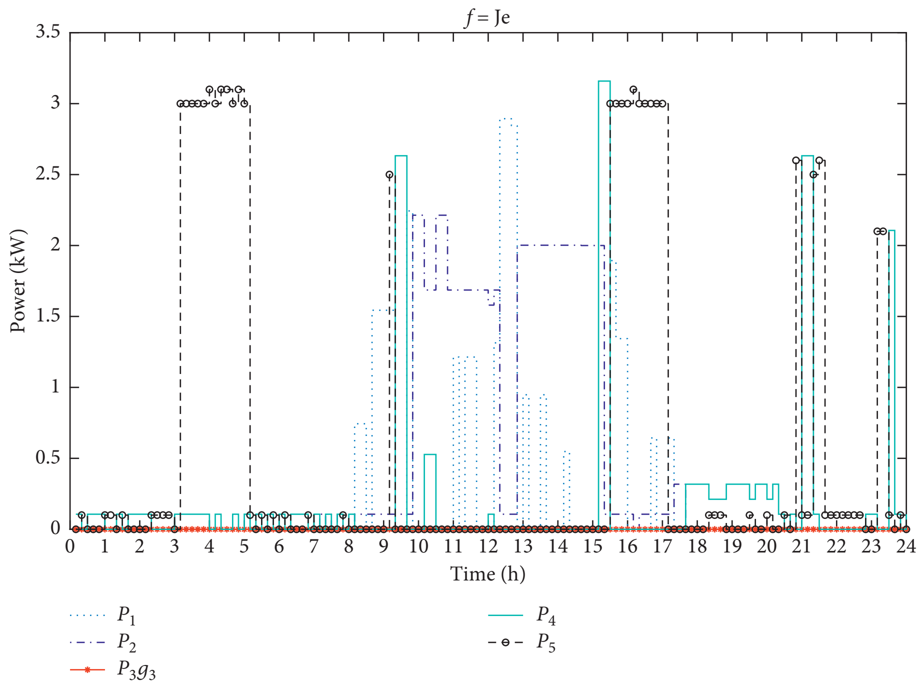


Figure 3: Power dispatch of the new design in case (i).

such that the usage of the battery is encouraged. The importance factors $\gamma_i = 1$ for each i, i.e., involved appliances, are considered equally important in this case. The efficiencies of the charge controller, AC charger, and inverters are given as well.

The power grid supply is described as follows: The maximum household current is 60 A, which is limited by the utility company. The charging power from the grid is considered constant in this case, which is 5 kW; that is, $P_3(k)$ can only be 0 or 5 kWh. Furthermore, the TOU tariff is adopted from the study in [13], as shown in Table 4.

5.2. Simulations. Simulations are programmed in C++ with the following running environment: the CPU is Inter Core i3-8100 CPU@3.60 GHz, the RAM is 16 GB, and the system is Windows 10 × 64. Three cases are adopted by adjusting the weighting factors in (21).

(i) $\lambda_e = 1, \lambda_c = 0$, and $\lambda_b = 0$ such that the optimization employs a single objective, i.e., the renewable energy penetration.

(ii) $\lambda_e = 0$, $\lambda_c = 1$, and $\lambda_b = 0$ such that only the cost minimization objective is optimized.

(iii) $\lambda_e = 1$, $\lambda_c = 1$, and $\lambda_b = 1$ such that the multi-objective optimization of (21) is implemented where J_c, J_e, and β are equally considered.

In each case, there are two demonstrated results. One result is from the proposed approach, and the other one is from the previous power dispatch model [15] as the comparative results. Both results are reported from the average of 20 runs, taking advantage of the proposed CSO-based optimizer. In the CSO algorithm, the swarm size is 1500 and the iteration number is 10000. The neighborhood field is defined to be the nearest superior particle and inferior particle. The details of such a neighborhood field can be referred to [30, 31].

5.3. Results and Analysis. The results are reported as follows:

(i) In the first case, the grid power supply is minimized to be 14 kWh, at the energy cost of R16.57 and inconvenience indicator of 14.25. The power dispatch is depicted in Figure 3. In the comparative result, the minimal grid power is 15.07 kWh, at the energy cost of R14.29 and inconvenience indicator of 14.59. The power dispatch is depicted in Figure 4. The improvement of the objective is 7.1%. The average running time is 18.43 minutes.

(ii) In the second case, the cost is minimized to be R7.06, while the overall power supply from the grid is 14.37 kWh. The inconvenience indicator is 13.89. The power dispatch is depicted in Figure 5. In the comparative result, the minimal cost is R7.72, with

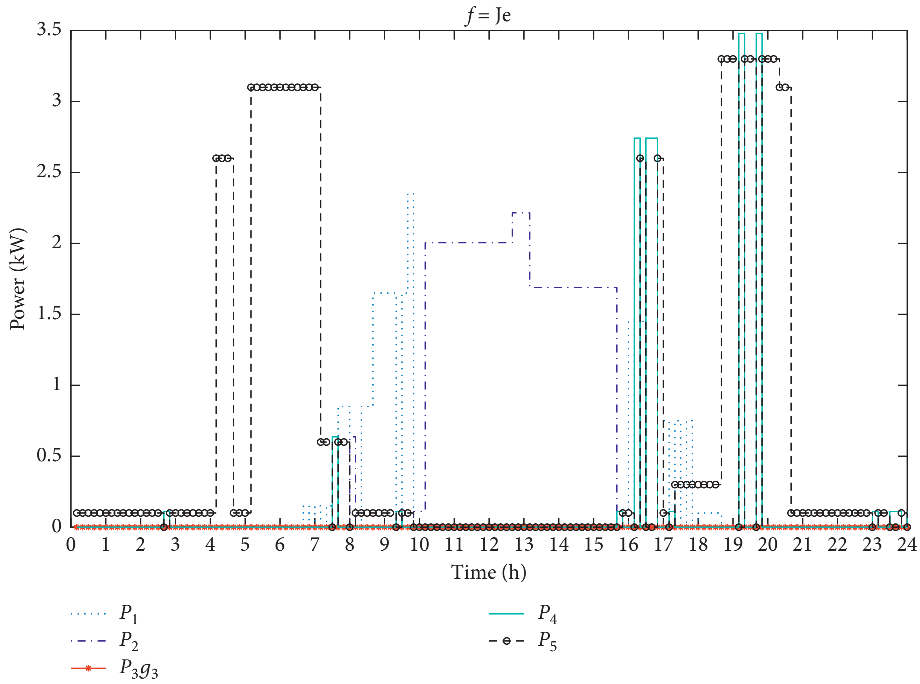


Figure 4: Power dispatch of the previous design in case (i).

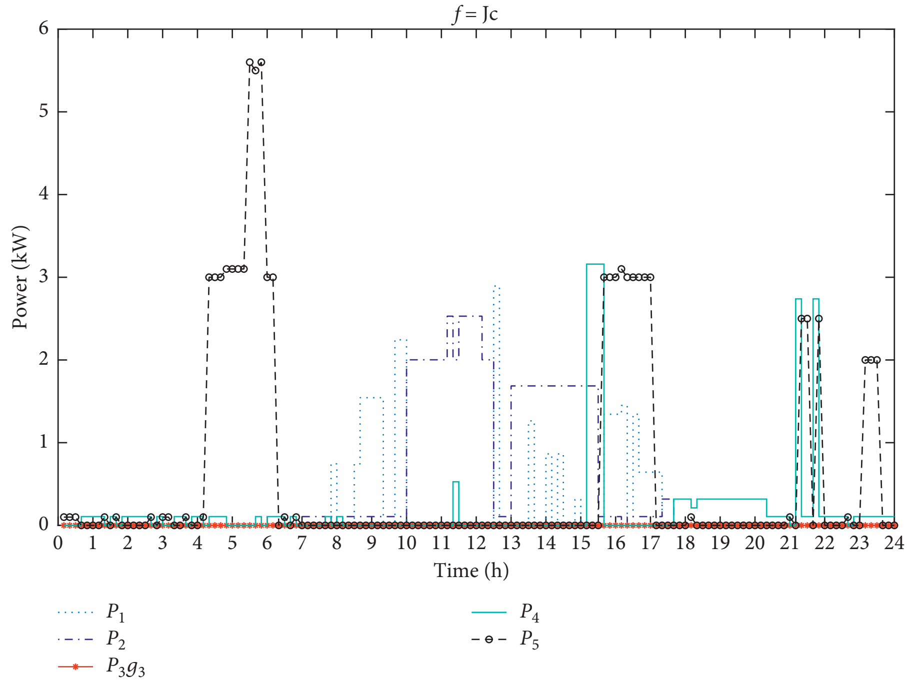


Figure 5: Power dispatch of the new design in case (ii).

the grid power supply of 15.57 kWh and inconvenience indicator of 15.03. The power dispatch is depicted in Figure 6. The improvement of the objective is 8.6%. The average running time is 18.36 minutes.

(iii) In the third case, the weighted sum of all objectives is minimized. The optimized objective function value is 34.64, when the energy cost is R8.29 and the grid power supply is 15.53 kWh, and the inconvenience indicator is 10.82. The power dispatch is depicted in Figure 7. In the comparative result, the objective function value is 39.97, where the energy cost is 9.53 and grid power is 18.78 kWh, with the inconvenience indicator of 11.66 as well. The power dispatch is depicted in Figure 8. The improvement of the objective is 13.3%. The average running time is 18.37 minutes.

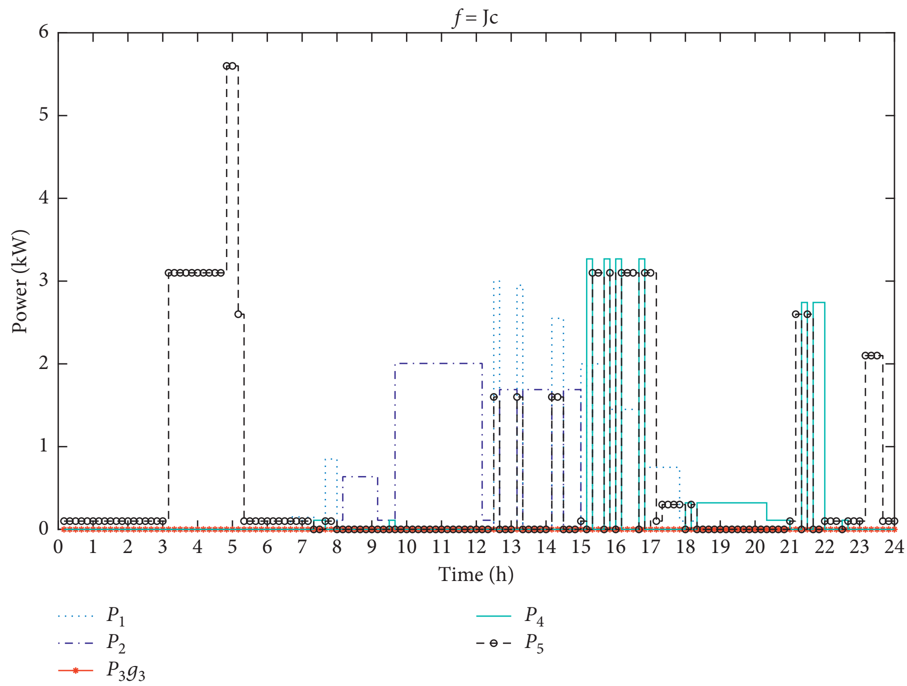


Figure 6: Power dispatch of the previous design in case (ii).

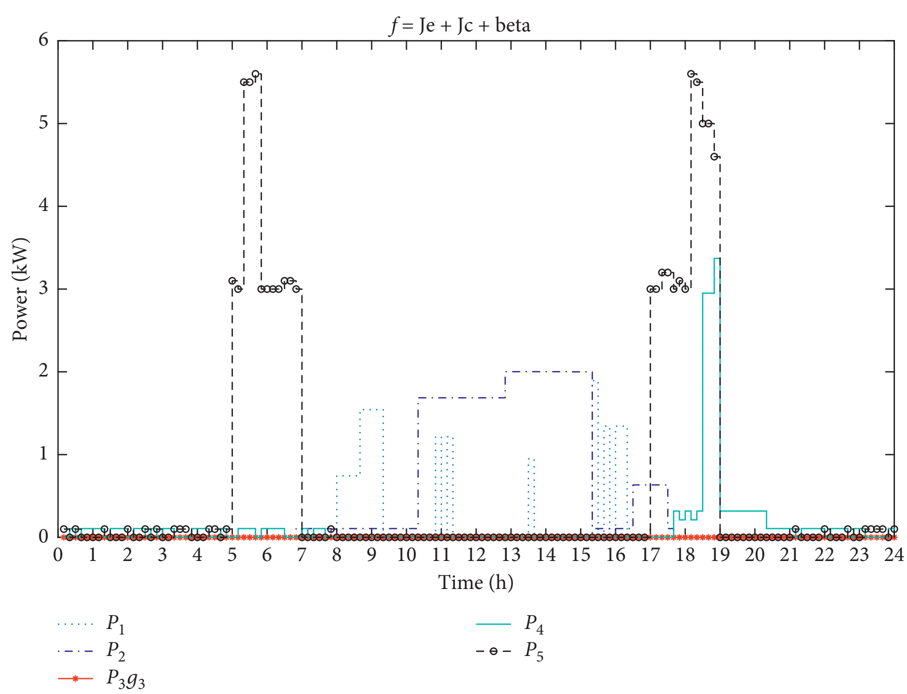


Figure 7: Power dispatch of the new design in case (iii).

From Figures 3, 5, and 7, overlaps between different power supplies can be observed especially during peak hours, while from Figures 4, 6, and 8, none of the time slots allows multiple power supplies. In these cases, the grid power outputs are smoothened under the dispatch with the proposed approach. From the comparative results, supplies become more intermittent because of the contradiction between the renewable penetration objective and the supply constraint. Allowing the combination of multiple power supplies reduces such an intermittent performance while, according to the energy and economy performances in case studies, making better use of the renewable energy sources.

As a conclusion, the proposed approach outperforms the previous model in all cases. When comparing the results of the new and previous designs, it appears that the flexible power dispatch allows more appliances to be scheduled to

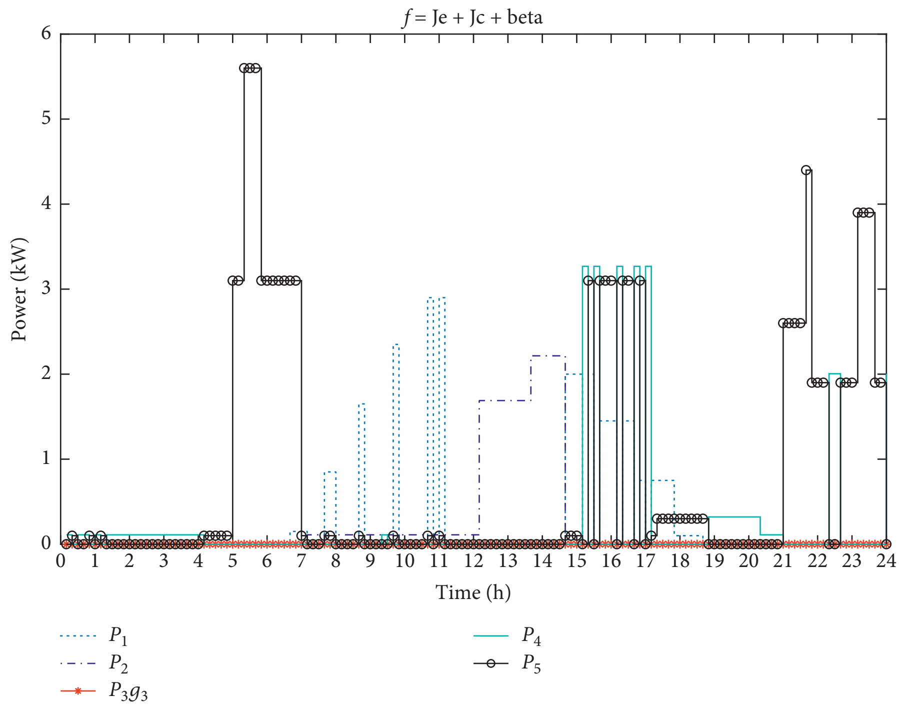


Figure 8: Power dispatch of the previous design in case (iii).

standard and off-peak hours, resulting in a lower energy cost. The objectives J_e and J_c manifest a certain level of trade-off. This is resulted from the employment of the TOU tariff, where some off-peak hours may be infeasible for the PV system because of its intermittent nature. When comparing the results among the three cases, it can be found that the first two cases can achieve lower energy cost and grid power consumption because of ignoring the inconvenience indicator β. It invokes an interesting topic for future studies that how to strike a balance between the conflicted interests of user satisfaction and energy efficiency in such a hybrid energy system management.

6. Conclusions

This paper investigates a technoeconomic optimization problem for a domestic grid-connected PV-battery hybrid energy system, via extending and improving a previously proposed system design. According to the previous design, the power dispatch is decided for the totality of the electrical loads. In the new model, appliances that comprise the electrical loads are supplied and managed, respectively, via additional power lines and switches from each power source. Furthermore, the appliance time scheduling is incorporated into such a flexible power dispatch. In this way, the system achieves better energy efficiency and economic performances via the technoeconomic optimization. The performances are evaluated by three optimization objectives: minimizing energy cost, maximizing renewable energy penetration, and increasing user satisfaction, over a finite horizon. There are nonlinear objective functions and constraints, as well as discrete and continuous decision variables, in such an optimization problem. As a result, the problem becomes an MINLP one at a large scale, which is difficult to solve. A competitive swarm optimizer-based numerical solver is thereby designed and employed.

In order to verify that the new design does improve the performances, a case study is investigated, where the power dispatch and appliance time scheduling on a daily basis are applied to a typical South African household hybrid system. There are three optimization cases, each with different objective functions, including only energy cost minimization, only renewable energy penetration maximization, and a weighted sum of the three objectives. Simulations are applied in these cases, where comparative results are also obtained via optimization of the previous system design. The same solver and system configurations are employed. In all cases, the results from the new design outperform results from the previous design. The improvement ranges from 7.1% to 13.3% and manifests that further energy efficiency and economic benefits can be achieved by the proposed approach. Furthermore, the solver generally takes around 18.4 minutes to obtain the solution. It verifies that the proposed approach has the potential for application in a real-time context.

There are several future works to investigate based on the current-stage results. Firstly, the performance evaluations, such as the battery wear cost and the renewable energy penetration, are simplified. More practical indicators can be introduced in the future. Secondly, uncertainties from the environment and user demands are inevitable in practice. The real-time feedback mechanism can be introduced to overcome such uncertainties. Thirdly, the game theory-based power dispatch and load scheduling considering conflicted interests call further study. Lastly, the CSO-based numerical solver can be further investigated to improve the algorithm performance.

Data Availability

All relevant data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Nature Science Foundation of China under Grant 61803162 and the Fundamental Research Funds for the Central Universities under Grant HUST: 2017KFYXJJ178.

References

[1] J. Soares, T. Pinto, F. Lezama, and H. Morais, "Survey on complex optimization and simulation for the new power systems paradigm," *Complexity*, vol. 2018, Article ID 2340628, 32 pages, 2018.

[2] B. Zhu, H. Tazvinga, and X. Xia, "Switched model predictive control for energy dispatching of a photovoltaic-diesel-battery hybrid power system," *IEEE Transactions on Control Systems Technology*, vol. 23, pp. 1229–1236, 2014.

[3] S. Shaahid and M. Elhadidy, "Economic analysis of hybrid photovoltaic–diesel–battery power systems for residential loads in hot regions—a step to clean future," *Renewable and Sustainable Energy Reviews*, vol. 12, pp. 488–503, 2008.

[4] P. Nema, R. Nema, and S. Rangnekar, "A current and future state of art development of hybrid energy system using wind and pv-solar: a review," *Renewable and Sustainable Energy Reviews*, vol. 13, pp. 2096–2103, 2009.

[5] X. Li, J. Lai, and R. Tang, "A hybrid constraints handling strategy for multiconstrained multiobjective optimization problem of microgrid economical/environmental dispatch," *Complexity*, vol. 2017, Article ID 6249432, 12 pages, 2017.

[6] T. Cheng, M. Chen, Y. Wang et al., "Adaptive robust method for dynamic economic emission dispatch incorporating renewable energy and energy storage," *Complexity*, vol. 2018, Article ID 2517987, 13 pages, 2018.

[7] X. Chen, B. Xu, and W. Du, "An improved particle swarm optimization with biogeography-based learning strategy for economic dispatch problems," *Complexity*, vol. 2018, Article ID 7289674, 15 pages, 2018.

[8] R. Kallel, G. Boukettaya, and L. Krichen, "Demand side management of household appliances in stand-alone hybrid photovoltaic system," *Renewable Energy*, vol. 81, pp. 123–135, 2015.

[9] S. Phiri and K. Kusakana, "Demand side management of a grid connected pv-wt-battery hybrid system," in *Proceedings of the 2016 International Conference on the Industrial and Commercial Use of Energy (ICUE)*, pp. 45–51, IEEE, Cape Town, South Africa, August 2016.

[10] H. Tazvinga, X. Xia, and J. Zhang, "Minimum cost solution of photovoltaic–diesel–battery hybrid power systems for remote consumers," *Solar Energy*, vol. 96, pp. 292–299, 2013.

[11] H. Tazvinga, B. Zhu, and X. Xia, "Energy dispatch strategy for a photovoltaic–wind–diesel–battery hybrid power system," *Solar Energy*, vol. 108, pp. 412–420, 2014.

[12] D. Setlhaolo and X. Xia, "Optimal scheduling of household appliances with a battery storage system and coordination," *Energy and Buildings*, vol. 94, pp. 61–70, 2015.

[13] S. M. Sichilalu and X. Xia, "Optimal energy control of grid tied pv–diesel–battery hybrid system powering heat pump water heater," *Solar Energy*, vol. 115, pp. 243–254, 2015.

[14] D. Setlhaolo and X. Xia, "Combined residential demand side management strategies with coordination and economic analysis," *International Journal of Electrical Power & Energy Systems*, vol. 79, pp. 150–160, 2016.

[15] F. Yang and X. Xia, "Techno-economic and environmental optimization of a household photovoltaic-battery hybrid power system within demand side management," *Renewable Energy*, vol. 108, pp. 132–143, 2017.

[16] C. Kagiri, E. M. Wanjiru, L. Zhang, and X. Xia, "Optimized response to electricity time-of-use tariff of a compressed natural gas fuelling station," *Applied Energy*, vol. 222, pp. 244–256, 2018.

[17] R. Cheng and Y. Jin, "A competitive swarm optimizer for large scale optimization," *IEEE Transactions on Cybernetics*, vol. 45, pp. 191–204, 2014.

[18] J. Kennedy, "Particle swarm optimization," in *Encyclopedia of machine learning*, C. Sammut and G. I. Webb, Eds., Springer, Boston, MA, USA, pp. 760–766, 2010.

[19] S. Leva, A. Dolara, F. Grimaccia, M. Mussetta, and E. Ogliari, "Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power," *Mathematics and Computers in Simulation*, vol. 131, pp. 88–100, 2017.

[20] H. Bindner, T. Cronin, P. Lundsager, J. F. Manwell, U. Abdulwahid, and I. Baring-Gould, *Lifetime Modelling of Lead Acid Batteries*, Forskningscenter Risoe, Roskilde, Denmark, 2005.

[21] C. Wei, M. Benosman, and T. Kim, "Online parameter identification for state of power prediction of lithium-ion batteries in electric vehicles using extremum seeking," *International Journal of Control, Automation and Systems*, 2019, In press.

[22] C. Yin, X. Huang, S. Dadras et al., "Design of optimal lighting control strategy based on multi-variable fractional-order extremum seeking method," *Information Sciences*, vol. 465, pp. 38–60, 2018.

[23] F. Cheng, L. Qu, W. Qiao, C. Wei, and L. Hao, "Fault diagnosis of wind turbine gearboxes based on DFIG stator current envelope analysis," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1044–1053, 2018.

[24] X. Xia and J. Zhang, "Operation efficiency optimisation modelling and application of model predictive control," *IEEE/CAA Journal of Automatica Sinica*, vol. 2, pp. 166–172, 2015.

[25] R. Cheng, "Nature inspired optimization of large problems," Ph.D. thesis, University of Surrey, Guildford, UK, 2016.

[26] Q. Zhao and L.-y. Zhang, "Research on the effect of Dpso in team selection optimization under the background of big data," *Complexity*, vol. 2018, Article ID 1386407, 14 pages, 2018.

[27] J. Liu, H. Ma, X. Ren, T. Shi, P. Li, and X. Ma, "The continuous-discrete pso algorithm for shape formation problem of multiple agents in two and three dimensional space," *Applied Soft Computing*, vol. 67, pp. 409–433, 2018.

[28] L. Wang and X.-l. Zheng, "A knowledge-guided multi-objective fruit fly optimization algorithm for the multi-skill resource constrained project scheduling problem," *Swarm and Evolutionary Computation*, vol. 38, pp. 54–63, 2018.

[29] H. Tazvinga and T. Hove, *Photovoltaic/Diesel/Battery Hybrid Power Supply System*, VDM Publishers, Pretoria, South Africa, 2010.

[30] Z. Wu and T. W. Chow, "Neighborhood field for cooperative optimization," *Soft Computing*, vol. 17, pp. 819–834, 2013.

[31] N. Ao, M. Zhao, Q. Li, S. Qu, and Z. Wu, "Network characteristics for neighborhood field algorithms," *Neural Computing and Applications*, 2019.

Control Engineering Practice 90 (2019) 85–100

Contents lists available at ScienceDirect

Control Engineering Practice

journal homepage: www.elsevier.com/locate/conengprac

Distributed control for a multi-evaporator air conditioning system

Jun Mei [a,b,*], Xiaohua Xia [b]

[a] School of Mathematics and Statistics, South-Central University for Nationalities, Wuhan 430074, China

[b] Centre of New Energy Systems, Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa

ARTICLE INFO

Keywords:
Multi-evaporator air conditioning system
Two-layer distributed controls
Model predictive control
Time-of-use
Energy/demand reducing

ABSTRACT

An autonomous hierarchical distributed control (AHDC) strategy is proposed for a building multi-evaporator air conditioning (ME A/C) system in this paper. The objectives are to minimize peak demand and energy costs, and to reduce communication resources, computational complexity and conservativeness while maintaining both thermal comfort and indoor air quality (IAQ) in acceptable ranges. The building consists of multiple connected rooms and zones. The proposed control strategy consists of two layers. The upper layer is an open loop optimizer, which only collects local measurement information and solves a distributed steady state resource allocation problem to autonomously and adaptively generate reference points, for low layer controllers. This is achieved by optimizing the demand and energy costs of a multi-zone building ME A/C system under a time-of-use (TOU) rate structure, while meeting the requirements of each zone's thermal comfort and IAQ within comfortable ranges. The lower layer also uses local information to track the trajectory references, which are calculated by the upper layer, via a distributed model predictive control (DMPC) algorithm. The control strategy is distributed at both layers because they use only local information from the working zone and its neighbors. Simulation results are provided to illustrate the advantages of the designed control schemes.

1. Introduction

It is well known that many environmental problems are linked to energy consumption. The energy consumed by the building sector accounts for 40% of the total energy consumption in the world (UNEP Sustainable Buildings & Climate Initiative, 2009). Among all building energy consumers, air conditioning (A/C) systems are responsible for the largest share, which represents close to 50% of the total electricity use in the building sector.

In recent years, many researchers have focused on reducing energy consumption of building heating, ventilation and air conditioning (HVAC) systems (Lee & Braun, 2008; Tang, Wang, Shan, & Cheung, 2018). Meanwhile, indoor comfort is also important for buildings, since it directly affects the occupants' working efficiency. The effective control of HVAC systems has the potential of reducing energy consumption or cost and improving indoor thermal comfort and air quality (IAQ). In Atthajariyakul and Leephakpreeda (2004), the authors proposed a method of real-time determination of an optimal indoor-air condition for the HVAC system to consider indoor thermal comfort and IAQ for occupants simultaneously with efficient energy consumption. However, this method is only tested around the desired points; we do not know if this method can be used without the desired points.

Model predictive control (MPC) has been verified as one of the most successful advanced control strategies, which is capable of improving energy efficiency and thermal comfort in buildings (Castilla, Álvarez, Normey-Rico, & Rodríguez, 2014; Cigler, Prívara, Váňa, Žáčeková, & Ferkl, 2012; Ma, Qin, & Salsbury, 2014; Maasoumy & Sangiovanni-Vincentelli, 2012; Mei & Xia, 2017b; Wallace et al., 2012). An energy-optimized open loop optimization and the MPC schemes were proposed (Mei & Xia, 2017a; Mei, Xia, & Song, 2018) for a direct expansion (DX) A/C system to improve energy efficiency while maintaining indoor thermal comfort and IAQ within comfort levels. Other advantages of MPC for building HVAC systems include robustness, tunability and flexibility (Oldewurtel et al., 2012). Despite MPC having superior performance to other control strategies, the size of the optimization problem increases rapidly when the dimension of the building A/C systems is large. Centralized MPC techniques were proposed (Hu & Karava, 2014; Maasoumy, Razmara, Shahbakhti, & Sangiovanni-Vincentelli, 2014; Mei & Xia, 2018; Razmara, Maasoumy, Shahbakhti, & Robinett III, 2015) for multi-zone HVAC systems to improve energy efficiency and thermal comfort. In the centralized control structure case, all the subsystems are controlled by one MPC law. The model used for prediction includes the coupling elements. When a centralized MPC algorithm is used for controlling HVAC systems in a large number of rooms, its algorithm is impractical since the optimization problems may not be solved in a reasonable time and the control systems are not easy to maintain. To reduce computational time, one of the effective predictive control strategies is a decentralized MPC approach (Elliott & Rasmussen, 2013). Large-scale control problems are decomposed into

* Corresponding author.
E-mail address: junmei027@gmail.com (J. Mei).

https://doi.org/10.1016/j.conengprac.2019.06.017
Received 3 January 2019; Received in revised form 8 May 2019; Accepted 22 June 2019
Available online 28 June 2019

Nomenclature

A_1	heat transfer area in the dry-cooling region of the DX evaporator, m^2
A_2	heat transfer area in the wet-cooling region of the DX evaporator, m^2
A_{win}	represents the total window area, m^2
C_a	specific heat of air, kJ kg^{-1} °C^{-1}
C_c	CO_2 concentration in the conditioning space, ppm
C_{load}	pollutant load, m^3/s
C_s	CO_2 concentration of air supply, ppm
d	cross-sectional area of zone, m^2
G	amount of CO_2 emission by a person, L/h
h_{fg}	latent heat of vaporization of water, kJ/kg
h_{r1}	enthalpy of refrigerant at evaporator inlet, kJ/kg
h_{r2}	enthalpy of refrigerant at evaporator outlet, kJ/kg
h_s	enthalpy leaving the DX evaporator, kJ/kg
k_P, k_I	proportional and integral coefficients
m_r	mass flow rate of refrigerant, kg/s
M_{load}	moisture load in the conditioned space, kg/s
$Occp$	number of occupants
Q_{load}	sensible heat load in the conditioned space, kW
Q_{rad}	solar radiative heat flux density, W/m^2
R	thermal resistance, °C/kW
T_d	air temperature leaving the dry-cooling region on air side of the DX evaporator, °C
T_{mix}	mixing temperature between the outside air and return air, °C
T_s	air temperature leaving the DX evaporator, °C
T_w	temperature of the DX evaporator wall, °C
T_z	air temperature in the conditioned space, °C
T_0	temperature of the outdoor air, °C
V	volume of the conditioned space, m^3
v_a	indoor air velocity, m/s
V_{h1}	air side volume in the dry-cooling region on air side of the DX evaporator, m^3
V_{h2}	air side volume in the wet-cooling region on air side of the DX evaporator, m^3
v_f	air volumetric flow rate, m^3/s
W_{mix}	mixing moisture content of outside air and return air, kg/kg
W_s	moisture content of air leaving the DX evaporator, kg/kg
W_z	moisture content of air-conditioned space, kg/kg
W_0	moisture content of the outdoor air, kg/kg

Greek letters

α_{dc}	heat transfer coefficient between air and the DX evaporator wall in the dry-cooling region, kW m^{-2} °C^{-1}
α_{wc}	heat transfer coefficient between air and the DX evaporator wall in the wet-cooling region, kW m^{-2} °C^{-1}
ε_{win}	transmissivity of glass of window
ρ	density of moist air, kg/m^3

Subscripts

i	room number

Abbreviations

AHDC	autonomous hierarchical distributed control
DMPC	distributed model predictive control
EEV	electronic expansion valve
HVAC	heating, ventilation and air conditioning
IAQ	indoor air quality
ME A/C	multi-evaporator air conditioning
MPC	Model predictive control
NLP	nonlinear programming
PMV	predicted mean vote
PSA	pressure swing absorption
TABS	thermally activated building systems
TOU	time-of-use

several independent control problems, which can take care of the local control parameters (Atam, 2016). However, the results demonstrated that the control performance loss was 28.58%. A distributed control approach is capable of balancing these issues. The structure of the distributed control is similar to a decentralized law, but is essentially a different approach (Zhang, Shi, Yan, Malkawi, & Li, 2017). The distributed control decomposes the centralized control to a group of local agents communicating with its neighbors, which makes it possible to be used for large-scale dynamically coupled systems. A communication network that allows collaboration among local control laws, which allows the improvement of global system performance compared to a decentralized structure. Moreover, computational demand should be significantly reduced compared to the centralized structure (Zheng, Li, & Qiu, 2013).

Owing to the advantages of distributed model predictive control (DMPC), this strategy was proposed to reduce the computational demand and handle the coupling among subsystems (Ma, Anderson, & Borrelli, 2011; Morosan, Bourdai, Dumur, & Buisson, 2011; Morosan, Bourdais, Dumur, & Buisson, 2010; Scherer et al., 2014). A DMPC was proposed in Ma et al. (2011) to improve the energy efficiency of the HVAC system while keeping zone temperature within the comfort range. In the study, the nonlinear optimal control problem is formulated and solved through sequential quadratic programming. Then the subproblem is decomposed further by adopting a subgradient approach. A local controller reaches the optimal solution by repeatedly negotiating with its neighbors in every sampling period, which inevitably increases the demand for calculation. In Morosan et al. (2010), the DMPC algorithm, only required the predicted output exchanged with its neighbors for every sampling period. However, this algorithm can only obtain Nash equilibrium, which may not be the optimal solution. In Morosan et al. (2011), the authors proposed a DMPC algorithm to control multi-source multi-zone temperatures. In order to attenuate the online computational burden, the DMPC algorithm was implemented based on Benders' decomposition. The results show that the computational and convergence times of this algorithm are superior to the centralized MPC. However, the energy efficiency of the DMPC method is not particularly good compared to the centralized MPC strategy. Furthermore, this type of distributed structure does not converge to the optimal solution, as in Scherer et al. (2014) which was an agent-based suboptimal controller; the drawback is transmitted to the decomposition algorithm.

In addition to improving energy efficiency while maintaining building multi-zones' thermal comfort within comfort range, DMPC strategies based on energy scheduling were proposed in Long, Liu, Xie, and Johansson (2016) and Radhakrishnan, Srinivasan, Su, and Poolla (2018). In Long et al. (2016), the authors proposed a method that combined the closed-loop centralized and distributed structures together to design a hierarchical control scheme to balance the computational complexity and conservativeness. In the study, the upper layer controller collects temperature and predictive information of all rooms and zones, which implies that the centralized scheduling (CS) needs to communicate with all rooms. The upper layer optimization problem is nonlinear, and solving it for a large building using centralized approaches is computationally cumbersome, leading to scalability issues. Furthermore, implementing centralized approaches requires transmission of zone-levels models and sensor information to the CS, leading to engineering difficulties and increasing information exchange. In the lower layer, the distributed controller only uses one room's information and its neighbor off-line reference signals. This may cause loss of control accuracy in receding horizon. Moreover, the trajectory references in the optimization objectives are given and fixed over a 24-h period, as in Ma et al. (2011), Morosan et al. (2010) and Scherer et al. (2014). Centralized and distributed MPC controllers following fixed trajectory references were also reported in other field (Zafra-Cabeza, Maestre, Ridao, Camacho, & Sánchez, 2011). In our previous work (Mei et al., 2018), the results demonstrated that the MPC strategy following preprogrammed time-varying reference points can save more in energy consumption and cost when compared with a fixed trajectory reference. More recently, in Radhakrishnan et al. (2018), the authors proposed adaptive learning and distributed control together to improve the energy efficiency and thermal comfort for multi-zone HVAC systems. The optimal references are preprogrammed and time-varying, while the presented zone thermal dynamics of a multizone building did not consider the interaction between rooms. Moreover, this distributed optimization algorithm is solved by using the subgradient method.

Advanced building structures are extremely complicated, with widely equipped multi-evaporator (ME) A/C systems. An ME A/C, which is DX based, consists of an outdoor compressor and condensing, and multiple indoor units including electronic expansion valves (EEVs) and evaporators (Xu, Yan, Deng, Xia, & Chan, 2013). Experimental results have illustrated that the control performance of the novel capacity control algorithm is further improved in comparison with its previous work. However, controlling indoor air temperature by using the novel capacity control algorithm could still be subject to significant fluctuations under certain operating conditions because of using a temperature dead-band, time-delay for compressor start-up. The interaction with other indoor units may be an important impact factor but was rarely considered. To improve the energy efficiency of a multi-zone building ME A/C system, thermal comfort and IAQ levels, a suitable optimization method is required for making each room's temperature, humidity and CO_2 concentration consistent with their desired references. To realize it, we consider a case that each DX unit can exchange information with its neighbors.

To overcome the above issues, in this paper we present an autonomous hierarchical distributed control (AHDC) method for a multi-zone building ME A/C system which not only considers how to maintain multiple zones' thermal comfort and IAQ within comfortable ranges but also considers reduction of communication resources, computational complexity and conservativeness reduction, and energy consumption and costs. Meanwhile, the peak-average-ratio (PAR) can also be considered in this paper. Moreover, the proposed comfort control considers thermal comfort and IAQ and the coupling effects of them. This control strategy consists of two layers. The upper layer is open loop scheduling that collects only a room's measurement information containing room cooling and pollutant loads, weather conditions, end-user services including demand and energy rates, thermal comfort and IAQ levels and operation profiles. Then the upper layer formulates and solves a steady-state optimization problem for minimizing the demand and energy costs of the multi-zone building ME A/C system under a time-of-use (TOU) rate structure of electricity over a 24-h period using nonlinear programming (NLP) algorithm. We make an assumption that the multi-zones are similar in the occupancies, functions and purposes; in this situation, one can distributively design an optimal scheduler. This scheduling generates time-varying trajectory references and communicates with the whole connected network through neighbors. All rooms then transmit their references to the lower layer controllers. The lower layer designed as DMPC controllers also uses local information to formulate and solve local optimization problems to track the autonomously and adaptively time-varying trajectory reference signals calculated by the upper layer. For simplicity, we make an assumption that all state variables are measured, thus full state feedbacks are considered. Our future work will consider designing observers in case some variables are not measured. The way we handle the upper layer is different from that of Long et al. (2016) and Zafra-Cabeza et al. (2011), which needs to collect all rooms' measurement information. It is also different from the distributed controllers in Ma et al. (2011), Morosan et al. (2011, 2010), Scherer et al. (2014) and Zheng et al. (2013), which collect information from a zone and its neighbors. The proposed control scheme can be realized with reduced, cheaper and short-range communication modules, and depending on the communication topology, a receiver only. While in the conventional control schemes (Long et al., 2016; Ma et al., 2011; Morosan et al., 2011, 2010; Scherer et al., 2014; Zafra-Cabeza et al., 2011; Zheng et al., 2013), it may require full-swing communication modules, i.e., with both a transmitter and receiver, which require external service providers in long-range data communication modules. The lower layer designs a new distributed controller for a zone such that this subsystem depends entirely on the zone by introducing a new input variable over a short-term horizon. This distributed control scheme is desirable in practice and can be easily implemented by our previous control algorithm (Mei et al., 2018). The results show that the proposed control scheme is superior to the previous control strategy on energy efficiency.

Our principal contributions can be summarized as follows:

(1) We first propose two-layer distributed control strategies that not only reduce more energy demand and costs in comparison with previous works but also maintain both thermal comfort and IAQ of multi-zone within comfortable ranges. These levels of performance are demonstrated in the case study.

(2) The proposed steady state distributed control and closed-loop distributed control schemes have the potential of reducing the complexity of computation and the hardware of communication modules in comparison with the centralized, non-distributed control schemes and hierarchical distributed control schemes.

(3) A novel approach for the lower layer closed-loop distributed control is designed to obtain a new feedback controller. This is achieved by introducing new input variables such that the closed-loop distributed control subsystems can be converted to a subsystem that depends entirely on one zone and our previous MPC algorithm developed for a single zone can be used.

(4) This study considers the predicted mean vote (PMV) index as an indicator of both thermal comfort and IAQ.

This paper is organized as follows: In Section 2, the nonlinear dynamical models and energy models for the multi-zone building ME A/C system, the PMV index and the system constraints are presented. The proposed AHDC method for the multi-zone building ME A/C system is proposed in Section 3. Simulation results are provided in Section 4. Section 5 concludes this paper.

2. System model

2.1. An ME A/C system in buildings

The schematic of an ME A/C system is illustrated in Fig. 1. The ME A/C system includes dampers, DX evaporators, an air-cooled tube-plate-finned condenser, a variable speed compressor, EEVs, variable

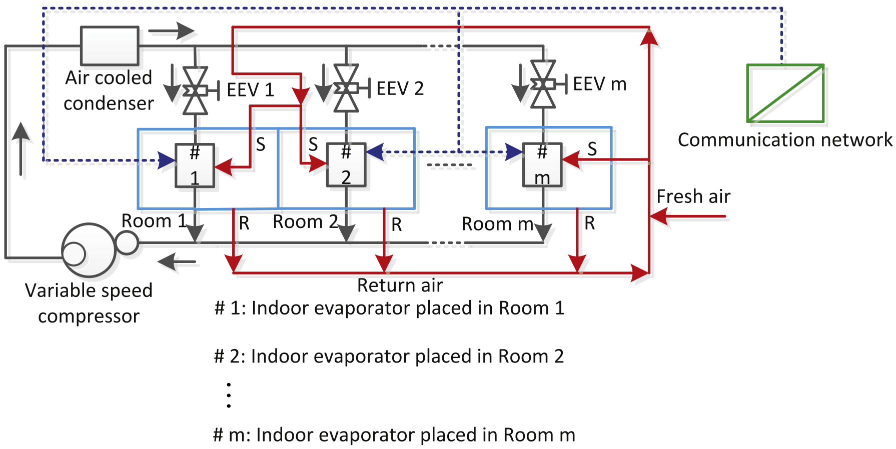


Fig. 1. Schematic diagram of an ME A/C system.

speed centrifugal supply fans with pressure swing absorption (PSA) boxes, and a damper for mixing return air from the ME A/C system with outside air. The variable speed supply fan adjusts its own speed based on the air flow rate/opening controlled by EEV to control cooled air to each room. Each indoor unit placed in the room has an EEV and an evaporator. The PSA box regulates the conditioned air flow rate and absorbs CO_2 contaminant concentration for improving the fresh air ratio. Each indoor unit is connected to the variable speed compressor and the outlet of the air-cooled condenser. The indoor air unit recirculates return air from building spaces and mixes it with outside air. The proportion of return air to outside air is controlled by damper positions in the ME A/C system. The mixed air is cooled by the cooling coils.

Because of the complex nature of air flow and the heat transfer process, ME A/C systems are usually modeled as time-varying nonlinear partial differential equations (Vakiloroaya, Ha, & Skibniewski, 2013), which are not suitable for control and optimization. Therefore, the following assumptions are made to simplify the modeling.

(1) The air in each room and outdoor environment is well mixed immediately so that the temperature, humidity and CO_2 concentration distributions are uniform.

(2) The heat capacity of air is constant.

2.2. Dynamic model of the ME A/C system

According to the above configuration, we use an undirected connected graph structure to represent the rooms and their dynamic couplings as described below. We associate the ith room with the ith node of the system. The mathematical dynamic models for the multi-zone building ME A/C system via the relationship between air enthalpy, temperature and the moisture content leaving the evaporator i of unit i as $h_{s,i} = C_a T_{s,i} + h_{fg} W_{s,i}$ are described as follows. In this paper, we only consider the interaction between rooms by sensible heat gain.

$$C_a \rho V_i \frac{\mathrm{d}T_{z,i}}{\mathrm{d}t} = \sum_{j=1}^{m} \frac{T_{z,j} - T_{z,i}}{R_{ij}} + \frac{T_0 - T_{z,i}}{R_i} + C_a \rho v_{f,i}(T_{s,i} - T_{z,i}) + Q_{load,i}, \tag{1a}$$

$$\rho V_i \frac{\mathrm{d}W_{z,i}}{\mathrm{d}t} = \rho v_{f,i}\left(\frac{h_{s,i} - C_a T_{s,i}}{h_{fg}} - W_{z,i}\right) + M_{load,i}, \tag{1b}$$

$$C_a \rho V_{h1,i} \frac{\mathrm{d}T_{d,i}}{\mathrm{d}t} = C_a \rho v_{f,i}(T_{mix} - T_{d,i}) + \alpha_{dc,i} A_{1,i}\left(T_{w,i} - \frac{T_{mix} + T_{d,i}}{2}\right), \tag{1c}$$

$$\begin{aligned}\rho V_{h2,i} \frac{\mathrm{d}h_{s,i}}{\mathrm{d}t} &= \alpha_{wc,i} A_{2,i}\left(T_{w,i} - \frac{T_{d,i} + T_{s,i}}{2}\right) + h_{fg} \rho v_{f,i}\left(W_{mix} - \frac{h_{s,i} - C_a T_{s,i}}{h_{fg}}\right) \\ &\quad + C_a \rho v_{f,i}(T_{d,i} - T_{s,i}),\end{aligned} \tag{1d}$$

$$\begin{aligned}C_{w,i} \rho_{w,i} V_{w,i} \frac{\mathrm{d}T_{w,i}}{\mathrm{d}t} &= \alpha_{dc,i} A_{1,i}\left(\frac{T_{mix} + T_{d,i}}{2} - T_{w,i}\right) \\ &\quad + \alpha_{wc,i} A_{2,i}\left(\frac{T_{d,i} + T_{s,i}}{2} - T_{w,i}\right) - \\ &\quad (h_{r2,i} - h_{r1,i}) m_{r,i},\end{aligned} \tag{1e}$$

$$V_i \frac{\mathrm{d}C_{c,i}}{\mathrm{d}t} = \left(k_P v_{f,i} + k_I \int_0^{T_I} v_{f,i} \mathrm{d}s\right)(C_{s,i} - C_{c,i}) + G_i \cdot Occp_i, \tag{1f}$$

where zone $i \in \{1, 2, \ldots, m\}$, $T_{z,i}$ and $W_{z,i}$ are the air temperature and moisture content of zone i, respectively; $T_{z,j}$ means the air temperature of neighboring zone i. $C_{c,i}$ denotes the CO_2 concentration of zone i, $C_{s,i}$ represents the CO_2 concentration of supply air to zone i. $T_{s,i}$ and $W_{s,i}$ are the air temperature and moisture content leaving the indoor unit i, respectively; T_0 and W_0 are the outside air temperature and moisture content, respectively. $T_{d,i}$ is the air temperature leaving the dry-cooling region on the air side of the DX evaporator of indoor unit i, $T_{w,i}$ is the temperature of the DX evaporator wall in indoor unit i, $h_{s,i}$ is the enthalpy leaving the DX evaporator of indoor unit i. $v_{f,i}$ is the air volumetric flow rate of the supply fan i, $m_{r,i}$ is the mass flow rate of refrigerant to the indoor unit i. $h_{r1,i}$ and $h_{r2,i}$ are the enthalpies of refrigerant at the DX evaporator inlet and outlet of indoor unit i, respectively. V_i is the volume of zone i; $V_{h1,i}$ and $V_{h2,i}$ are the air side volumes in the dry-cooling region and wet-cooling region on the air side of the DX evaporator of indoor unit i, respectively. $C_{w,i}$, $\rho_{w,i}$ and $V_{w,i}$ are the specific heat of air, density of moist air and volume of the DX evaporator wall of indoor unit i, respectively. $\alpha_{dc,i}$ and $\alpha_{wc,i}$ are the heat transfer coefficients between air and the evaporator wall in the dry-cooling region and wet-cooling region of indoor unit i, respectively. $A_{1,i}$ and $A_{2,i}$ are the heat transfer areas in the dry-cooling region and wet-cooling region on the DX evaporator of indoor unit i, respectively, which are time-varying uncertainty and bounded parameters. $Occp_i$ is the number of occupants of zone i, G_i is amount of CO_2 emission rate of people at zone i. k_P and k_I are the parameter of the PI controller.

$R_{ij} = R_{ji}$ is the thermal resistance of the wall between zone i and j, R_i is the thermal resistance of the wall between zone i and the outside. If R_{ij} and R_i are not known from design specifications, they can be obtained via model identification (Bacher & Madsen, 2011; Jiménez, Madsen, & Andersen, 2008). T_{mix} and W_{mix} are the mixed air temperature and mixed moisture content before each DX evaporator cooling coil, respectively. The mixed air temperature and moisture content are calculated as follows:

$$T_{mix} = (1-\delta)T_0 + \delta \frac{\sum_{i=1}^{m} \upsilon_{f,i} T_{z,i}}{\sum_{i=1}^{m} \upsilon_{f,i}}, \; W_{mix} = (1-\delta)W_0 + \delta \frac{\sum_{i=1}^{m} \upsilon_{f,i} W_{z,i}}{\sum_{i=1}^{m} \upsilon_{f,i}}, \tag{2}$$

where δ is the mixing ratio between the outside air and return air. It is assumed that the return air temperature and moisture content are the weighted sums of the zone temperatures and moisture contents with weights, being the air flow rate of supply air to the corresponding zones. The return air is not recirculated when $\delta = 0$, and no outside fresh air is used when $\delta = 1$. δ can be employed to save energy through recirculation but it has to be less than one to guarantee minimal outdoor fresh air delivered to the rooms. Note that the first equation of (2) is taken from Ma, Matuško, and Borrelli (2015). It is assumed that the mixed moisture content has a similar description in the second equation of (2). The airside convective heat transfer coefficients for the louvre-finned evaporator under both dry-cooling and wet-cooling regions on the air side of the evaporator i are calculated as follows (Chen & Deng, 2006):

$$\alpha_{dc,i} = j_{dc} \rho \upsilon_{a,i} \frac{C_a}{Pr^{\frac{2}{3}}}, \; \alpha_{wc,i} = j_{wc} \rho \upsilon_{a,i} \frac{C_a}{Pr^{\frac{2}{3}}}, \; i = 1, 2, \ldots, m, \tag{3}$$

where Pr is the Prandtl number, j_{dc} and j_{wc} are the Colburn factors in the cooling mode. The air velocity $\upsilon_{a,i}$ is described as follows:

$$\upsilon_{a,i} = \frac{\upsilon_{f,i} - \varepsilon_i}{d_i}, \; i = 1, 2, \ldots, m,$$

where d_i (m^2) is the cross-sectional area of zone i, ε_i is the non-desired air velocity through the door or window to pass in and out of the air to zone i, $\upsilon_{a,i}$ is the indoor air velocity of room i.

The above models (1a)–(1e) without considering outside air temperature and humidity entering into system for a single room were first built in Qi and Deng (2008). The above models (1a)–(1f), absorbing CO_2 by an independent PSA box for a single room, were built in Mei and Xia (2017a). The above models (1a)–(1f) for a single room, absorbing CO_2 by using a PI controller based on a supply fan, were built in Mei et al. (2018). On the right-hand side of (1a), the first term denotes the heat transfer between zone i and all neighbors of zone i; the second term means the heat transfer between zone i and the outside wall. The PI controller in Eq. (1f) is designed based on the air volumetric flow rate of the supply fan. It can be used for controlling the indoor CO2 concentration. In addition, the PI controller has the potential of reducing the complexity of computation and the cost of hardware.

Remark 1. Higher-order resistance–capacitance (RC) models were developed in Maasoumy et al. (2014) and Razmara et al. (2015). For simplicity, we only consider the first-order RC model in this paper. Though the higher-order RC models maybe more accurate than the first-order model, it is more difficult to use the current methods to solve the distributed control problem. Most existing works to solve the distributed control problem assume that interaction terms are either disturbances or negligible. We will study the distributed control problem of the higher-order RC models in the future.

Remark 2. The building DX A/C system's cooling and pollutant loads can be expressed in Mei et al. (2018) and used as measurement information for an open loop controller in the upper optimization. The building loads are affected by some parameters (such as T_0, W_0, $Q_{rad,i}$, $Occp_i$, internal heat gain $Q_{int,i}$ and moisture ventilation load $M_{int,i}$). The prediction of these parameters can be obtained through a weather forecast station, historical data and schedules. Though the multi-zone buildings' cooling and pollutant loads cannot be accurately predicted, the designed AHDC strategy in the next section includes the DMPC controllers that are capable of handling the prediction errors.

To make the ME A/C system cooperatively control multi-zones' thermal comfort and air quality, we suppose that the ME A/C system is equipped with a communication network based on wireless communication. In this network, they can share information (e.g., $T_{z,i}$, $W_{z,i}$ and $C_{c,i}$) with one another, which is shown in Fig. 1. The information flow between them is modeled as a network graph $\mathcal{G} = (\mathcal{V}, \vartheta, \mathcal{A})$, where $\mathcal{V} = \{1, 2, \ldots, m\}$ is the index set of different rooms and zones of the ME A/C system, $\vartheta \subset \mathcal{V} \times \mathcal{V}$ is the edge set of ordered pairs of the ME A/C system, and $\mathcal{A} = [a_{ij}] \in \mathbb{R}^{m \times m}$ is the adjacency matrix with entries $a_{ij} = 1$ or $a_{ij} = 0$. If the ME A/C subsystem i can receive information from the ME A/C subsystem j, then $(j, i) \in \vartheta$, $a_{ij} = 1$ and the ME A/C subsystem j is called the network neighbor of the ME A/C subsystem i, denoted by $j \in \mathcal{V}_i$, where $\mathcal{V}_i = \{j \in \mathcal{V} | a_{ij} = 1\}$. If the ME A/C subsystem i cannot have access to the information of the ME A/C subsystem j, then $(j, i) \notin \vartheta$, $a_{ij} = 0$ and $j \notin \mathcal{V}_i$. Self-connection is not considered for $\mathcal{G}$, i.e., $a_{ii} = 0$, $\forall i \in \mathcal{V}$. A graph $\mathcal{G}$ is undirected if $a_{ij} = a_{ji}$ for any $i, j \in \mathcal{V}$. In this paper, the network graph $\mathcal{G} = (\mathcal{V}, \vartheta, \mathcal{A})$ of the ME A/C system is assumed to be undirected and connected (Yu & Xia, 2017).

All the DX units of the ME A/C system adjust their comfort levels adaptively by acquiring the adjacent information. The neighbors of each DX unit can be defined in many different ways. In this paper, the following way is based on the effect of thermal resistance and is defined as follows:

$$\mathcal{V}_i = \{j : |R_{ij}| < \varepsilon_0, i \neq j\}, \tag{4}$$

where the parameter ε_0 is a predefined threshold, $\mathcal{V}_i$ is the set of neighbors of room i.

The system dynamic equations (1) can be written as equations of the following:

$$\dot{x}_i = f_i(x_i, x_{-i}, u_i, \omega_i), \; i = 1, 2, \ldots, m, \tag{5}$$

where the vector denoted as $x_i \triangleq [h_{s,i}, T_{z,i}, T_{d,i}, T_{w,i}, W_{z,i}, C_{c,i}]^T$ is the state of the subsystem S_i; $u_i = [\upsilon_{f,i}, m_{r,i}]^T$ are the constrained control signals; $\omega_i \triangleq [Q_{load,i}, M_{load,i}, C_{load,i}]^T$ represent the load variables of room i; and x_{-i} concatenate the states of all subsystems S_j ($j \in \mathcal{V}$) of the subsystem S_i, i.e., $x_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_m)$. The functions $f_i(x_i, x_{-i}, u_i, \omega_i)$ ($i = 1, 2, \ldots, m$) are defined as follows:

$$f_i(x_i, x_{-i}, u_i, \omega_i) = \begin{bmatrix} \frac{\alpha_{wc,i} A_{2,i}(T_{w,i} - \frac{T_{d,i}+T_{s,i}}{2}) + h_{fg} \rho \upsilon_{f,i}(W_{mix} - \frac{h_{s,i} - C_a T_{s,i}}{h_{fg}}) + C_a \rho \upsilon_{f,i}(T_{d,i} - T_{s,i})}{\rho V_{h2,i}} \\ \frac{\sum_{j=1}^{m} \frac{T_{z,j} - T_{z,i}}{R_{ij}} + \frac{T_0 - T_{z,i}}{R_i} + C_a \rho \upsilon_{f,i}(T_{s,i} - T_{z,i}) + Q_{load,i}}{C_a \rho V_i} \\ \frac{C_a \rho \upsilon_{f,i}(T_{mix} - T_{d,i}) + \alpha_{dc,i} A_{1,i}(T_{w,i} - \frac{T_{mix}+T_{d,i}}{2})}{C_a \rho V_{h1,i}} \\ \frac{\alpha_{dc,i} A_{1,i}(\frac{T_{mix}+T_{d,i}}{2} - T_{w,i}) + \alpha_{wc,i} A_{2,i}(\frac{T_{d,i}+T_{s,i}}{2} - T_{w,i}) - (h_{r2,i} - h_{r1,i}) m_{r,i}}{C_{w,i} \rho_{w,i} V_{w,i}} \\ \frac{\rho \upsilon_{f,i}(\frac{h_{s,i} - C_a T_{s,i}}{h_{fg}} - W_{z,i}) + M_{load,i}}{\rho V_i} \\ \frac{(k_P \upsilon_{f,i} + k_I \int_0^{T_I} \upsilon_{f,i} \mathrm{d}s)(C_{s,i} - C_{c,i}) + G_i \cdot Occp_i}{V_i} \end{bmatrix}. \tag{6}$$

2.3. Simplified energy models of the ME A/C system

The power consumers of the multi-zone building ME A/C system include the dampers, condenser fan, compressor and DX cooling coils. The power to drive the dampers is assumed to be negligible. The condenser fan power P_{con} is approximated as a second-order polynomial function of the total mass flow rate of refrigerant ($m_r = \sum_{i=1}^{m} m_{r,i}$) driven by the fan

$$P_{con} = c_0 + c_1 m_r + c_2 m_r^2, \tag{7}$$

where c_0, c_1 and c_2 are the parameters to be identified by curve-fitting of experimental data in Vakiloroaya, Samali, and Pishghadam (2014).

The power consumption of the evaporator fans P_{eva} based on the energy conservation law is expressed as follows:

$$P_{eva} = \sum_{i=1}^{m}(a_0 + a_1 v_{f,i}, + a_2 v_{f,i}^2 + a_3 T_{s,i} + a_4 T_{s,i}^2 + a_5 Q_{c,i} + a_6 Q_{c,i}^2 + a_7 v_{f,i} T_{s,i} + a_8 v_{f,i} Q_{c,i} + a_9 T_{s,i} Q_{c,i}), \tag{8}$$

where the coefficients a_i $(i = 0, 1, \dots, 9)$ are constant and can be determined by curve-fitting of experimental data in Vakiloroaya et al. (2014). $Q_{c,i}$ is the summation of the sensible and latent heat loads in room i.

The power consumption of the compressor P_{comp} is determined by Wallace et al. (2012):

$$P_{comp} = \sum_{i=1}^{m} \frac{m_{r,i}(h_{r2,i} - h_{r1,i})}{\overline{\eta}}, \tag{9}$$

where $\overline{\eta}$ is the combined total efficiency of the compressor (known parameters).

The total electric power consumption P_{tot} of the multi-zone building ME A/C system at time t then is calculated as

$$P_{tot} = P_{con} + P_{eva} + P_{comp}. \tag{10}$$

2.4. PMV index

The PMV index was proposed by Fanger (1972) and is used as a thermal comfort indicator. Fanger's index quantifies thermal sensation experienced by numerous people. The sensation is represented by a scale ranging from −3 (cold) to +3 (hot). The PMV index can be determined by personal and environmental factors. The personal factors consist of metabolic rate M_r (W/m^2) and clothing insulating I_{cl} (m^2°C/W). The environmental factors comprise air temperature T_z, air humidity (or moisture content) W_z, air velocity v_a and mean radiant temperature T_r. The function of the conventional PMV index for a single zone is depicted by

$$PMV = g(T_z, W_z, v_a, M_r, I_{cl}, T_r), \tag{11}$$

where the specific expression can be described in Fanger (1972).

Conventionally, the PMV index is an indicator of indoor air temperature and humidity (Castilla et al., 2011, 2014; Cigler et al., 2012; Freire, Oliveira, & Mendes, 2008). The CO_2 concentration, air temperature and humidity have become three major indicators of thermal comfort and IAQ. The separate control of the PMV index and CO_2 concentration was studied in Atthajariyakul and Leephakpreeda (2004) and Wang and Jin (2000). However, three coupling effects of indoor air temperature, humidity and CO_2 concentration cannot be ignored in many cases. In fact, indoor humidity was correlated with CO_2 concentration according to measurement results reported in Gladyszewska-Fiedoruk (2013). Furthermore, the experimental investigation (Lin, Chiu, & Chen, 2015) suggested that the value of the PMV index was affected by control of the indoor CO_2 concentration. To our best knowledge, very little work exists in the literature that proposes mathematical equations among the indoor air temperature, relative humidity and CO_2 concentration. We propose simplified mathematical equations such that the PMV index includes indoor thermal comfort and CO_2 concentration in this study.

M_r is the rate of metabolism, which denotes the amount of energy used by a person per unit of time. From the study of Weir (1949), the metabolic rate is directly related to a person's energy output, which can be expressed by calorie output per hour and a body's surface area

$$M_r = K_p / S_p, \tag{12}$$

where K_p denoting a person's heat output per hour is the calorie of 1L of oxygen consumed, S_p is the body surface area and can be expressed as (Weir, 1949)

$$S_p = 0.007184 H^{0.725} W^{0.425}, \tag{13}$$

where H and W are the height (cm) and weight (kg) of a person, respectively. For 1L oxygen consumed, we have (Weir, 1949)

$$\begin{cases} 1L\ \mathrm{O_2}\ \text{consumed} = a + b + c = 1, \\ 1L\ \mathrm{CO_2}\ \text{produced} = R = a + 0.802b + 0.718c, \\ K_p = 5.047a + 4.463b + 4.735c, \end{cases} \tag{14}$$

where a is the carbohydrate, b denotes the protein and c represents the fat which is obtained by 1L of oxygen metabolizing. The third equation of (14) can be reduced to the following one

$$K_p = 3.9 \times \mathrm{L\ O_2}\ \text{used} + 1.1 \times \mathrm{L\ CO_2}\ \text{produced} = 3.9 * V_o + 1.1 * G, \tag{15}$$

where V_o is the amount of oxygen consumed per unit of hour (l/h). This equation was widely cited and can be used for estimating the energy expenditure, oxygen consumed and CO_2 produced (Christensen, Frey, Foenstelien, Aadland, & Refsum, 1983; Kinney, Morgan, Domingues, & Gildner, 1964; Treuth, Adolph, & Butte, 1998).

Under normal conditions, when a body is at rest and in nutritional equilibrium, the global respiratory ratio is $m_{CO_2}/m_{O_2} = 0.83$ as reported in Djongyang, Tchinda, and Njomo (2010). Since this study investigates thermal comfort and IAQ of offices, we assume $G/V_o = 0.83$. One can then obtain

$$K_p = \frac{481.3}{83} * G, \tag{16}$$

According to (12), one can obtain that the metabolic rate in human metabolism of room i denoted by M_{r_i} has the following equation

$$M_{r_i} = \frac{481.3}{83 S_{p_i}} * G_i,\ i = 1, 2, \dots, m. \tag{17}$$

where S_{p_i} is the body surface area of room i.

Based on the Eqs. (17) and (1f), M_{r_i} under a steady state of the CO_2 concentration in room i can be expressed by

$$M_{r_i} = \frac{481.3}{83 S_{p_i} \cdot Occp_i}(k_P v_{f,i} + k_I \int_0^{T_I} v_{f,i} \mathrm{d}s)(C_{c,i} - C_{s,i}),\ i = 1, 2, \dots, m. \tag{18}$$

This equation implies that the metabolic rate can reflect indoor CO_2 concentration produced.

Then the PMV_i index is the function of the following variables:

$$PMV_i = g_i(T_{z,i}, W_{z,i}, C_{c,i}, v_{f,i}, I_{cl}, T_r),\ i = 1, 2, \dots, m. \tag{19}$$

It can be noted from this equation that the PMV index can be used as an indicator of thermal comfort and IAQ of room i.

The Eq. (18) represents a condition under which the steady states of the system are reached, i.e., there is a relationship between the metabolic rate, CO_2 concentration and air volumetric flow rate at steady states. In fact, for the same activity of a person, his respiratory change is determined by the indoor air temperature or/and humidity. A person's metabolic rate is directly reflected by a respiratory change. The high or low temperature or/and humidity can cause the occupant to breathe out either more or less CO_2, thus the indoor air temperature or/and humidity can influence the metabolic rate. On the other hand, the air volumetric flow rate determines the indoor air temperature and humidity and their eventual steady states. Therefore, the air volumetric flow rate is indirectly related to the metabolic rate.

Remark 3. Most previous works used the PMV index as a thermal comfort indicator. From function (19), it can be seen that the modified PMV index has been extended and used as an indicator of both thermal comfort and IAQ in the normal office buildings.

2.5. Constraints

The multi-zone building ME A/C system is subject to thermal comfort and IAQ constraints, and cooling operational constraints are defined as below.

(C1) $\mathrm{PMV}_i \in [\underline{\mathrm{PMV}}_i, \overline{\mathrm{PMV}}_i]$, $i = 1, 2, \ldots, m$. Each room's thermal comfort and IAQ are within the comfort ranges.

(C2) $\delta \in [\underline{\delta}, \overline{\delta})$. The upper and lower bounds limit the ratio of the outside air entering the system.

(C3) $T_{z,i} \in [\underline{T}_{z,i}, \overline{T}_{z,i}]$, $W_{z,i} \in [\underline{W}_{z,i}, \overline{W}_{z,i}]$, $C_{c,i} \in [\underline{C}_{c,i}, \overline{C}_{c,i}]$, $i = 1, 2, \ldots, m$. Each room's air temperature, moisture content and CO_2 concentration are within the required ranges for occupants in the cooling mode.

(C4) $T_{s,i} \in [\underline{T}_{s,i}, \overline{T}_{s,i}]$, $W_{s,i} \in [\underline{W}_{s,i}, \overline{W}_{s,i}]$, $i = 1, 2, \ldots, m$. The bounds of the air temperature and moisture leaving the DX evaporator are limited because of the physical characteristics of the coils and the air cooling coils of the DX evaporators. Besides, the upper bounds $\overline{T}_{s,i}$ and $\overline{W}_{s,i}$ are less than $T_{z,i}$ and $W_{z,i}$ respectively since they are used for cooling and dehumidifying of each room. The bound of the air enthalpy $h_{s,i}$ satisfies: $h_{s,i} \in [C_z \underline{T}_{s,i} + h_{fg} \underline{W}_{s,i}, C_a \overline{T}_{s,i} + h_{fg} \overline{W}_{s,i}]$.

(C5) $\sum_{i=1}^{m} v_{f,i} T_{s,i} \le \sum_{i=1}^{m} v_{f,i} T_{mix}$, $\sum_{i=1}^{m} v_{f,i} W_{s,i} \le \sum_{i=1}^{m} v_{f,i} W_{mix}$. The mixed air temperature and moisture content after each DX evaporator can only decrease.

(C6) $T_{d,i} \le T_{mix}$, $T_{w,i} \le T_{d,i}$, $W_{s,i} \le W_{mix}$, $i = 1, 2, \ldots, m$. Air temperature and moisture content after each DX dry-cooling and wet-cooling regions can only decrease, respectively.

(C7) $v_{f,i} \in [\underline{v}_{f,i}, \overline{v}_{f,i}]$, $m_{r,i} \in [\underline{m}_{r,i}, \overline{m}_{r,i}]$, $i = 1, 2, \ldots, m$. The upper bounds of the air volumetric flow rate $\overline{v}_{f,i}$ and the mass flow rate of refrigerant $\overline{m}_{r,i}$ of each room are limited by the physical characteristics of the multi-zone building ME A/C system. The lower bounds $\underline{v}_{f,i} > 0$ and $\underline{m}_{r,i} > 0$ are matched minimum operation and ventilation demands.

The constraints in (C1)–(C7) are compactly written as

$$x_i \in \mathbb{X},\ u_i \in \mathbb{U},\ h_{1,i}(x_i, u_i) \le 0,\ h_{2,i}(x_i) \le 0 \text{ and } PMV_i \in \mathbb{F}, \quad i = 1, 2, \ldots, m. \tag{20}$$

where $\mathbb{X}$, $\mathbb{U}$, $\mathbb{P}$ and $\mathbb{F}$ are bounded sets, $h_{1,i}(x_i, u_i)$ and $h_{2,i}(x_i)$ can be written as functions of the state and input variables, where they correspond to constraints in (C5) and (C6).

3. Controller design

To facilitate the description of the proposed AHDC strategy for the nonlinear systems (5), the notation ^{u} will be used for the upper layer control and ^{l} will be used for the lower layer DMPC. We will abbreviate the upper layer open loop controller to UOPC while the lower layer DMPC as LDMPC for short. t_k^u denotes the sampling time instant of the UOPC and t_k^l represents that of the lower level DMPC; assume $c(k,q) \triangleq kM + q$, where M is a positive integer number corresponding to the number of sampling instants of the LDMPC between two sampling instants of the UOPC; $t_k^u \triangleq t_{c(k,0)}^l$; $\delta^u \triangleq t_{k+1}^u - t_k^u$ and $\delta^l \triangleq t_{k+1}^l - t_k^l$ denote the sampling period of the UOPC and LDMPC, respectively; $\delta^u = M\delta^l$. T^l denotes the prediction horizon of the LDMPC, which satisfies $\delta^u \ge T^l$.

Throughout the rest of this paper, we denote the long-term scale horizon as $[0, K^u]$, and $K^u = n\delta^u$ ($n \in N^+$). Fig. 2 shows the time index of the two layers and that the upper layer sends information to the lower layer.

3.1. Upper level: steady state optimization problem

In reality, each zone has desired air temperature, humidity and CO_2 concentration, the reference points of which are determined by users. The objective of the upper layer considered in this paper is to minimize the total electricity bills in the building, which reflect demand and energy costs under the TOU rate structure, and to generate optimal reference points of air temperature, humidity and CO_2 concentration for each zone for the lower layer. More specifically, we consider the following centralized steady-state optimization problem:

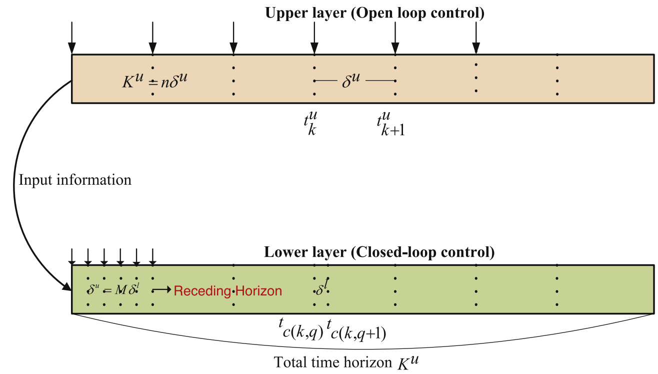


Fig. 2. Simplified schematic of two-layer time index.

$$X^*(t_k^u) = \arg\min_{x(t_k^u), u(t_k^u)} \Bigg(\underbrace{\sum_{i=1}^{m} \Bigg[w_1 \sum_{k=1}^{n} \big(E_c(t_k^u) P_{tot,i}(t_k^u) \delta^u \big)}_{\text{energy cost}} \tag{21a}$$

$$+ \underbrace{w_2 \big(D_c(t_k^u) \max_{1 \le k \le n} \{ P_{tot,i}(t_k^u) \} \big) \Bigg]}_{\text{demand cost}} \Bigg),$$

subject to the following constraints:

$$f_i(x_i(t_k^u), x_{-i}(t_k^u), u_i(t_k^u), \omega_i(t_k^u)) = 0,\ i = 1, 2, \ldots, m, \tag{21b}$$

$$|PMV_i(t_k^u)| \le \alpha,\ i = 1, 2, \ldots, m, \tag{21c}$$

$$x_i(t_k^u) \in \mathbb{X}_i,\ u_i(t_k^u) \in \mathbb{U}_i,\ h_{1,i}(x_i(t_k^u), u_i(t_k^u)) \le 0,\ h_{2,i}(x_i(t_k^u)) \le 0, \quad i = 1, 2, \ldots, m, \tag{21d}$$

where $t_k^u \in [0, K^u]$, $x(t_k^u) = [x_1(t_k^u), \ldots, x_m(t_k^u)]^T$ is the system state, $u(t_k^u) = [u_1(t_k^u), \ldots, u_m(t_k^u)]^T$ is the control input. The total energy consumption P_{tot} is expressed in (10) and the PMV function is described in (19). Constant α is the comfort bounded of the value of the PMV index. $E_c(t_k^u)$ is the TOU electricity rate at time step t_k^u, and $D_c(t_k^u)$ is the demand charge rate at time step t_k^u. w_i $(i = 1, 2)$ denote the positive weighting factors and $f_i(x_i(t_k^u), x_{-i}(t_k^u), u_i(t_k^u), \omega_i(t_k^u))$ are defined in (6). $X^*(t_k^u)$ is a global optimal solution of the optimization problem (21).

Before investigating the distributed steady state optimization problem, we make an assumption on the system model.

Assumption 1. The optimal problem (21) admits a solution, of which the steady state of temperature, humidity and CO_2 concentration for each zone are approximately the same.

This assumption is valid in many practical situations where the different zones serve the same functions and purposes; for example in an office environment, the comfort requirements are subject to the same standards, ambient conditions and energy regulatory and pricing structure and are therefore normally the same.

This assumption may not hold in cases where buildings have different functional zones such as offices and ancillary equipment spaces. The similar algorithms can be extended to the cases when different functional zones can be grouped into homogeneous ones.

Secondly, under a steady state, the total heat gain from neighboring zones is sometimes less dominant compared with that from the outside plus the indoor heat gain in every zone. As reported in Mei et al. (2018), the TOU rate structure is also the main factor to dominate the steady state optimization solutions. Therefore, in the optimization problem

(21), we can ignore the interacting terms $\Sigma_{j=1,j\neq i}^{m}\frac{T_{z,j}-T_{z,i}}{R_{ij}}$ in (1a) or $\Sigma_{j=1,j\neq i}^{m}\frac{T_{z,j}-T_{z,i}}{R_{ij}}$ in (21b). Thereby, a simplified optimization problem (22) is considered for one zone i only as follows:

$$X_i^r(t_k^u) = \arg\min_{x_i(t_k^u),u_i(t_k^u)} \Big(w_1 \underbrace{\sum_{k=1}^{n}\big(E_c(t_k^u)P_{tot,i}(t_k^u)\delta^u\big)}_{\text{energy cost}} \tag{22a}$$

$$+ \underbrace{w_2\big(D_c(t_k^u)\max_{1\le k\le n}\{P_{tot,i}(t_k^u)\}\big)}_{\text{demand cost}} \Big),$$

subject to the following constraints:

$$\widetilde{f}_i(x_i(t_k^u), u_i(t_k^u), \omega_i(t_k^u)) = 0, \tag{22b}$$

$$|PMV_i(t_k^u)| \le \alpha, \tag{22c}$$

$$x_i(t_k^u) \in \mathbb{X}_i,\ \ u_i(t_k^u) \in \mathbb{U}_i,\ \ h_{1,i}(x_i(t_k^u), u_i(t_k^u)) \le 0,\ \ h_{2,i}(x_i(t_k^u)) \le 0, \tag{22d}$$

where $t_k^u \in [0, K^u]$, $X_i^r(t_k^u)$ is a local optimal solution, and i means that the optimization problem (22) only needs the measurement information of room i. Here, $\widetilde{f}_i(x_i, u_i, \omega_i)$ is described by

$$\widetilde{f}_i(x_i, u_i, \omega_i) = \begin{bmatrix} \dfrac{\alpha_{wc,i}A_{2,i}(T_{w,i}-\frac{T_{d,i}+T_{s,i}}{2})+h_{fg}\rho v_{f,i}(W_{mix}-\frac{h_{s,i}-C_aT_{s,i}}{h_{fg}})+C_a\rho v_{f,i}(T_{d,i}-T_{s,i})}{\rho V_{h2,i}} \\ \dfrac{\frac{T_0-T_{z,i}}{R_i}+C_a\rho v_{f,i}(T_{s,i}-T_{z,i})+Q_{load,i}}{C_a\rho V_i} \\ \dfrac{C_a\rho v_{f,i}(T_{mix}-T_{d,i})+\alpha_{dc,i}A_{1,i}(T_{w,i}-\frac{T_{mix}+T_{d,i}}{2})}{C_a\rho V_{h1,i}} \\ \dfrac{\alpha_{dc,i}A_{1,i}(\frac{T_{mix}+T_{d,i}}{2}-T_{w,i})+\alpha_{wc,i}A_{2,i}(\frac{T_{d,i}+T_{s,i}}{2}-T_{w,i})-(h_{r2,i}-h_{r1,i})m_{r,i}}{C_{w,i}\rho_{w,i}V_{w,i}} \\ \dfrac{\rho v_{f,i}(\frac{h_{s,i}-C_aT_{s,i}}{h_{fg}}-W_{z,i})+M_{load,i}}{\rho V_i} \\ \dfrac{(k_P v_{f,i}+k_I\int_0^{T_I} v_{f,i}\mathrm{d}s)(C_{s,i}-C_{c,i})+G_i\cdot Occp_i}{V_i} \end{bmatrix}. \tag{23}$$

We have five important remarks for the optimization problem (22).

- In (22a), the term regarding the end-user services contains two parts, i.e., the energy cost of the multi-zone building ME A/C system given by $\sum_{k=1}^{n}\left[E_c(t_k^u)P_{tot,i}(t_k^u)\delta^u\right]$ (weighted by w_1) aims to minimize energy cost, the peak demand $D_c(t_k^u)\max_{1\le k\le n}\{P_{tot,i}(t_k^u)\}$ (weighted by w_2) aims to reduce demand cost.
- The weighting factors w_1 and w_2, which are determined by users, are to balance the two objectives. Specifically, if preferring more demand reduction, they can increase w_2 and decrease w_1 and vice versa.
- It can be seen in (22a) that the energy and demand rates $E_c(t_k^u)$ and $D_c(t_k^u)$ depend on the TOU. The rate structures are determined by utilities for various types of customers. For some rate plans, customers have the flexibility to choose peak periods so that they can save cost by optimizing energy use during specific time periods.
- This steady state optimization problem is different from our previous work (Mei & Xia, 2017a; Mei et al., 2018). In Mei and Xia (2017a), an open loop optimal control algorithm was proposed to minimize energy consumption by setting temperature, humidity and CO_2 concentration. In Mei et al. (2018), an open loop steady state optimal control algorithm is autonomously and adaptively setting optimal temperature, humidity and CO_2 concentration references, which could be time-varying to minimize energy cost and the PMV index. This study proposes an open loop optimal controller that minimizes the energy and demand charge costs under the PMV index within acceptable ranges. It reaches the same conclusions as Mei et al. (2018) in scheduling the reference setpoints. On the other hand, this study considers a DR action, which can further improve energy efficiency and reduce energy cost, it is different from our previous work (Mei & Xia, 2017a; Mei et al., 2018) without consideration of that action.
- The optimal solution applies to one zone, and the resulting reference setpoints are then communicated to the whole network through connecting neighbors. Therefore, the scheduling is implementable in a distributed manner.

The ADSMS is a hierarchical distributed way that aims at achieving energy and cost savings in ME A/C operations without compromising occupancy comfort levels. The information communication for the simplified ADSMS is illustrated in Fig. 3. The idea here is to consider comfort as a service for occupants. The zones (using zone modules (ZMs)) are customers seeking this service (called token), and a distribution system operator (DSO) is the service provider (called provider). There are four steps in the ADSMS that are explained in the following:

(1) Master: The DSO collects one zone's measurement information (cooling and pollutant loads, weather and occupancy), then computes and transmits optimal reference signals to this zone by a communication network.

(2) Slave: The neighboring zones receive communication information using numerous ZMs from the driving system. Then neighboring zones then communicate to whole zones through connecting neighbors.

(3) Token requests: The main aim of the ZM is to run an MPC using forecast information (weather condition, occupancy and cooling and pollutant loads) plus sensor readings (temperature and humidity, thermostat and CO_2 sensors) to compute the minimal energy consumption and cost required without breaching comfort ranges.

(4) Coordination: After each room receives communication information, each DX unit employs a DMPC algorithm to optimize the transient process of reaching thermal comfort and satisfying IAQ demands while minimize energy consumption and costs.

Remark 4. For ease of implementation, the min–max problem in (22a) is converted into the standard nonlinear programming described below so that it can be conveniently solved by the Matlab built-in functions. A new variable z_P is introduced to represent the peak demand of the day for zone i only as follows:

$$z_{P,i} = \max_{1\le k\le n}\{P_{tot,i}(t_k^u)\}. \tag{24}$$

By simplifying the objective to the form in (25), the optimization problem in (22a) can be rewritten as

$$\min\ \Big(w_1\sum_{k=1}^{n}E_c(t_k^u)P_{tot,i}(t_k^u)\delta^u + w_2 D_c(t_k^u)z_{P,i}\Big). \tag{25}$$

3.2. Lower level: DMPC

To conclude, the goal of the lower layer is to design the tracking rule $u(t_k^u)$ in a distributed way so that each subsystem of (5) can reach its steady states according to the changing environment during the day.

The UOPC transmits the reference signals, $x^r(s;t_k^u) = [x_1^r(s;t_k^u), \ldots, x_m^r(s;t_k^u)]^T, u^r(s;t_k^u) = [u_1^r(s;t_k^u), \ldots, u_m^r(s;t_k^u)]^T, T_{s,i}^r(s;t_k^u), \delta(s;t_k^u)$, to the LDMPC for $s \in [t_k^u, t_{(k+1)}^u), i = 1, 2, \ldots, m$. Here, $x^r(t_k^u) \triangleq x^r(t_k^u;t_k^u)$. In the lower layer, the DMPC controllers are designed to steer for the multi-zone building ME A/C system to track the trajectory references calculated by the upper layer. The linearized dynamic subsystem S_i for the nonlinear systems (5) around the trajectory references at sampling time instant $t_{c(k,q)}^l$ can be written as given below. In (26), the interacting terms in non-neighboring zone are ignored because of our definition of neighbors in (4).

$$\begin{cases} \delta\dot{x}_i(s) = A_{ii}(t_{c(k,q)}^l)\delta x_i(s) + \sum_{j\in\mathcal{V}_i} A_{ij}(t_{c(k,q)}^l)\delta x_j(s) + B_i(t_{c(k,q)}^l)\delta u_i(s), \\ y_i(s) = C_{ii}\delta x_i(s) + y_i^r(s), \quad s \in [t_{c(k,q)}^l, t_{c(k,q)}^l + T^l),\ i = 1, 2, \ldots, m, \end{cases} \tag{26}$$

where $A_{ii}(t_{c(k,q)}^l) = \frac{\partial f_i}{\partial x_i}(x_i^r(t_{c(k,q)}^l), u_i^r(t_{c(k,q)}^l))$, $A_{ij}(t_{c(k,q)}^l) = \frac{\partial f_i}{\partial x_j}(x_j^r(t_{c(k,q)}^l), u_j^r(t_{c(k,q)}^l))$, $B_i(t_{c(k,q)}^l) = \frac{\partial f_i}{\partial u_i}(x_i^r(t_{c(k,q)}^l), u_i^r(t_{c(k,q)}^l))$ for $j \in \mathcal{V}_i$. $\delta x_i(s)$ and

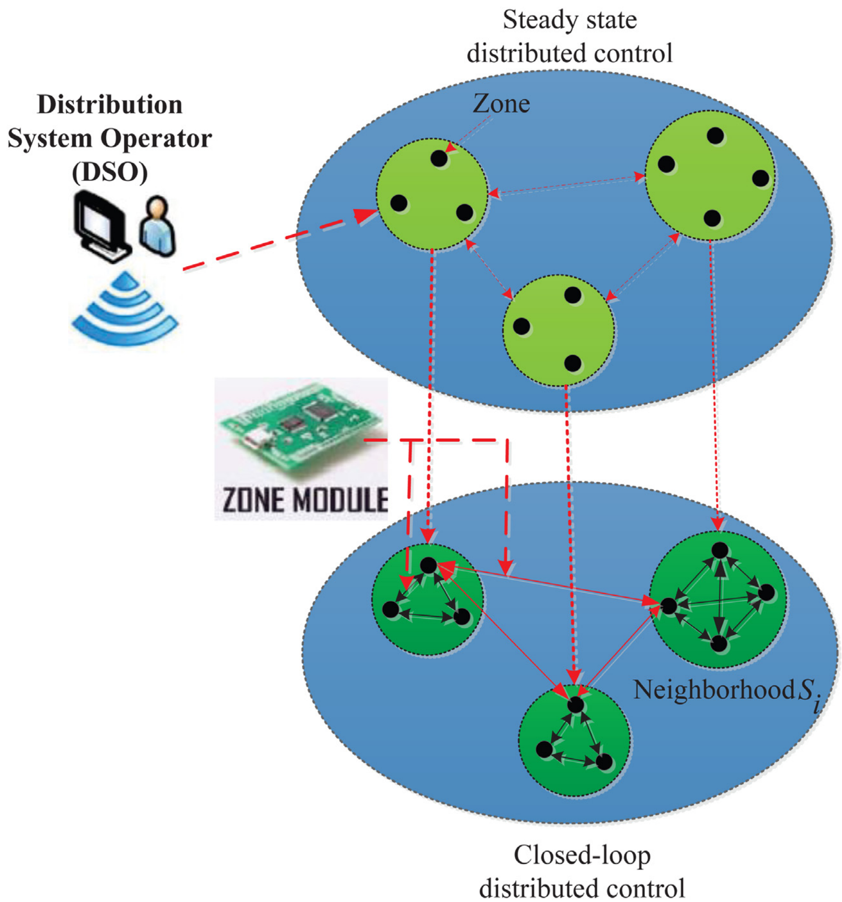


Fig. 3. Autonomous demand-side management scheduling architecture.

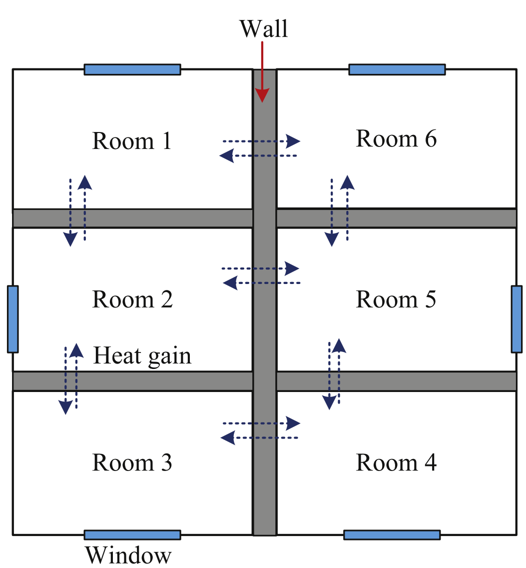


Fig. 4. Schematic of six-rooms building with the thermal network for all zones and its surrounding walls.

$\delta u_i(s)$ are the deviations of state and input from their trajectory references, respectively; $y_i = [T_{z,i}, W_{z,i}, C_{c,i}]^T$ are the original output variables; $y_i^r = [T_{z,i}^r, W_{z,i}^r, C_{c,i}^r]^T$ are the trajectory references in the lower layer, which are calculated in the upper layer.

The predicted subsystem S_i can be written as follows:

$$\begin{cases} \delta \dot{x}_i^p(s; t_{c(k,q)}^l) = A_{ii}(s; t_{c(k,q)}^l)\delta x_i^p(s; t_{c(k,q)}^l) \\ \qquad + \sum_{j\in\mathcal{V}_i} A_{ij}(t_{c(k,q)}^l)\delta\hat{x}_j(s; t_{c(k,q)}^l) + \\ \qquad B_i(s; t_{c(k,q)}^l)\delta u_i^p(s; t_{c(k,q)}^l), \\ y_i^p(s; t_{c(k,q)}^l) = C_{ii}\delta x_i^p(s; t_{c(k,q)}^l) + y_i^r(s; t_{c(k,q)}^l), \ s \in [t_{c(k,q)}^l, t_{c(k,q)}^l + T^l), \\ \quad i = 1, 2, \ldots, m, \end{cases} \tag{27}$$

where $\delta x_i^p(s; t_{c(k,q)}^l)$, $\delta u_i^p(s; t_{c(k,q)}^l)$ and $y_i^p(s; t_{c(k,q)}^l)$ are the predicted state, input and output trajectories at time step $t_{c(k,q)}^l$, $\delta\hat{x}_j(s; t_{c(k,q)}^l)$ is the assumed state sequence of S_i at time step $t_{c(k,q)}^l$.

The MPC algorithm is designed for the lower layer to minimize the optimization objective after reaching trajectory references as well as to handle building external disturbances and to compensate for the model mismatch. Let

$$\delta u_i^p(s; t_{c(k,q)}^l) = -\sum_{j\in\mathcal{V}_i} K_j(s; t_{c(k,q)}^l)\delta\hat{x}_j(s; t_{c(k,q)}^l) + v_i^p(s; t_{c(k,q)}^l), \\ s \in [t_{c(k,q)}^l, t_{c(k,q)}^l + T^l), \tag{28}$$

where $i = 1, 2, \ldots, m$, $K_j(s; t_{c(k,q)}^l)$ is the gain matrix from zone j, $v_i(s; t_{c(k,q)}^l)$ is a new input variable for zone i, then (27) is converted to (29) as follows:

$$\begin{cases} \delta \dot{x}_i^p(s; t_{c(k,q)}^l) = A_{ii}(s; t_{c(k,q)}^l)\delta x_i^p(s; t_{c(k,q)}^l) + B_i(t_{c(k,q)}^l)v_i^p(s; t_{c(k,q)}^l), \\ y_i^p(s; t_{c(k,q)}^l) = C_{ii}\delta x_i^p(s; t_{c(k,q)}^l) + y_i^r(s; t_{c(k,q)}^l), \ s \in [t_{c(k,q)}^l, t_{c(k,q)}^l + T^l), \\ \quad i = 1, 2, \ldots, m. \end{cases} \tag{29}$$

Many standard approaches exist in Ma et al. (2011) and Scherer et al. (2014) for the system (29), which depends entirely on one zone i. In this paper, we are using the MPC strategy proposed by our previous work (Mei et al., 2018), then the proposed optimization objective is given by

$$\min_{v_i^p(s;t^l_{c(k,q)})} \overline{J}_i^l = \int_{t^l_{c(k,q)}}^{t^l_{c(k,q)}+T^l} \Big(\left\| y_i^p(s;t^l_{c(k,q)}) - y_i^r(s;t^l_{c(k,q)}) \right\|^2_{Q_i} + \left\| v_i^p(s;t^l_{c(k,q)}) \right\|^2_{R_i} \Big)\, ds + \left\| y_i^p(t^l_{c(k,q)}+T^l;t^l_{c(k,q)}) - y_i^r(t^l_{c(k,q)}+T^l) \right\|^2_{P_i}, \quad i = 1,2,\ldots,m, \tag{30a}$$

subject to:

$$\delta x_i^p(s;t^l_{c(k,q)}) = A_{ii}(s;t^l_{c(k,q)})\delta x_i^p(s;t^l_{c(k,q)}) + B_i(s;t^l_{c(k,q)}) v_i^p(s;t^l_{c(k,q)}), \quad i = 1,2,\ldots,m, \tag{30b}$$

$$y_i^p(s;t^l_{c(k,q)}) = C_{ii}\delta x_i^p(s;t^l_{c(k,q)}) + y_i^r(s;t^l_{c(k,q)}),\ i = 1,2,\ldots,m, \tag{30c}$$

$$x_i^p(s;t^l_{c(k,q)}) \in \mathbb{X},\ v_i^p(s;t^l_{c(k,q)}) \in \mathbb{V},\ i = 1,2,\ldots,m, \tag{30d}$$

where $s \in [t^l_{c(k,q)}, t^l_{c(k,q)} + T^l)$, $\overline{J}_i^l$ is the lower layer objective function i, the controllers $\delta u_i^p(s;t^l_{c(k,q)})$ obtained are distributed. Q_i, R_i, P_i are the weighting matrix, $\mathbb{V}$ is a bounded set of the new input variable v_i. The convergence for the above finite horizon periodic MPC optimization problem (30) can be proved by the results in Xia, Zhang, and Elaiw (2011) and Zhang and Xia (2011).

The implementation strategy of the proposed AHDC algorithms for a multi-zone building ME A/C system can be summarized as follows:

The algorithm 1 in our previous work (Mei et al., 2018) is adopted to solve the upper layer distributed steady state optimization problem.

Algorithm: The lower layer DMPC algorithm can be given below.

(1) At sampling time instant t_k^u, $k = 0,1,\ldots,n$, UOPC receives each local neighbor's measurement information.

(2) UOPC computes the state trajectory $x^r(s;t_k^u) = [x_1^r(s;t_k^u),\ldots,x_m^r(s;t_k^u)]^T$, $s \in [t_k^u, t_{k+1}^u)$ and its corresponding control input trajectory $u^r(s;t_k^u) = [s;u_1^r(t_k^u),\ldots,u_m^r(s;t_k^u)]$, $s \in [t_k^u, t_{k+1}^u)$, which are transmitted to LDMPC, to obtain linearized systems (29).

(3) At sampling time instant $t^l_{c(k,q)}$, LDMPCi receives the state measurement $x_i(s;t^l_{c(k,q)})$ and $x_{-i}(s,t^l_{c(k,q)})$ from its neighbors, gives an initial point $x_i(0)$ ($k = q = 0$) and computes the optimal control input $v_i^*(s;t^l_{c(k,q)})$ of the optimization problems (30) over the prediction horizon $[t^l_{c(k,q)}, t^l_{c(k,q)} + T^l]$.

(4) The first solution $v_i^*(s;t^l_{c(k,q)})$ is used through (28) to update $\delta u_i^p(s;t^l_{c(k,q)})$ as the initial condition over the next prediction horizon $[t^l_{c(k,q+1)}, t^l_{c(k,q)} + \delta^l]$.

(5) If $0 \le q < M$, $q = q+1$ and go to (3); else $k = k+1$, $q = 0$ and go to (1).

4. Case study

In this section, a six-room model is considered to simulate the performance of the proposed AHDC strategy in special climate conditions in Cape Town, South Africa. The simulations are conducted during normal operation of an office building with normal occupancy. The six rooms are connected and the undirected graph is $\mathcal{G} = \{\mathcal{V},\mathcal{A}\}$ where $\mathcal{V} = \{1,2,3,4,5,6\}$ and $\varepsilon_0 = 5$. $R_{12} = R_{21} = R_{23} = R_{32} = R_{34} = R_{43} = R_{45} = R_{54} = R_{56} = R_{65} = R_{61} = R_{16} = 4 < \varepsilon_0$, $R_{13} = R_{31} = R_{24} = R_{42} = R_{35} = R_{53} = R_{46} = R_{64} = R_{51} = R_{15} = R_{62} = R_{26} = 8 > \varepsilon_0$, $R_{14} = R_{41} = R_{25} = R_{52} = R_{36} = R_{63} = 12 > \varepsilon_0$, then the neighbors of zone i are depicted in Table 1. As an illustrating example, Fig. 4 shows the schematic of a six-room building with the thermal network. It can be verified that the network is connected.

Table 1
The neighborhood definition of zones.

Room (i)	Neighbors ($\mathcal{V}_i$)	Room (i)	Neighbors ($\mathcal{V}_i$)
1	2,6	2	1,3
3	2,4	4	3,5
5	4,6	6	5,1

Table 2
Model parameters of the ME A/C system.

Notations	Values	Notations	Values
ρ	1.2 kg/m^3	h_{fg}	2450 kJ/kg
V_i	77 m^3	ε_{win}	0.45
V_{h1}	0.04 m^3	V_{h2}	0.16 m^3
k_{spl}	0.0251 kJ/m^3	C_a	1.005 kJ kg^{-1} °C^{-1}
$A_{0,i}$	22.07 m^3	R_i	15 °C/kW

Table 3
Coefficients of energy models.

Notations	Values	Notations	Values
$a_0 = 900.5$	$a_1 = -8.1$	$a_2 = 6.18$	$a_3 = -0.15$
$a_4 = -4.61$	$a_5 = 0.02$	$a_6 = -0.2$	$a_7 = 0.01$
$a_8 = 0.12$	$a_9 = 0.09$	$c_0 = 138.1$	$c_1 = 0.52$
$c_2 = -2.3$			

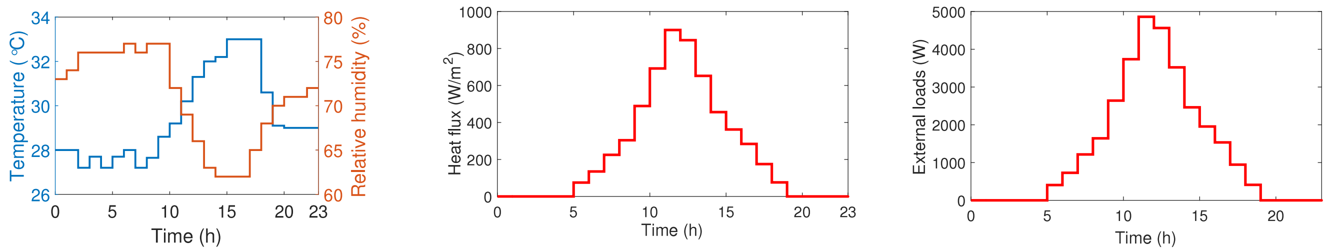
Table 4
Values of system constraints.

Notations	Values	Notations	Values
$\overline{T}_{s,i}$	22 °C	$\underline{T}_{s,i}$	8 °C
$\overline{T}_{z,i}$	26 °C	$\underline{T}_{z,i}$	22 °C
$\overline{T}_{d,i}$	22 °C	$\underline{T}_{d,i}$	10 °C
$\overline{T}_{w,i}$	22 °C	$\underline{T}_{w,i}$	10 °C
$\overline{W}_{z,i}$	12.3/1000 kg/kg	$\underline{W}_{z,i}$	9.85/1000 kg/kg
$\overline{C}_{c,i}$	800 ppm	$\underline{C}_{c,i}$	650 ppm
$\overline{W}_{s,i}$	9.85/1000 kg/kg	$\underline{W}_{s,i}$	7.85/1000 kg/kg
$\overline{v}_{f,i}$	0.8 m^3/s	$\underline{v}_{f,i}$	0.05 m^3/s
$\overline{m}_{r,i}$	0.11 kg/s	$\underline{m}_{r,i}$	0.005 kg/s
$\overline{h}_{s,i}$	46.3 kJ/kg	$\underline{h}_{s,i}$	27.3 kJ/kg
α	0.5		

4.1. Parameter selection

The volume of each room space is 77 m^3. The model parameters of the multi-zone building ME A/C system are given in Table 2. The coefficients of the power consumption models for the condenser (7) and evaporators (8) are calibrated through the regression analysis of the available measured data in Vakiloroaya et al. (2014), which are shown in Table 3. It is assumed that the combined total efficiency of the compressor $\overline{\eta}$ is 0.85. Each room has a window with the area of 4 m^2. For the proposed AHDC strategy considered below, the system variable constraints are given by bounds in Table 4, and we constrain the value of each room's PMV index in the range of $[-0.5, 0.5]$ to ensure that the multi-zone building ME A/C system is able to control each room's thermal comfort and IAQ at the required levels for occupants. The weighting factors are defined as $w_1 = 1$, $w_2 = 1$. In our previous work (Mei et al., 2018), the simulation results demonstrated that the open loop optimal controller and the MPC scheme are not sensitive to the model parameters A_1 and A_2 of the single-zone DX A/C system within any ranges of $[aA_0, bA_0]$ where $0 \le a,b \le 1$ and $a \le b$. This result can be extended to the multi-zone building ME A/C system. Hence, $A_{1,i} = 0.15A_{0,i}$ and $A_{2,i} = 0.85A_{0,i}$, $i \in \mathcal{V}$ are chosen in this paper.

The simulation runs from 0:00 to 23:59. The environmental temperature and relative humidity information are obtained from a meteorological station located in Cape Town, South Africa. The outside air temperature and relative humidity profiles are plotted in Fig. 5(a). The predicted solar radiative heat flux density profile is shown in Fig. 5(b). The external sensible heat load of each room is depicted in



(a) (b) (c)

Fig. 5. (a) Outside temperature and relative humidity. (b) Radiative heat flux. (c) External sensible heat load.

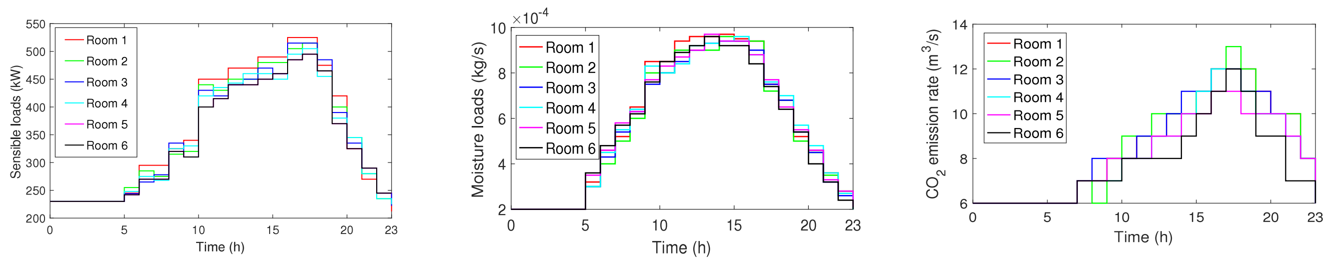


Fig. 6. Certainty internal sensible, certainty moisture and CO_2 emission loads.

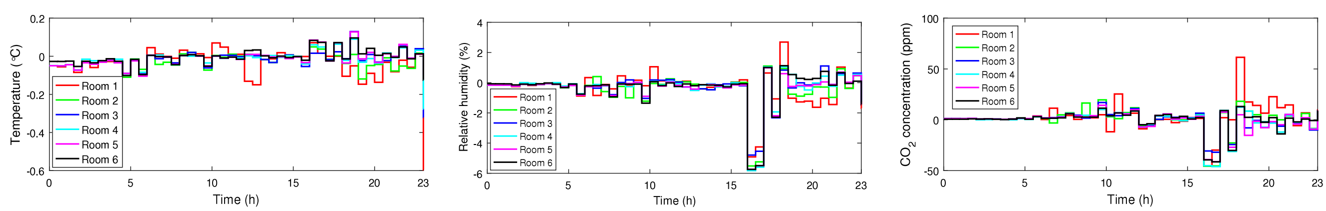


Fig. 7. The steady state errors in six-room building under the sampling periods 1 h and 0.5 h.

Table 5
Time-of-use electricity rates.

Summer	Period	Energy charge ($/kWh)	Demand charge ($/kWh)
Peak	12:00-18:00	0.20538	11.889
Standard	08:00-12:00, 18:00-21:00	0.05948	2.352
Off-Peak	21:00-08:00	0.03558	1.007

Fig. 5(c). The certainty internal sensible and latent heat loads and the CO_2 emission load of each room over a 24-h period are predicted in Fig. 6. The certainty loads mean the sensible heat and moisture loads from lighting, equipment and applications. The values of Figs. 5–6 at every hour are commensurately quantized for the lower layer.

It is assumed that the building operates under the TOU rate plan shown in Table 5. Since there is a big difference in the demand charges between peak and off-peak hours, energy cost savings can be expected if significant amounts of peak power consumption are shifted to non-peak hours.

4.2. Comparison of optimal scheduling control strategies

To illustrate the performance of the proposed AHDC, comparisons with other control strategies are considered for scheduling the operation of the multi-zone building ME A/C system. The first approach is the DMPC algorithm based on given setpoints of air temperature, humidity and CO_2 concentration, aiming at minimizing energy consumption, referred as S1 (Mei & Xia, 2017a). The second approach is the DMPC algorithm based on energy cost and the value of the PMV index minimization, referred as S2 (Mei et al., 2018). The proposed approach is the DMPC algorithm based on demand and energy cost minimization, referred as S3. To simplify the comparison, among the three strategies, the multi-zone building ME A/C system operation profiles are generated by employing an NLP algorithm under the same outside and inside conditions. The control parameters are listed below: The sampling time $\Delta = 2$ min is adopted to discretize the nonlinear multi-zone building ME A/C system. The prediction horizon of the lower layer DMPC scheme is set as $N = 15$; the sampling periods of UOPC and LDMPC are 1 h and 2 min, respectively. The total simulation time K^u is 24 h. To illustrate the sampling period without affecting the control accuracy, the steady state solutions under the sampling periods 1 h and 0.5 h are plotted in Fig. 7. It can be seen from Fig. 7 that the control accuracy is rarely affected by the setting sampling period. Table 6 lists the combinations of the optimization and control strategies in the three scenarios. The test results for the three scenarios are shown in Section 4.3.

4.3. Simulation test results

The performances of the three scenarios are compared through MATLAB simulations with historical weather data for a specific day. Fig. 8 shows the steady state profiles of air temperature, relative humidity and CO_2 concentration of each room, which are obtained by solving the distributed coordination optimization problem (22) and

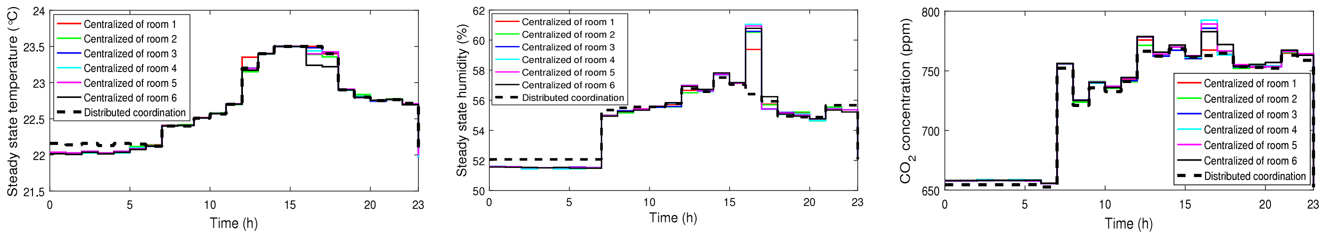


Fig. 8. The steady state in each room under the distributed and centralized optimal controller.

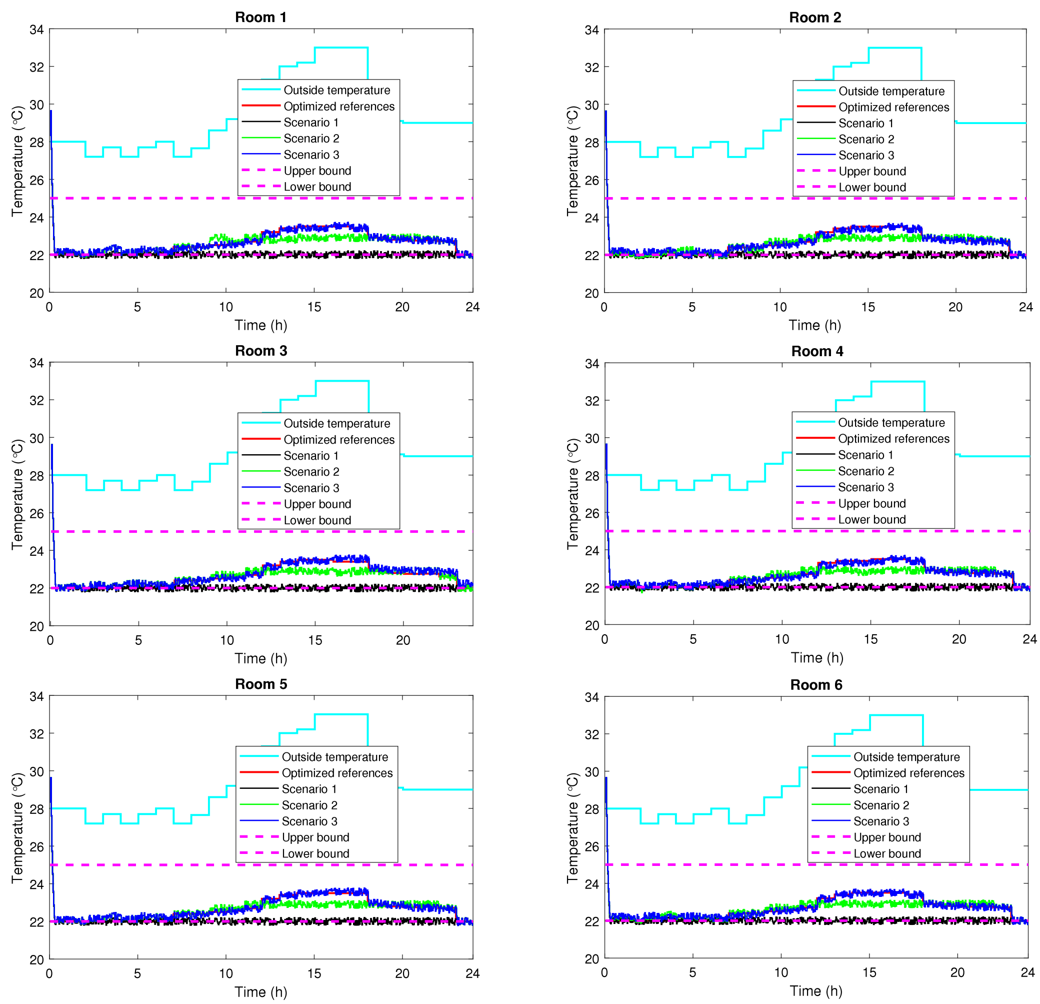


Fig. 9. Each zone's temperature profile for a 24-h period.

Table 6
Comparison of different control strategies.

Scenarios	Upper layer optimization	Low layer control	Setpoint	DR action
S1	Energy consumption	DMPC	Given	
S2	Energy cost+PMV	DMPC	Autonomous	
S3	Energy cost+demand cost	DMPC	Autonomous	✓

the centralized optimization (21). It can be seen from Fig. 8 that the distributed steady state is close to the centralized steady states of each room; the deviations are small and can be accepted by occupants (Assumption 1 is valid). The scheduling is thus effective.

The tracking reference points of the air temperature of each room with the three control strategies are depicted in Fig. 9 over a 24-h period. The tracking reference points of relative humidity of each room with the three control strategies are illustrated in Fig. 10 over a 24-h period. The tracking reference points of CO_2 concentration of each room with the three control strategies are shown in Fig. 11. Figs. 9–11 also show that the optimized reference points are adaptively preprogrammed by employing scenarios S2 and S3. We observe that each room's air temperature, relative humidity and CO_2 concentration, by using the proposed control strategy, are tracking and maintaining their reference points. It can be seen from Figs. 9–11 that the reference points of air temperature, relative humidity and CO_2 concentration of each room with scenarios S2 and S3 are raised during standard hours. The reason is that the controllers of scenarios S2 and S3 are automatically adjusting their reference points upward during standard hours according to the energy price policy and DR action respectively, such that the energy cost and energy consumption are minimized while both thermal comfort and IAQ are still maintained within comfort ranges. The pre-cooling and pre-decreasing CO_2 contaminant concentration

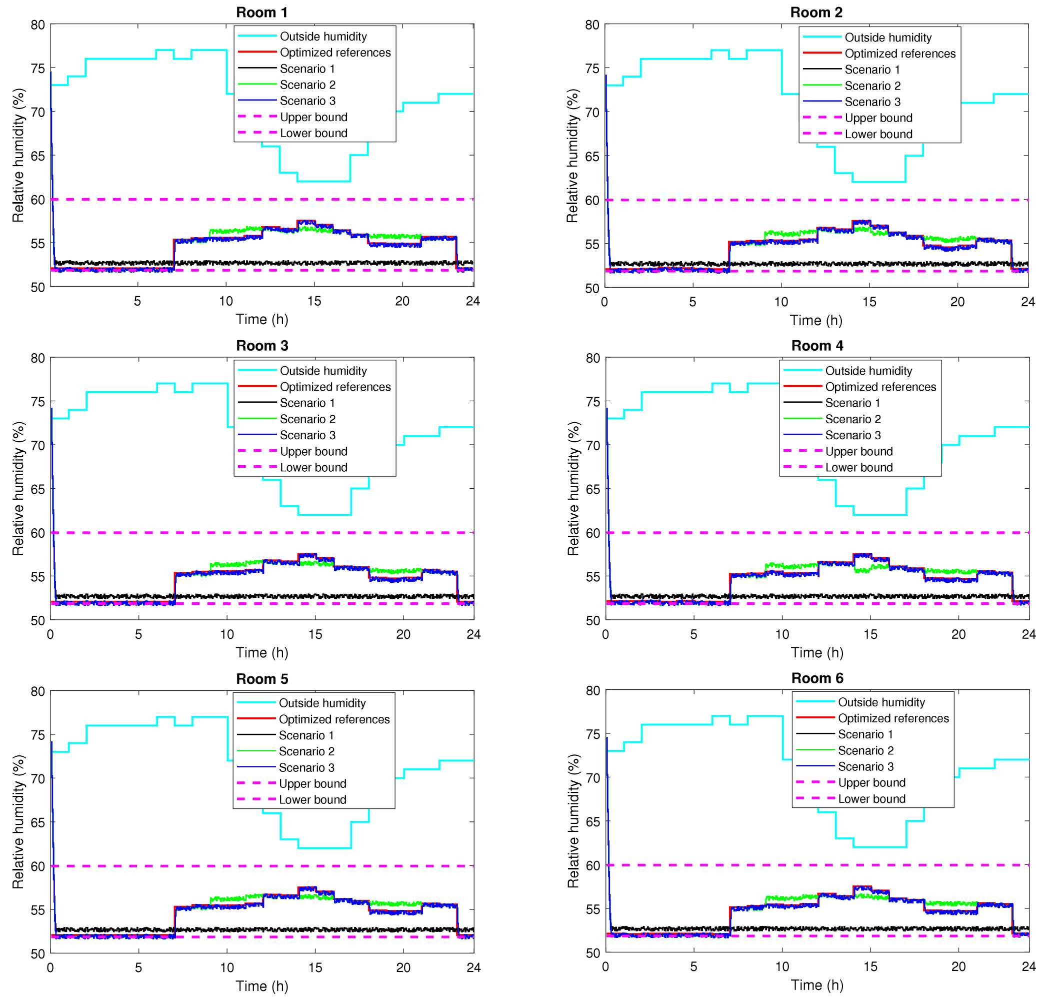


Fig. 10. Each zone's relative humidity profile for a 24-h period.

automatically starts in the morning simultaneously. This is because the energy costs for operating a multi-zone building ME A/C system during off-peak hours are lower than other periods. In the morning, the air temperature, humidity and CO_2 concentration reference points of all rooms are kept at the lower bounds of the comfort regions to store cooling and lower CO_2 contaminant concentration until the peak hours. As soon as the peak hours start, the reference points increase to the upper bounds, hence minimizing the demand in the afternoon by taking DR action. After more cooling and pollutant loads occur simultaneously during peak hours, the reference points are automatically set higher to turn off the cooling and increase the CO_2 contaminant concentration. We also observe that the time-varying reference points of air temperature, relative humidity and CO_2 concentration of each room are always maintained in the comfort regions over a 24-h period with the proposed control strategy. We further observe that after reaching their reference points, the proposed controllers are maintaining the reference points with small variation ranges. Therefore, the proposed control strategy is capable of handling the changing cooling and pollutant loads over a 24-h period and maintaining thermal comfort and IAQ at comfort levels. From Fig. 12, it can be observed that the values of the PMV index for the six rooms lie within the expected range [−0.5, 0.5], which indicates that the indoor air temperature, humidity and CO_2 concentration are controlled within their comfort ranges. It can be observed from Fig. 12, with the control method in Freire et al. (2008), the PMV index is controlled at the desired value, which indicates that the indoor air temperature and humidity are at their desired references, but it may not demonstrate that the indoor air CO2 concentration is within a comfort range.

Table 7
Comparison of different control strategies.

Strategy	Energy consumption (kWh)	Energy cost ($)
S1	124.56	10.67
S2	80.34	6.98
S3	79.78	5.66

To show the advantage of the proposed AHDC strategy over the other two control strategies in shifting demands from peak periods to non-peak periods, the power consumption under the peak and non-peak periods for the three control strategies are shown in Fig. 13. Table 7 summarizes the total energy consumption and cost for the multi-zone building ME A/C system under the three control strategies. From Table 7, it can be seen that with control strategies S2 and S3, more energy consumption and costs are reduced in comparison with control strategy S1. The reason is that each room's air temperature, humidity and CO_2 concentration reference points are adaptively and optimally preprogrammed under control strategies S2 and S3. We observe from Table 7 that the energy consumptions with control strategies S2 and S3 are almost the same, while the energy costs are different. It implies that the proposed control strategy S3 is capable of reducing more energy costs but not of reducing energy consumption in comparison with control strategy S2. It can be seen from Fig. 13 that under the proposed

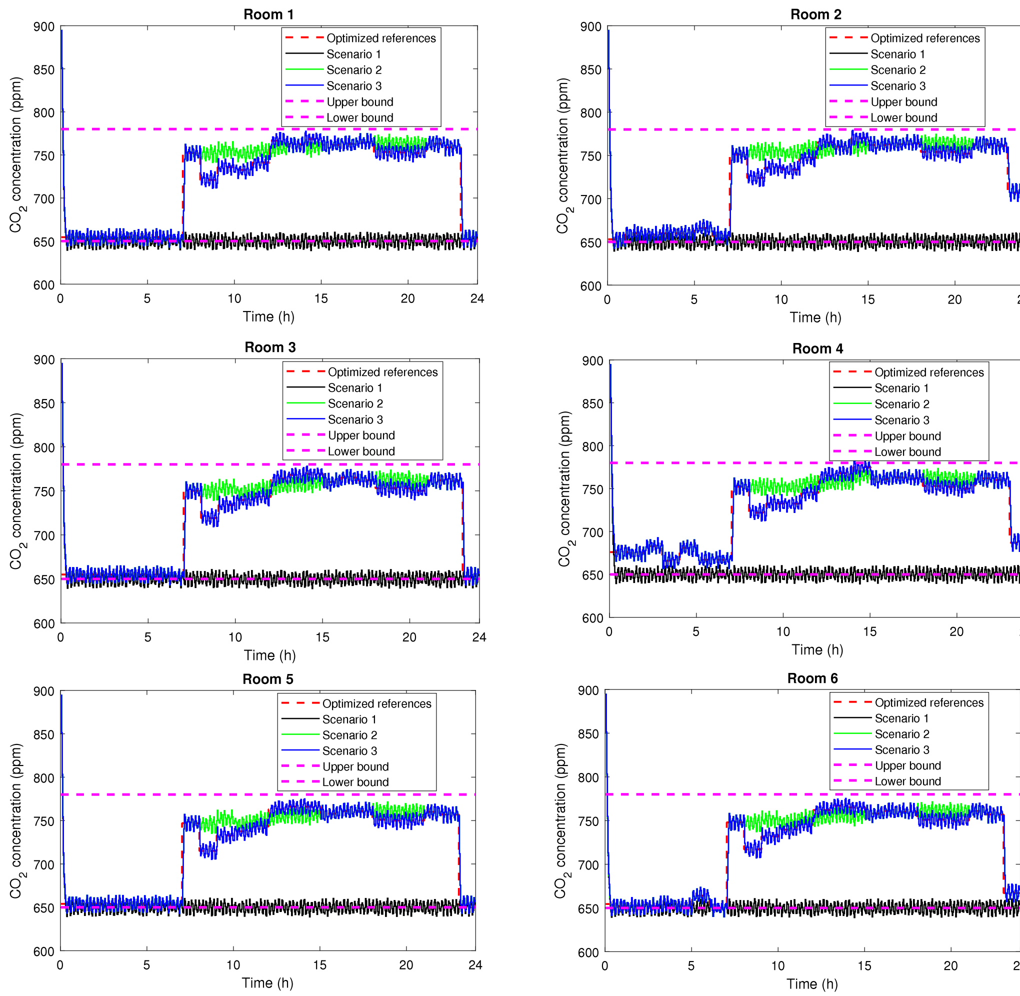


Fig. 11. Each zone's CO_2 concentration profile for a 24-h period.

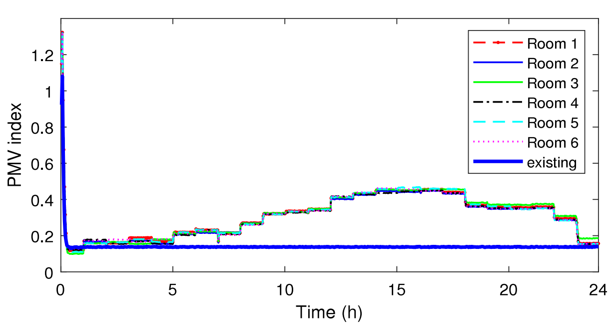


Fig. 12. Profile of the value of the PMV index for the six rooms over a 24-h period.

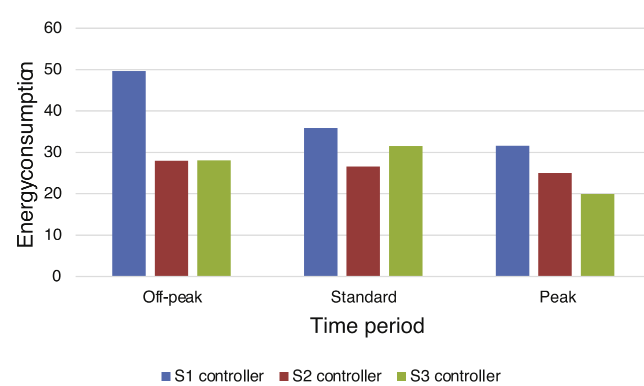


Fig. 13. Energy consumption in three time periods with the three control strategies.

control strategy S3 with DR action, more energy costs are reduced during peak hours in comparison with control strategies S1 and S2. The reason is that the proposed control strategy S3 is automatically shifting peak demands to non-peak periods. Meanwhile, energy consumption with the proposed control strategy S3 is more than that with control strategy S2 during standard periods because the energy cost in standard periods is lower than that in peak periods. Consequently, minimizing total energy costs and shifting demand are achieved over a 24-h period while maintaining both thermal comfort and IAQ at the required levels. Therefore, according to the above comparisons, the proposed control strategy S3 achieves a lower proportion of demand cost during peak hours and shows successful demand shifting and energy cost reduction.

Furthermore, to show the performance of the proposed distributed control strategies over the previous distributed control scheme (Scherer et al., 2014), we will compare the two control methods in view of energy efficiency in this section. The distributed control strategy in Scherer et al. (2014) is based on the given reference of indoor air temperature and a linearization system of the HVAC system with a solar plant by fixing the fancoil air speed. The distributed controller is then steered for the HVAC system to follow the given reference with minimizing energy consumption. In order to compare the two control strategies, the control scheme (Scherer et al., 2014) should be employed

Table 8
Compared with the previous control strategy.

Strategy	Energy consumption (kWh)	Energy cost ($)
Previous control (Scherer et al., 2014)	128.75	10.98
Proposed control	79.78	5.66
Saving	38%	48.5%

to steer the ME A/C system to follow the given reference and fixing volume flow rate of supply air. The comparison results are depicted in Table 8. It can be seen from the table that the proposed control strategy can reduce more energy consumption and cost in comparison with the previous control strategy (Scherer et al., 2014). The reason is that the proposed control scheme shifts the peak demand from the peak hours to off-peak hours by adaptively programming each room's setpoints of air temperature, humidity and CO_2 concentration.

5. Conclusion

This paper presents an AHDC strategy to the problem of minimizing demand and energy costs, as well as reducing communication resources, computational complexity and conservativeness for a multi-zone building ME A/C system while maintaining both thermal comfort and IAQ within comfort ranges. The developed control strategy is an improvement over the current control methods, in which the air temperature, humidity and CO_2 concentration references of each zone are adaptively preprogrammed optimal operation profiles for the multi-zone ME A/C system to minimize the energy and demand costs. The lower layer DMPC controllers steer the multi-zone building ME A/C system to follow and maintain the autonomously preprogrammed references; meanwhile, the energy and demand costs are reduced and shifted from the peak hours to non-peak hours. The simulation results show that the designed DMPC controller optimize the transient processes reaching the steady state and over the previous distributed control method in view of energy efficiency. They also demonstrated that the proposed AHDC strategy gives the controller the ability to handle model parameters uncertainty and time-varying weather conditions. The proposed control strategy is suitable for a cluster of similarly purposed buildings, thus requiring less and cheaper communication resources to implement.

Declaration of competing interest

The authors declare that they have no conflict of interest.

References

Atam, E. (2016). Decentralized thermal modeling of multi-zone buildings for control applications and investigation of submodeling error propagation. *Energy and Buildings*, *126*, 384–395.

Atthajariyakul, S., & Leephakpreeda, T. (2004). Real-time determination of optimal indoor-air condition for thermal comfort, air quality and efficient energy usage. *Energy and Buildings*, *36*(7), 720–733.

Bacher, P., & Madsen, H. (2011). Identifying suitable models for the heat dynamics of buildings. *Energy and Buildings*, *43*(7), 1511–1522.

Castilla, M., Álvarez, J. D., Berenguel, M., Rodríguez, F., Guzmán, J. L., & Pérez, M. (2011). A comparison of thermal comfort predictive control strategies. *Energy and Buildings*, *43*(10), 2737–2746.

Castilla, M., Álvarez, J. D., Normey-Rico, J. E., & Rodríguez, F. (2014). Thermal comfort control using a non-linear MPC strategy: A real case of study in a bioclimatic building. *Journal of Process Control*, *24*(6), 703–713.

Chen, W., & Deng, S. (2006). Development of a dynamic model for a DX VAV air conditioning system. *Energy Conversion and Management*, *47*(48–49), 2900–2924.

Christensen, C. C., Frey, H. M. M., Foenstelien, E., Aadland, E., & Refsum, H. E. (1983). A critical evaluation of energy expenditure estimates based on individual O2 consumption/heart rate curves and average daily heart rate. *The American Journal of Clinical Nutrition*, *37*(3), 468–472.

Cigler, J., Prívara, S., Váňa, Z., Žáčeková, E., & Ferkl, L. (2012). Optimization of predicted mean vote index within model predictive control framework: Computationally tractable solution. *Energy and Buildings*, *52*, 39–49.

Djongyang, N., Tchinda, R., & Njomo, D. (2010). Thermal comfort: A review paper. *Renewable & Sustainable Energy Reviews*, *14*(9), 2626–2640.

Elliott, M. S., & Rasmussen, B. P. (2013). Decentralized model predictive control of a multi-evaporator air conditioning system. *Control Engineering Practice*, *21*(12), 1665–1677.

Fanger, P. O. (1972). *Thermal comfort: Analysis and applications in environmental engineering*. Copenhagen, Denmark: Danish Technical Press.

Freire, R. Z., Oliveira, G. H. C., & Mendes, N. (2008). Predictive controllers for thermal comfort optimization and energy savings. *Energy and Buildings*, *40*(7), 1353–1365.

Gladyszewska-Fiedoruk, K. (2013). Correlations of air humidity and carbon dioxide concentration in the kindergarten. *Energy and Buildings*, *62*, 45–50.

Hu, J., & Karava, P. (2014). A state-space modeling approach and multi-level optimization algorithm for predictive control of multi-zone buildings with mixed-mode cooling. *Building and Environment*, *80*, 259–273.

Jiménez, M. J., Madsen, H., & Andersen, K. K. (2008). Identification of the main thermal characteristics of building components using MATLAB. *Building and Environment*, *43*(2), 170–180.

Kinney, J. M., Morgan, A. P., Domingues, F. J., & Gildner, K. J. (1964). A method for continuous measurement of gas exchange and expired radioactivity in acutely Ill patients. *Metabolism*, *13*(3), 205–211.

Lee, K., & Braun, J. E. (2008). Development of methods for determining demand-limiting setpoint trajectories in buildings using short-term measurements. *Building and Environment*, *43*(10), 1755–1768.

Lin, S., Chiu, S., & Chen, W. (2015). Simple automatic supervisory control system for office building based on energy-saving decoupling indoor comfort control. *Energy and Buildings*, *86*, 7–15.

Long, Y., Liu, S., Xie, L., & Johansson, K. H. (2016). A hierarchical distributed MPC for HVAC systems. In *2016 American control conference* (pp. 2385–2390).

Ma, Y., Anderson, G., & Borrelli, F. (2011). A distributed predictive control approach to building temperature regulation. In *2011 American control conference* (pp. 2089–2094).

Ma, Y., Matuško, J., & Borrelli, F. (2015). Stochastic model predictive control for building HVAC systems: Complexity and conservation. *IEEE Transactions on Control Systems Technology*, *23*(1), 101–116.

Ma, J., Qin, S. J., & Salsbury, T. (2014). Application of economic MPC to the energy and demand minimization of a commercial building. *Journal of Process Control*, *24*(8), 1282–1291.

Maasoumy, M., Razmara, M., Shahbakhti, M., & Sangiovanni-Vincentelli, A. (2014). Handling model uncertainty in model predictive control for energy efficient buildings. *Energy and Buildings*, *77*, 377–392.

Maasoumy, M., & Sangiovanni-Vincentelli, A. (2012). Total and peak energy consumption minimization of building HVAC systems using model predictive control. *IEEE Design and Test of Computers*, *29*(4), 26–35.

Mei, J., & Xia, X. (2017a). Energy-efficient predictive control of indoor thermal comfort and air quality in a direct expansion air conditioning system. *Applied Energy*, *195*, 439–452.

Mei, J., & Xia, X. (2017b). A reduced model for direct expansion air conditioning system and energy efficiency MPC control of indoor climate. In 13th *International conference on control and automation* (pp. 624–629).

Mei, J., & Xia, X. (2018). Multi-Zone building temperature control and energy efficiency using autonomous hierarchical control strategy. In *2018 IEEE 14th international conference on control and automation* (pp. 884–889).

Mei, J., Xia, X., & Song, M. (2018). An autonomous hierarchical control for improving indoor comfort and energy efficiency of a direct expansion air conditioning system. *Applied Energy*, *221*, 450–463.

Morosan, P. D., Bourdai, R., Dumur, D., & Buisson, J. (2011). A distributed MPC strategy based on Benders' decomposition applied to multi-source multi-zone temperature regulation. *Journal of Process Control*, *21*(5), 729–737.

Morosan, P., Bourdais, R., Dumur, D., & Buisson, J. (2010). Building temperature regulation using a distributed model predictive control. *Energy and Buildings*, *42*(9), 1445–1452.

Oldewurtel, F., Parisio, A., Jones, C. N., Gyalistras, D., Gwerder, M., Stauch, V., et al. (2012). Use of model predictive control and weather forecasts for energy efficient building climate control. *Energy and Buildings*, *45*, 15–27.

Qi, Q., & Deng, S. (2008). Multivariable control-oriented modeling of a direct expansion (DX) air conditioning (A/C) system. *International Journal of Refrigeration*, *31*(5), 841–849.

Radhakrishnan, N., Srinivasan, S., Su, R., & Poolla, K. (2018). Learning-based hierarchical distributed HVAC scheduling with operational constraints. *IEEE Transactions on Control Systems Technology*, *26*(5), 1892–1900.

Razmara, M., Maasoumy, M., Shahbakhti, M., & Robinett III, R. D. (2015). Optimal exergy control of building HVAC system. *Applied Energy*, *156*, 555–565.

Scherer, H. F., Pasamontes, M., Guzmán, J. L., Álvarez, J. D., Camponogara, E., & Normey-Rico, J. E. (2014). Efficient building energy management using distributed model predictive control. *Journal of Process Control*, *24*(6), 740–749.

Tang, R., Wang, S., Shan, K., & Cheung, H. (2018). Optimal control strategy of central air-conditioning systems of buildings at morning start period for enhanced energy efficiency and peak demand limiting. *Energy*, *151*, 771–781.

Treuth, M. S., Adolph, A. L., & Butte, N. F. (1998). Energy expenditure in children predicted from heart rate and activity calibrated against respiration calorimetry. *American Journal of Physiology-Endocrinology and Metabolism*, *275*(1), 12–18.

UNEP Sustainable Buildings & Climate Initiative (2009). *Building and climate change: Summary for decision-makers, Paris CEDEX 09*. France: Sustainable United Nations.

Vakiloroaya, V., Ha, Q. P., & Skibniewski, M. (2013). Modeling and experimental validation of a solar-assisted direct expansion air conditioning system. *Energy and Buildings*, *66*, 524–536.

Vakiloroaya, V., Samali, B., & Pishghadam, K. (2014). Investigation of energy-efficient strategy for direct expansion air-cooled air conditioning systems. *Applied Thermal Engineering*, *66*(1–2), 84–93.

Wallace, M., Mcbride, R., Aumi, S., Mhaskar, P., House, J., & Salsbury, T. (2012). Energy efficient model predictive building temperature control. *Chemical Engineering Science*, *69*(1), 45–58.

Wang, S., & Jin, X. (2000). Model-based optimal control of VAV air-conditioning system using genetic algorithm. *Building and Environment*, *35*(6), 471–487.

Weir, J. B. (1949). New methods for calculating metabolic rate with special reference to protein metabolism. *The Journal of Physiology*, *109*, 1–9.

Xia, X., Zhang, J., & Elaiw, A. (2011). An application of model predictive control to the dynamic economic dispatch of power generation. *Control Engineering Practice*, *19*, 638–648.

Xu, X., Yan, P., Deng, S., Xia, L., & Chan, M. (2013). Experimental study of a novel capacity control algorithm for a multi-evaporator air conditioning system. *Applied Thermal Engineering*, *50*(1), 975–984.

Yu, H., & Xia, X. (2017). Adaptive leaderless consensus of agents in jointly connected networks. *Neurocomputing*, *241*, 64–70.

Zafra-Cabeza, A., Maestre, J. M., Ridao, M. A., Camacho, E. F., & Sánchez, L. (2011). A hierarchical distributed model predictive control approach to irrigation canals: A risk mitigation perspective. *Journal of Process Control*, *21*(5), 787–799.

Zhang, X., Shi, W., Yan, B., Malkawi, A., & Li, N. (2017). Decentralized and distributed temperature control via HVAC systems in energy efficient buildings. arXiv preprint arXiv:1702.03308, [Online] Available: https://arxiv.org/abs/1702.03308.

Zhang, J., & Xia, X. (2011). A model predictive control approach to the periodic implementation of the solutions of the optimal dynamic resource allocation problem. *Automatica*, *47*(2), 358–362.

Zheng, Y., Li, S., & Qiu, H. (2013). Networked coordination-based distributed model predictive control for large-scale system. *IEEE Transactions on Control Systems Technology*, *21*(3), 991–998.

作者简介

夏小华，笔名野蕃哥哥（Yebogogo）。 南非比勒陀利亚大学工程、建筑环境和信息技术学院，电气、电子和计算机系教授，新能源系统中心主任，南非国家节能减排研究基地主任，武汉大学长江讲座教授，东北大学特聘教授，IEEE Fellow，南非 A 级教授，科学院、工程院院士。多年从事控制系统中非线性观测器的理论和设计、非线性和复杂控制理论方面的研究。他成功地为南非 Spoornet 开发了超长重载火车分布动力纵向运行控制系统；在艾滋病疫苗和药效上，有效地引入系统、模型辨识和控制方法；在南非电力总公司 Eskom 和 Exxaro, AngloGold 等矿业公司从事重大能源系统优化和节能减排以及能效评估方面理论与多项专题研究工作和工程项目。创建公司 Onga Energy，帅先获 ISO17020 牌照，能效评估曾占南非市场三分之一。现任 Applied Energy，Advances in Applied Energy 和 Annual Reviews in Control 杂志编委（Editorial Board Member），曾是《自动化学报》副主编，Automatica, Acta Automatica Sinica，IEEE Transactions on Automatic Control 和 IEEE Transactions on Circuits and Systems II 的副编辑（Associate Editor），SAIEE Africa Research Journal 的编辑。