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

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.

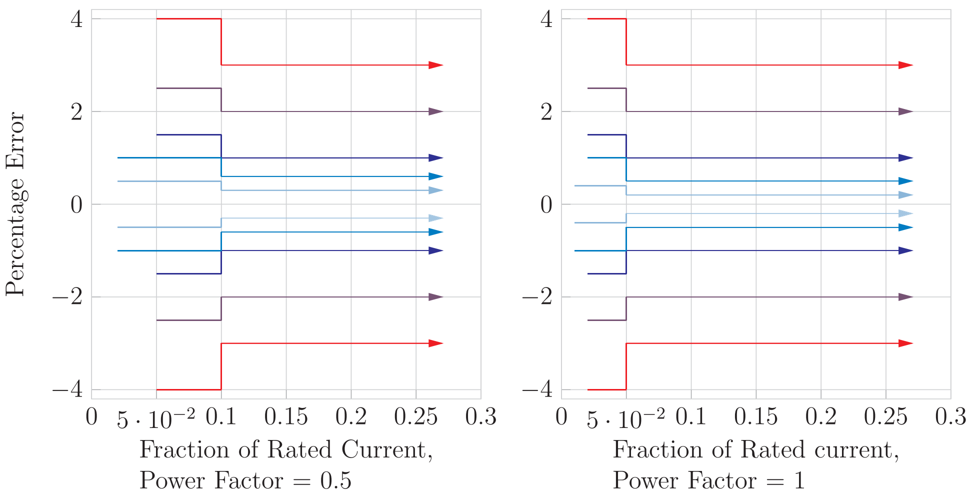

Fig. 1. Comparison of different IEC accuracy class meters [66–68] for transformer-connected single or polyphase meters with balanced loads under sinusoidal conditions.

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

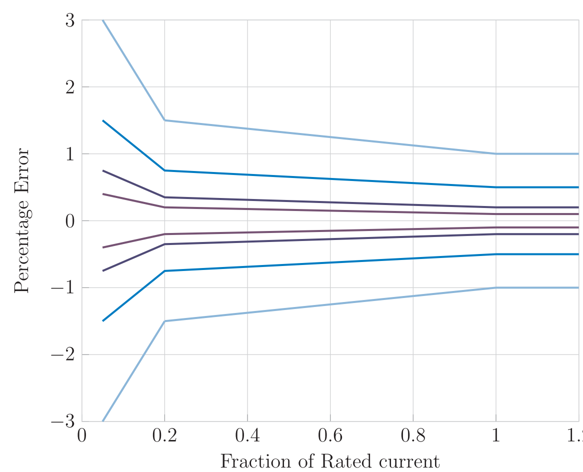

Fig. 2. Instrument Current Transformer accuracies according to IEC 60044-8 [70]. For Class 3 and Class 5, the limits are flat at 3% and 5% respectively.

used. One can consider less accurate and costly options in these applications.

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.

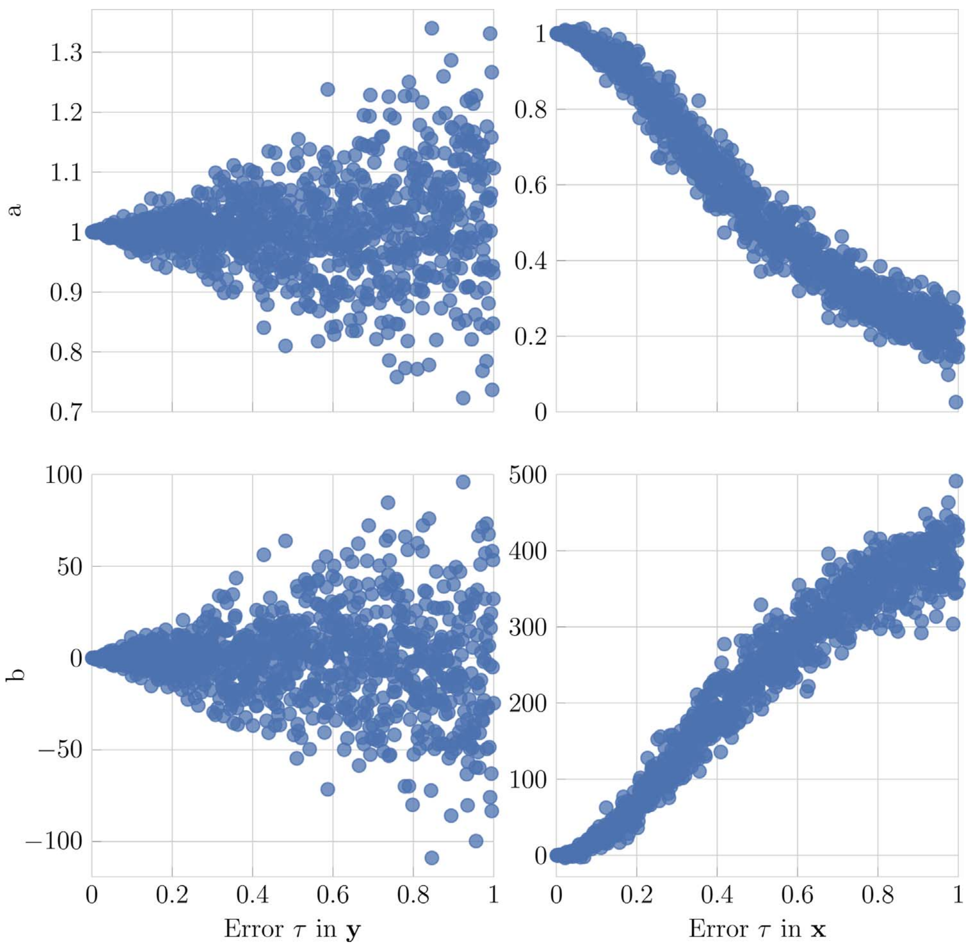

Fig. 3. OLS parameter estimates for $\mathbf{y} = \mathrm{a}\mathbf{x} + \mathrm{b}$, where $\mathrm{a} = 1$ and $\mathrm{b} = 0$, given measurement error τ in the form (2) and (3).

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.