Automation in Construction 41 (2014) 40–49

Contents lists available at ScienceDirect

Automation in Construction

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

The gap between predicted and measured energy performance of buildings: A framework for investigation

Pieter de Wilde *

Chair of Building Performance Analysis, Environmental Building Group, Plymouth University, Plymouth, United Kingdom
Plymouth University, School of Architecture, Design and Environment, Environmental Building Group, Chair of Building Performance Analysis, Roland Levinsky Building Room 301A, Drake Circus, Plymouth, Devon, PL4 8AA, United Kingdom

ARTICLE INFO

Article history:
Accepted 8 February 2014
Available online 2 March 2014

Keywords:
Energy performance gap
Prediction
Measurement
Building design and engineering

ABSTRACT

There often is a significant difference between predicted (computed) energy performance of buildings and actual measured energy use once buildings are operational. This article reviews literature on this 'performance gap'. It discerns three main types of gap: (1) between first-principle predictions and measurements, (2) between machine learning and measurements, and (3) between predictions and display certificates in legislation. It presents a pilot study that attempts an initial probabilistic probe into the performance gap. Findings from this pilot study are used to identify a number of key issues that need to be addressed within future investigations of the performance gap in general, especially the fact that the performance gap is a function of time and external conditions. The paper concludes that the performance gap can only be bridged by a broad, coordinated approach that combines model validation and verification, improved data collection for predictions, better forecasting, and change of industry practice.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Within the building industry, there is an increasing concern about a mismatch between the predicted energy performance of buildings and actual measured performance, typically addressed as 'the performance gap' [1–4]. Rapid deployment of automated meter reading (AMR) technology, typically now harvesting data at hourly or even half-hourly intervals, is making the performance gap more and more visible. The magnitude of this gap is significant, with reports suggesting that the measured energy use can be as much as 2.5 times the predicted energy use [4]. Increased pressure on the industry to address the challenges of environmental issues and rising energy prices makes it important to address this performance gap, with clients and the general public expecting new high performance buildings to meet increasingly stringent energy efficiency targets. While it seems reasonable to allow for some variation in both predictions and measurements due to the realities of uncertainties (inherent in predictions) and data scatter (inherent in measurements), the evidence seems to point to the gap presently being too wide to be acceptable.

Bridging the gap between predicted and measured performance is crucial if the design and engineering stage is to provide serious input to the delivery of buildings that meet their (quantified) ambitions, such as High Performance Buildings, Zero Carbon and Net Zero Energy Buildings. Bridging the gap is also crucial if the industry wants to deliver buildings that are robust towards change, that maintain a good performance throughout their lifetime, and that are engineered to adapt to changing use conditions in terms of 'occupant proofing' or 'climate change proofing'. Furthermore, it is a key prerequisite to novel modes of building delivery and facility management, enabling concepts such as performance based building, or performance-contracting, where occupants purchase a working environment with specified comfort boundaries rather than hardware (building and systems) that might – or might not – deliver such an environment [5,6]. In a wider context, the performance gap erodes the credibility of the design and engineering sectors of the building industry, and leads to general public scepticism of new High Performance Building concepts.

Energy efficiency is only one of the various performance aspects of buildings; it is highly likely that similar performance gaps exist between predicted and measured indoor air quality, thermal comfort, acoustic performance, daylighting levels and others. However, building industry and research presently focus on the energy performance gap; this might be due to the fact that energy metering is more prevalent and easier to implement than measurement of the other aspects. This paper aligns itself with this general focus on energy.

Energy performance of buildings can be studied at various levels of resolution. The primary view used in most studies is annual energy use of the whole building for heating and cooling purposes. However, one needs to be very careful in terms of including or excluding additional energy use for appliances, lighting, hot water and others. Energy efficiency can also be studied at higher temporal resolution using monthly, weekly, daily or even hourly data. A further differentiation relates to the

* Tel.: +44 1752 586115; fax: +44 1752 585155.
E-mail address: pieter.dewilde@plymouth.ac.uk.

http://dx.doi.org/10.1016/j.autcon.2014.02.009
0926-5805/© 2014 Elsevier B.V. All rights reserved.

object of study in energy performance analysis; at the design stage (prediction) this is linked to design intent, whereas post construction the measurement only applies to a building as actually constructed (instantiation). The energy performance gap typically concerns predicted performance of the design intent with observed performance of the realized building over the year.

Some discrepancy between prediction and measurement is inevitable due to numerical errors in simulation, and experimental variation in any observation [7], but getting reasonable agreement has been a key aim of tool developers ever since the inception of energy performance prediction methods, which started in the 1960s [8]. A good historical overview of various efforts in this direction is provided by Strachan et al. which although focussed on the development of the ESP-r simulation program is also applicable to many other similar efforts [9]. Their paper describes the validation of ESP-r in the context of a series of IEA Annexes (BESTEST) and related validation projects starting from the late 1970s. The key approaches used in much of this work are analytical validation, inter-program comparison and empirical validation, with the latter mostly based on results obtained from dedicated test cells; see for instance [10]. Yet these validation approaches are not without criticism. For instance Williamson [11] has pointed out that the analytical approach requires strong constraints and thus often does not reflect the real world, while inter-program comparison does not guarantee that any of the tools studied reflects what happens in the real world. But typically empirical validation is only possible for simple situations, not for full building complexity. In general, the field of verification, validation and testing (sometimes abbreviated as VVT) is still under development [12–14]. Interest is now also showing at the measurement side, most notably through the International Performance Measurement and Verification Protocol (IPMVP) [15,16]. Rapid developments in monitoring techniques and data mining techniques, including cheap sensors, radio-frequency identification (RFID) tags, and ubiquitous positioning, provide an increasingly high resolution map of reality and hence set higher benchmarks for performance predictions [17].

Indications of the 'performance gap' as addressed in this work started to appear from the mid-1990s [18], with a continuous coverage to the present day [1,19–22]. It must be noted that this performance gap is positioned in a different context than the above validation efforts: it addresses the differences between prediction and measurement of the energy performance of a complete building, including the full complexities of sub-systems, control settings, occupant behaviour, climate conditions, and others. Also, it is important to emphasize that in true prediction, made when the project still is in the design stage, there is typically only a description of a building, but no actual object—apart from a case where the design involves the renovation of an existing building; see for instance Sanguinetti [23].

This article develops a framework for further investigation of the magnitude of the performance gap, and for R&D, efforts towards narrowing or bridging the gap. It first provides a critical review of current literature on the subject, both in terms of root causes and solutions, and then continues to develop a fundamental position that distinguishes three different views of the energy performance gap. From this perspective the discussion then focuses on a pilot study that attempts an initial probabilistic probe into the performance gap. Finally, findings from the pilot study are used to identify a number of key issues that need to be addressed within future investigations of the performance gap in general.

2. Root causes

Literature on the energy performance gap suggests various causes for the mismatch between prediction and measurements. These causes can be grouped in three main categories: causes that pertain to the design stage, causes rooted in the construction stage (including handover), and causes that relate to the operational stage. Note that the specific issues which cause a performance gap will vary from one building to another; in many cases there will be a combination of several issues.

Within the design stage, a first cause towards later performance discrepancies is frequently within the design itself. Issues can start from mis-communication about performance targets for the future building between client and design team, or between the members of the design team [2,24,25]. A further key problem is that design teams often cannot fully predict the future use (functions) of the buildings; operational requirements and conditions might thus be subject to significant change [4,25–28]. It is also possible that, in terms of energy performance, the building design itself is inadequate through poor thermal concept, overspecification (oversizing of HVAC components), or lack of appropriate detail. Even if the design itself is energy efficient, lack of attention to buildability, simplicity, sequencing of the construction process, or of appropriate detail might be a built-in source for later underperformance [3].

It has been suggested that there might be issues with energy saving technology for buildings, especially in those buildings that aim to be more efficient, 'green', or 'high performance' than the average design. Equipment might simply not perform as well as specified by the manufacturer, either by nature or by over-optimism on system acceptance by the intended users [1,24]. Novel and advanced systems might be specifically prone to underperformance and 'teething problems' [28]. Many energy saving systems appear to be overly complex, as are the controls for these systems [3]. Additionally, many systems in general are becoming increasingly dependent on software for their operation, requiring updating (evolving) of this software to keep pace with changes to the environment, thus adding an extra layer of complexity [29].

Obviously, the second cause of a performance gap within the design stage relates to modelling and simulation as they are the key components of any prediction. Any use of incorrect methods, tools or component models will result in unreliable predictions and a gap later down the line [2,4,25]. The correct use of tools alone is insufficient; the tool user/analyst/modeller also needs to have the right knowledge and skills and the ability to apply these in the right manner [30]. This includes a good overview of the application area of models and methods, and correct data input definition. Note that within a generally well-performing method there might be component models that still have issues [31]. Even with a correct model applied by a well-trained analyst, all predictions remain subject to fundamental uncertainties, especially with regards to variation in aspects such as actual weather conditions, occupancy schedule, internal heat gains, and plug loads [1,4].

Linking back to the design context, it is also suggested that there can be a mis-alignment between design and prediction. Quoting the Zero Carbon Hub report [3], "*Calculations and modelling are often divorced from design and the mechanisms for ensuring that modelling is an accurate reflection of what is built are weak*". The same source suggests that there is an unfortunate lack of formal error and accuracy testing of detailed design calculations, and that typically there is no modelling/calculation audit trail [ibid]. The Carbon Trust points out that it would be beneficial to test designs for robustness, ensuring that these designs can accommodate change of use and occupancy. This does not seem to be the case in current modelling practice [2]. Finally, Williamson [11] points out that present approaches do not take system performance deterioration into account, which again will lead to a mismatch between prediction and measurement.

A different group of causes for the performance gap arises from the actual construction process and the handover to the client. Many authors point out that the quality of building is often not in accordance with the specification, with insufficient attention to both insulation and airthightness [4,24,25]. Often, details are left unspecified and for the contractor to define, with potential risks for the creation of thermal bridges; or on-the-job solutions leading to unexpected extra wood in timber-frame walls that can change the overall performance [32]. Further discrepancy between design and actual building is introduced by change orders and value engineering [1,32]. Problems where actual construction does not meet specification might be hard to spot, as buildings often consist of various layers (for instance cavity walls); some problems will be evident from measurement but in many cases visual inspections are needed to establish actual issues [33]. Efforts are under way to improve the handover (commissioning) process, such as the 'Soft Landings' process [34]. Soft Landings is a process that has been developed in the UK which aims to keep designers and constructors involved in the performance of buildings beyond completion, through extended aftercare of up to three years. This is typically tied into a longer and more rigorous quality assurance drive, including pre- and post-handover process (continuous commissioning), yet problems in this area remain [2,3,25]. It is also noted that at present there still are issues concerning proper tests to assess the performance of new constructions once a building is complete [3].

Once a building is commissioned and in use, the operational side also contributes to the performance gap. Occupant behaviour is often different from the assumptions made in the design stage; this is often cited as the main reason for the performance gap [4,21,25,26,35]. Seminal studies on the behaviour of office occupants and how to model this have been conducted by Haldi and Robinson [36]. Further efforts to better understand and model occupant behaviour and appliance use are ongoing [35,37,38]. More specifically, assumptions regarding occupant behaviour often lead to a mismatch between input for any calculations/simulations and actual values for internal gain [35,39] and plug loads [24]. Technological developments can also cause a mismatch; for instance IT-related loads are often higher than anticipated [27]. Furthermore, the actual operation of the building is typically different from the idealized assumptions made in the design stage, both in terms of actual control settings (such as thermostat settings, operation hours, BEMS settings) as well as the broader scope of facility management (FM) [1,2,4,25,27,28,40].

At the operational stage, it is also important to realize that there always is uncertainty in experimental data. In a complex system like a building, there will typically be a whole network of sensors; it is a challenging task to ensure that all sensors operate properly and that data is registered in the correct manner [41]. While basic measurement such as energy consumption can be expected to be covered, there appears to be less coverage in terms of observing the overall state of the building, with gaps in the measurement science [42] and a lack of standardization and continuity of monitoring, analysis and control throughout the building life cycle [43].

Literature on the performance gap also hints at deeper underlying factors for the performance gap. This includes the overall industry culture of construction, where traditional processes are hard to change and where there are often issues with quality, integrity, and responsibility [3,33,44]. Poor client knowledge and labour skills are also contributing to the situation [3]; regulatory pressure to make buildings more energy efficient might in fact contribute to over-optimistic predictions and hence contribute to the performance gap [45]. More specifically, the lack of routine Post Occupancy Evaluation (POE) processes means that there is no mechanism in place to improve on past performance [4,34].

3. Ongoing efforts to bridge the gap

Suggestions on how to best bridge this gap presented in literature are generally aligned with the root causes, and cover design and prediction, construction, and measurement once buildings are operational. As the performance gap for a specific building can stem from any cause listed, there is a need to address the whole field [25]. As stated in the Zero Carbon Hub report: "*In developing a solution to the problem of underperformance it is clear that the focus should be on improving the robustness of design and construction, a task that will involve the whole of industry, not just those designers and developers in the front line*" [3].

Regarding design, efforts to bridge the gap mainly take the form of development of design guidance and reports that aim to raise awareness [2]. The suggestions provided are typically quite generic: raise awareness amongst clients and design teams [21], ensure that ambitions, design intent, and responsibilities are clearly communicated [2], develop design details that are thermally robust and leave no room for error during construction [32], and prevent complex technological solutions and controls [2]. Some authors believe that introducing more strict regulation for design, for instance making the passivhaus standard mandatory, would contribute significantly to reducing the gap [44]. The design stage should also prepare the ground for extended contractor involvement, beyond the construction process only, independent commissioning, monitoring, and POE [2,34].

Specific efforts to address the performance in terms of its prediction methods and tools are part of the continuous work to improve the quality of these tools and methods in general. Within this wider context, André et al. suggest the development of 'reference simulations' which in fact take the form of a set of validated building component/subsystem models that help increase the rigour of a building simulation in which these components are used [46]. This fits well with more generic work on the performance of component-based software, albeit noting that it is important to pay attention to both the components and the deployment platform [47]. Raftery et al. stress the importance of the feedback loop from measurement to prediction, stating that "*… in a research context, the calibration process can provide feedback to improve the quality of future design stage models by identifying common mistaken assumptions in these models and by developing best-practice modelling procedures*" [48]. Within the wider computing contexts, efforts are underway to link prediction and measurement in integrated building information systems; while this does not necessarily bridge the gap it at least works towards software environments where data from both sides can be compared and contrasted [49,50].

Within the construction process, efforts such as Building with Care [44] attempt to increase the quality of the delivery process. However, this requires a change of culture across the whole supply chain and hence is difficult to achieve [3,32]. It is likely that to achieve success in this area a combination of incentives and penalties is needed to change the industry culture, which might require legislative approaches [3]. Alternative construction methods such as off-site construction might also contribute to change [45]. Commissioning and aftercare should preferably be conducted through processes such as Soft Landings [34].

On the measurement side, there is an awareness that current monitoring approaches need further improvement. Turner and Frankel state that "*within each of the metrics, measured performance displays a large degree of scatter, suggesting opportunities for improved programs and procedures*" [1]. One avenue that is being pursued is the use of system level metering in order to gain more in-depth information [28]; this could take the form of separate metering approaches for the building fabric, building services, and occupants [46]. Others argue for a mixed approach: "*Certain applications require quantitative measurements, forensic investigation, qualitative insights or a combination of all the above. Sometimes imperfect measurements can be enough to inform us that there is a problem, but we will often need reliable data to back this up and persuade others of the severity of the problem*" [33]. In general, measurements should be spread out over time, taking a longitudinal approach and covering various seasons [28]. A specific performance evaluation approach combining qualitative and quantitative techniques (one of which is measurement) that received a lot of attention is POE. For instance Menezes et al. [4], Morant [25], Bell et al. [32] and Soft Landings [34] all consider it key to obtain more detailed information on occupancy, occupant behaviour and the related lighting, small power/plug load electricity consumption. However, there is presently no standard POE methodology; occupant behaviour is especially complex and hard to capture [51]. Even if POE data is available, translating this data into input for computational predictions in a way that reduces the performance gap is difficult. For instance, one can collect data from a large sample of buildings and use that to generate mean occupancy values, but this will not help to predict actual occupancy in a specific case. For a deeper discussion of this issue, see Mahdavi [52].

Finally, there are suggestions that work is needed to bridge the performance gap in building regulations. The Zero Carbon Hub report suggests that the industry needs to move to a situation where over 95% of houses meet the performance required [3]. The same report also suggests the need to redevelop performance indicators for building energy performance, which ought to include confidence factors that reflect the amount of control over the performance such as; accreditation of design, construction and systems; more rigorous post completion testing; and audit arrangements [ibid]. Within the European context, it is recognized that there is a need to develop tools and methods in order to improve energy performance certificates [53].

4. Methodology

From the review of literature on the performance gap, it becomes obvious that there are different viewpoints regarding the gap. Any research directed at a deeper investigation of the gap, or efforts to bridge the gap, requires a solid position in regards to these different viewpoints. Thus, Section 5 of this paper develops a classification that discerns three main types of performance gap.

The work reported here is based on the premise that uncertainties will be present in both energy performance predictions and in energy performance measurements. A fundamental approach for dealing with these uncertainties is to develop probability density distributions (PDF) for both the prediction and measurement, instead of using deterministic values. One then needs to study the equality of these two PDFs using statistical tests, such as, for instance, the two-sample Kolmogorov–Smirnov (K–S) test. See Fig. 1.

Unfortunately, the complexities of building energy performance in an operational context make it very difficult to apply this approach to the performance gap. To explain the constraints, this paper will present a pilot study into a campus building at Plymouth University in the UK: the Roland Levinsky Building. For this building, a detailed EnergyPlus model has been developed and used in past projects, giving a solid starting point for simulation studies [54]. Two years of detailed monitoring of energy use, including gas and electricity, is available for comparison, with data taken from the Building Energy Management System (BEMS) and logged at a 30 min interval.

Constraints and limitations identified from the pilot study will be used to develop a framework for further research into the performance gap. This covers approaches to quantify the magnitude of the performance gap, target and benchmark values for acceptable correspondence of prediction and measurement, as well as identification of fundamental aspects that need to be addressed in efforts to bridge the gap.

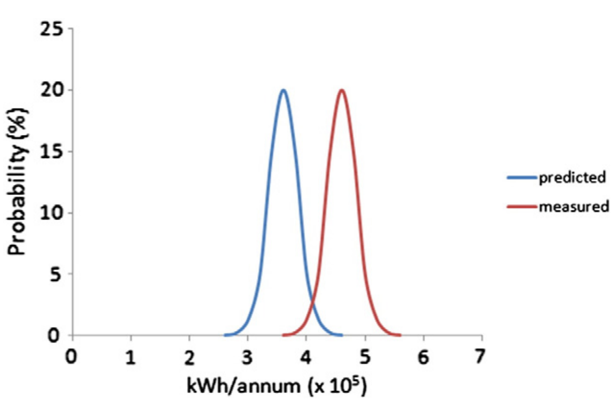

Fig. 1. Theoretical mapping of gap between predicted and measured energy performance.

5. Performance gap typology

While at first sight the concept of a performance gap between predicted and measured energy performance seems simple enough, in reality it is more complex: in the world of building design and engineering (as well as research), there are various approaches to both prediction and measurement. Additionally, there is another layer of building regulation which involves both prediction and measurement; this substantially complicates the discussion about the performance gap. The resulting complex landscape around the performance gap is depicted in Fig. 2.

Most energy performance prediction is based on first-principle modelling. Here knowledge about building physics and building systems is used to develop a physical model. This physical model is translated into a computational model, which is solved and used to generate predictions. For a more rigorous discussion about first-principle modelling, see for instance Underwood and Yik [55]. These first-principle models span a wide range, from stationary calculations, via semi-dynamic methods, to dynamic simulation. In general, the complexity of the model increases when moving from stationary methods to simulation. This not only pertains to the representation of indoor and outdoor conditions (as implied by the name, these are fixed at average values in stationary calculations, but vary dynamically with a time step – hourly or smaller – in simulation) but typically is also reflected in the representation of building systems, control settings and occupant behaviour. There is nothing to preclude a stationary calculation from yielding a decent prediction, provided that good input values are used in the calculation. Yet in general there appears to be a tendency to place more trust in models that represent all systems in more detail, in other words: dynamic energy simulation. For more information on building energy simulation, see Hensen and Lamberts [56].

There is also a second route for energy performance prediction, which is often overlooked in literature on the performance gap. This is based on developing a correlation between input parameters and output parameters without explicitly modelling the systems and physical processes (black box approach). This approach is broadly known as machine learning and covers techniques such as regression analysis, artificial neural networks, and support vector machines. An important issue is that machine learning techniques require training data; this can be provided by either measurement data, or from first-principle models. For recent work in this area, see for instance Tsanas and Xifara or Kusiak and Xu [57,58]. Obviously, better training sets – for instance obtained through a longer observation period – will provide better predictions [59]. An option only recently gaining attention is the use of first-principle modelling to train machine learning [60,61].

Measurement also needs further differentiation. Firstly, there is the issue of system resolution of the meters. The most basic metering will just capture the energy delivered to a building, such as kWh electricity or m^3 of gas. However, a building can also be equipped with metering of sub-systems, all the way down to individual components and appliances. Secondly, there is the option of capturing energy-related data, such as indoor and outdoor conditions, occupancy, and control systems. Thirdly, there is also variation in temporal resolution of the measurements, from a minimum of annual values all the way to intervals of only a few minutes or even seconds. Broadly speaking, annual facility meter readings are at the low end of the spectrum. Monthly readings, often looking into a limited number of sub-sets, is a clear next level and often can be linked to utility bill analysis. The high end of the spectrum is covered by Automated Meter Reading (AMR); the emerging frequency in the UK currently is half-hourly readings. AMR is often linked with information obtained from the Building Energy Management System (BEMS); it can also be expanded with bespoke meters. For a rigorous discussion of the state of the art in this area, see [62,63], or for a non-building specific discussion of performance monitoring in engineering systems, see [64].

Legislation is now also concerned with the energy performance of buildings; see for instance the European Energy Performance of

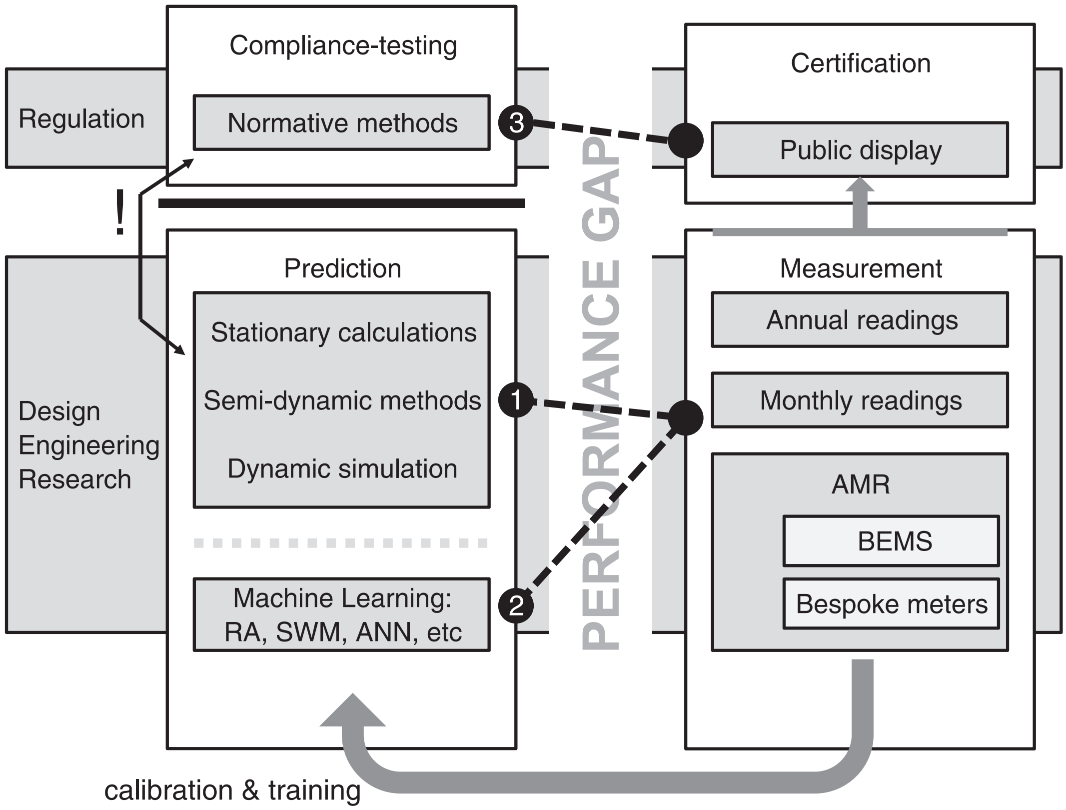

Fig. 2. Conceptual map of various views of the performance gap.

Buildings Directive (EPBD) or the US ASHRAE Standard 90.1. Typically legal frameworks require a normative 'pass or fail' energy use calculation in order to enforce a minimum standard of efficiency. The EPBD also requires actual energy use monitoring once the building is constructed and in operation. However, while this has some similarities with prediction and measurement as discussed above, there are fundamental differences with the design/engineering/research approach. Within the normative energy use calculation, regulation needs to impose an unambiguous ranking that allows to order designs according to preference and the ability to set strict pass-or-fail thresholds. This ranking needs to be fully reproducible and will have to be able to stand in court; it does not allow for any uncertainties or variation. On the measurement side, regulation takes meter readings into a unified format (for instance translating readings into Energy Use Intensity (EUI) figures) to report these in a way that allows cross-comparison of all building stock. See [65] for more information on building energy legislation in Europe and [66] for more information on energy performance benchmarking.

Typically the prediction side of regulation only looks into heating and cooling load, while the measurement side includes the impacts of plug loads, occupant behaviour, and even climate variation. This automatically creates a performance gap and makes the two incommensurable on principle, which is stressed in Fig. 2 by the shift in alignment between normative methods and public display. At the same time, there is an unfortunate trend that both sides often result in an energy label on a scale from A to G. Typically it appears that compliance testing yields better scores than energy display certification. Fig. 3 presents results from a sample of buildings in the UK that was selected by the author. All buildings in this sample have been highlighted as a high performance building, for instance through a high BREEAM rating, passivehouse certification, inclusion in the case study list of the UK's Concrete Centre, or having achieved a RIBA sustainability prize. Energy performance certificates (EPC, for compliance testing) and display energy certificates (DEC, for public display) were obtained from the UK's Non-Domestic Energy Performance Register (www.ndepcregister.com). Only in two out of twenty cases are labelling results the same; in all others the compliance rating is better than the display rating. Again, it must be stressed that in the UK EPC and DEC certificates are not capturing the same energy use; EPCs only address heating and cooling while DECs also include 'unregulated emissions' such as plug in appliances.

	Credentials	Building type	EPC	DEC
building 1	BREEAM Excellent	court	B	D
building 2	BREEAM Excellent	court	B	E
building 3	BREEAM Excellent	data centre	A	F
building 4	BREEAM Excellent	education	B	F
building 5	BREEAM Excellent	education	B	D
building 6	BREEAM Excellent	education	B	D
building 7	BREEAM Excellent	office	B	C
building 8	BREEAM Excellent	office	A+	E
building 9	BREEAM Outstanding	education	B	G
building 10	BREEAM Excellent	court	D	D
building 11	BREEAM Excellent	education	C	C
building 12	BREEAM Excellent	education	B	C
building 13	BREEAM Excellent	education	B	E
building 14	passivehouse	education	A+	B
building 15	concrete center case	education	B	E
building 16	concrete center case	education	B	F
building 17	RIBA prize	office	A	B
building 18	RIBA prize	office	B	C
building 19	RIBA prize	healthcare	B	E
building 20	RIBA prize	education	B	D

Fig. 3. Comparison of outcomes of legislative energy assessment in the UK. EPC: energy performance certificate (compliance testing). DEC: display energy certificate (public display).

There might be issues with input errors in the EPC calculations; however, as these 20 buildings are all flagship projects with strong environmental credentials this is less likely to be the case than for average developments. Still, even knowing the comparison is not like for like, the lack of alignment is quite stark. This can be expected to generate a lot of confusion amongst clients and the general public, who might not understand the underlying assumptions and detail. Such confusion does not increase the confidence in labels on either side of the eq. As an aside, it is interesting to note that the 20 buildings is a subset of a total of 163 buildings that had met the search criteria; the remaining 143 buildings (over 80%) did not show a complete set of EPC and DEC in the Non-Domestic Energy Performance Register. These findings are in line with generic criticism of (UK) performance certificates found in literature [67]. Note that in the UK the report by CIBSE TM54 [68] attempts to reduce this gap in certification through better data collection and explanation but still cannot take away the fundamental differences between EPC and DEC certificates.

Returning to the landscape of the performance gap, a special word of caution is needed regarding the computational methods used in regulation for compliance testing. Typically, these are based on physics and have their roots in semi-dynamic (monthly) methods. However, as soon as a method is used for regulation it becomes a ranking system with mandatory thresholds rather than a predictive method. The regulatory context imposes additional demands. For instance the outcome of compliance testing calculations should never allow for parallel calculation routes that can lead to different ranking outcomes; after all whether a proposed design is acceptable to the legislator should not depend on the computation method but only on the design itself.

One also needs to be cautious with feedback-loops from measurements back to prediction methods. This link is obvious for machine learning approaches that need data for training their algorithms. However, first-principle modelling is also intertwined with measurements: first of all because empirical tool validation is based on measurement results, and secondly because for existing buildings there is the principle of calibration of models—a practice where measurements are used to establish the value of model parameters. Note that calibration is not possible in a design stage, when the actual building does not yet exist. However, in many cases there is a feed-forward loop where expertise on past projects help to establish better parameter settings in subsequent projects [48].

Within this landscape, three different types of 'energy performance gap' then can be identified:

1. Type 1: mismatch between 'first principle' energy models, and measurements undertaken on actual buildings;
2. Type 2: mismatch between machine learning approaches, and measurements from real buildings;
3. Type 3: mismatch between the energy ratings provided by compliance test methods and energy display certificates as enshrined in regulation.

6. Pilot study

A pilot study, focussing on the Roland Levinsky Building at Plymouth University, has been undertaken to explore the feasibility of investigating and quantifying the performance gap while taking into account the uncertainties in both prediction and measurement. In the performance gap classification this is a study into Type 1, comparing 'first principle' models with actual measurements.

The Roland Levinsky Building is subject of a series of computational studies [54,69] and monitoring, and evaluation is ongoing. The building itself is a complex, 9 story facility. It provides offices for about 200 members of staff, in both academic and administrative roles. Most of the staff offices are open plan. The building also contains 2 lecture theatres and a cinema, as well as 25 smaller classrooms/studio spaces for use by the approximately 6000 students of the Faculty of Arts. The offices are used throughout the year on a typical academic regime, which sees use from early in the morning (7:00 AM) to late (23:00), on workdays as well as weekends. Student use follows a system with winter (1), spring (2) and summer (3) term, with most teaching in terms 1 and 2.

A model of the Roland Levinsky Building has been created in EnergyPlus and was previously used for studies of the impact of climate change on the performance of the building; this model includes 105 zones; the Input Data File (IDF) comprises 41,614 lines. The existing model was then combined with an Uncertainty Analysis workbench [70], allowing the propagation of various uncertainties pertaining to this EnergyPlus model. For more detail, see [69]. A sample of computational results is presented in Fig. 4. Fig. 4a shows histograms for the annual gas use for the building, which is exclusively used for heating purposes. Fig. 4b shows histograms for the annual electricity use that is directly related to running the HVAC systems: this is mainly the electricity consumption of the chillers and fan energy. Fig. 4c presents histograms for the annual total electricity usage; this includes plug loads for computing, lighting, as well as the buildings cafeteria. All histograms are based on a sample of 100 EnergyPlus simulations.

A comparison with measurement data is provided in Table 1. It is obvious that the measured values sit to the far right of the tail of the distribution as obtained from the simulations. Actual measurements for electricity consumption for the HVAC system are to the other end of the distribution, showing an overprediction. Predictions of total electricity consumption are in the same order of magnitude as metered data, with measurement points fitting within the histogram produced by the simulations.

From this initial work, the following lessons have been learned:

- Defining and propagating the uncertainties that are present in a computational model is a complex task. The Uncertainty Analysis workbench used in the pilot allows the inclusion of uncertainties in a long list of relevant physical parameters, including convective heat transfer coefficients, infiltration level, temperature gradients, material properties (for instance thickness, density, heat capacity), system coefficients of performance, design levels, flow rates, urban parameters, ground temperature and many others. In total well over 100 parameters were taken into account. Yet even so a whole range of uncertainties had to be excluded since it is presently impossible to quantify these in terms of a distribution type, lower and upper bounds. Uncertainties that were thus ignored include for instance the quality of workmanship on the construction site, (un)intentional changes, construction errors, occupant behaviour, performance degradation, and operational management procedures.
- Developing a PDF for measured data is far from straightforward. While in general one must assume that there will be variation due to a range of issues such as weather conditions, occupancy, sensor accuracy and others, the reality is that over a period of two years only two values for annual energy consumption are created. These values pertain to one specific realization of the building and all uncertainty parameters only. Developing a PDF that accounts for various uncertainties would require either a longitudinal monitoring study over a longer time, or parallel studies over buildings that stem from the same design.
- Upon comparing simulation results with measurement data, there is a natural tendency to re-run the simulations, taking into account differences between initial assumptions and some data available regarding the actual situation, for instance with respect to the climate conditions or occupancy behaviour. Yet in a 'purist' approach to capturing and quantifying the performance gap one needs to take extreme care not to allow a calibration of the model to take place that would not have been possible during the design stage, where such data will never be available.

Given the large differences between the simulation results and the actual measured energy consumption data results have been investigated in a different, more traditional way, by looking at one single annual simulation and moving to a higher temporal resolution by plotting

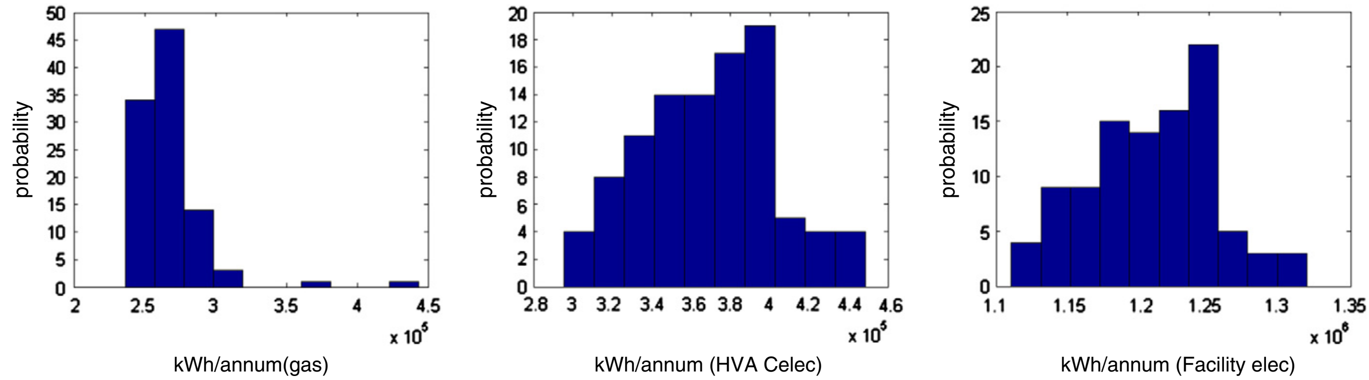

Fig. 4. Histograms for prediction of annual consumption of (a) gas for heating, (b) annual HVAC electricity use, and (c) total annual electricity use.

simulated and measured energy consumption per month in a bar chart, see Fig. 5. In this figure only gas consumption and electricity use for the HVAC systems have been plotted, in order to prevent the graph from being skewed by the (high) electrical base load. Note that the plot of monthly values for one annual simulation moves away from the probabilistic approach and looks at a single, deterministic simulation run instead.

Fig. 5 clearly demonstrates that the magnitude of the performance gap differs from month to month, and hence is a function of time. A review of Fig. 5 also shows that the trends for measurement and simulation of both electricity and gas per month are similar. On average, the values for gas use match quite closely, with simulation overpredicting about 5%; however, in terms of electricity use the simulation underpredicts significantly, in the order of 30%. As a consequence it seems logical to expect that there is an issue with the plug load in either the simulation or the actual building. However, without deeper investigation it is impossible to say whether this is an issue with the model or the actual situation: there could be incorrect assumptions in the model, there could be less efficient electric systems in the building, there could be an issue with the control settings of the real building, or any of these in combination. The only way to establish the actual reason for this difference is to delve deeper into the control settings of the BEMS, submetered data, and to conduct a survey of appliances used in the building, and to compare this with the assumptions made for the EnergyPlus model. However, once that is done the simulation efforts will include a calibration loop, whereas on the actual building one enters the realm of building energy management.

Further exploration has been conducted by means of techniques that are more common within the building energy management sector, by mapping the relation between climate as a driving factor and the facility energy use [71]. Again simulation results have been simplified to those of one single EnergyPlus run, using default values and ignoring all uncertainties. For both the simulation and the measurement years 2011 and 2012 the total energy consumption has been plotted against average monthly outdoor temperatures, with trend lines for each set of data. See Fig. 6. Note that while the building has a significant base load, as evident from Table 1, large glass facades on the North and South side of the building allow the outdoor temperature to remain the key driving factor.

Table 1
Comparison of simulated with metered data.

	Gas (kWh/annum)	Elec HVAC (kWh/annum)	Elec total (kWh/annum)
Simulated			
Median	2.31×10^5	3.69×10^5	1.21×10^6
Q1 (25% percentile)	0.22×10^5	3.46×10^5	0.12×10^6
Q2 (50% percentile)	0.23×10^5	3.69×10^5	1.21×10^6
Q3 (75% percentile)	2.43×10^5	3.93×10^5	1.24×10^6
Measured, 2011	8.81×10^5	1.20×10^5	1.19×10^6
Measured, 2012	6.68×10^5	1.29×10^5	1.21×10^6

In Fig. 6 it is clear that the measured energy use over 2011 and 2012 is consistent, with trend lines almost overlapping. The trend line from the EnergyPlus simulation however is significantly different, both in terms of gradient and intercept with the X-axis. It demonstrates that in colder years the performance gap will be wider than in warmer years. Note that Fig. 6 takes a simplified approach to the trend lines applying linear two-parameter regression only; more advanced modelling might involve the use of change-point models (piecewise linear regression models or spline fits); for further discussion see Kissock *et al.* [72]. Note that exponential, logarithmic, power or polynomial trends are not normally used in building energy management [71]. Fig. 6 demonstrates that the magnitude of the performance gap is dependent on outdoor conditions such as the average external temperature. At the comfort temperature of around 16 °C to 18 °C the trend lines cross; this is a situation where there is no need for heating or cooling. When monthly temperature drops the gap between predicted and measured energy use increases. Again, model calibration of the model would allow to establish why the simulation model predicts less impact of outdoor temperature on energy use (steeper slope for measurement lines in Fig. 5), and what specifics cause the difference in electricity consumption. However, in the typical situation of prediction at design stage and a subsequent performance gap, calibration is not possible. This is why this is not explored in further depth in this pilot study, but will be the subject of further analysis.

7. A framework for further investigation of the performance gap

Following from the literature, the pilot study, and focussing mainly on the Type 1 performance gap, the following issues need to be addressed in further research that aims to study the performance gap:

1. Efforts trying to quantify the performance gap need to accommodate the fact that the magnitude of the performance gap is dependent on time, contextual factors like climate and building use, as well as the temporal resolution at which the performance gap is studied, as demonstrated in the pilot study. It is also important to realize that each individual building will have a specific, individual performance gap. A general quantification therefore should take into account a larger sample of buildings and their respective predictions and measurements.

2. For efforts that aim to bridge the performance gap, there is a need to set a target for the equivalence of predicted and measured energy use that is sought. This target is closely related to the view one takes on the temporal and contextual resolution of the performance gap—for instance a generic percentage of agreement on annual energy use is not equivalent to one that requires a match depending on outside temperatures. It seems clear that this target needs to be in terms of

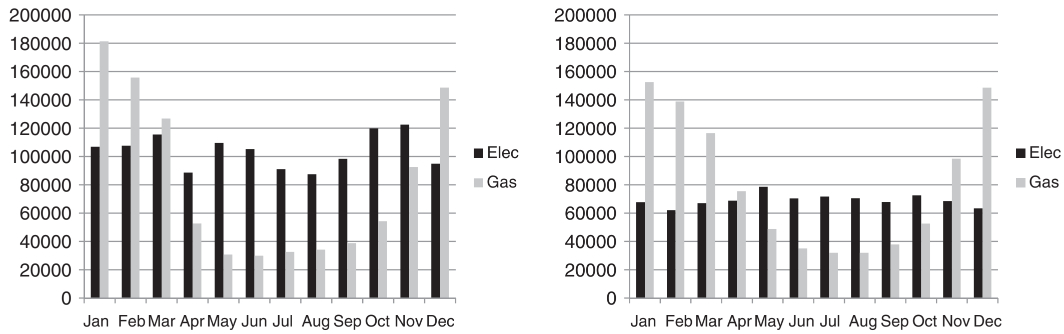

Fig. 5. Plot of monthly values for gas and electricity use. (Left) measured and (right) simulated values.

a confidence interval, but what level of confidence is feasible is presently unknown.

3. On the prediction side, there is a need to specify simulation outcomes in a manner that takes into account the uncertainties that cannot be known during the design stage, such as the impact of actual construction of a building, weather conditions, occupant behaviour, HVAC control systems, but which will impact on later measurement results. For more long-term prediction this also needs to include forecasting of technological trends (for instance the power required by office equipment), socio-economic developments (occupant density in buildings) and the effects of climate change and urban microclimatic developments.
4. On the measurement side, there is a need to capture data that pertains to the factors that cannot be fully predicted during design. In general, weather data is relatively easy to obtain. However, the other factors are much harder to capture. For instance identifying and quantifying construction errors require forensic investigation of the building at a level normally not attainable; capturing occupant behaviour requires a complex sensor network and can easily run into privacy issues. As a result, in most cases insufficient data is available.
5. In terms of comparing predictions and measurements, there is a need to develop a solid approach on how to deal with model calibration. In actual design situations there is no possibility to calibrate a model, since the real building does not yet exist. In other words, quantification of the true performance gap should be based on work that ensures strict separation between design stage models (with a priori knowledge of the building only) and calibrated models (with a posteriori knowledge through measurement of the actual building). At the same time, the later calibration of a design stage model could demonstrate the accuracy of that model within boundaries that were uncertain at the design stage.
6. Attention is needed to ensure that work on the performance gap aligns with actual building engineering practice. The more advanced probabilistic simulations at present tend to be conducted by researchers in an academic setting, yet in the end it is the performance gap between predictions done by experts in industry and monitoring data collected by facility/energy managers that needs to be addressed.
7. The performance gap that exists between machine learning approaches and measurements from real buildings needs further attention and quantification. While in this context the complexities pertaining to design and construction are removed, since machine learning is based on training an algorithm based on data from the a real building, uncertainties remain. Deep study of this area will provide a better baseline for the best possible predictions one can hope to make based on first-principle predictions.

To address these issues, a coordinated approach is needed that combines deep expertise in building energy modelling, building energy monitoring, validation and verification, as well as action research addressing application of appropriate findings to industry. Work in areas

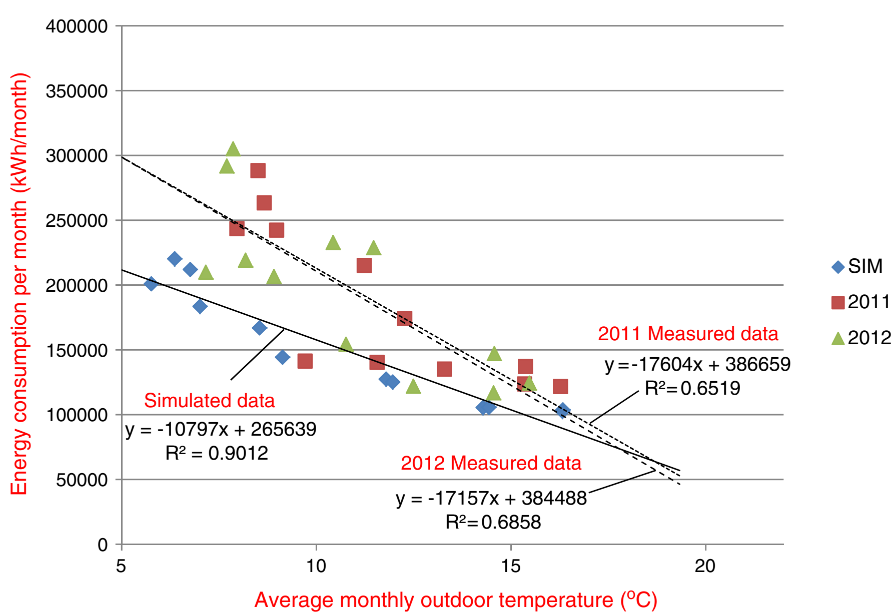

Fig. 6. Monthly energy consumption versus average outdoor temperature.

such as POE, uncertainty analysis and reliability engineering is essential for bridging the gap, but cannot bridge the energy performance gap in isolation.

Further studies into the performance gap can take three forms:

- The ideal approach is where one building design is used for the construction of a series of very identical buildings. The 1:*n* relation allows to use simulation to propagate uncertainties into a prediction model, as demonstrated in the pilot study. However, the *n* realizations of the building now allow to do parallel measurements, and to gather data that allows the construction of the histogram that could not be created for the pilot study. Unfortunately this situation occurs relatively seldom in real construction practice: most non-domestic buildings are bespoke designs, not series. Domestic buildings do appear in series, but are seldom subject of simulation as part of their engineering process; moreover, it can be expected that the impact of occupant behaviour is larger in dwellings.
- A second approach is to conduct a single building case study, taking a 1:1 relation between prediction and measurement. Here the prospect is to employ calibration (starting at the point where the pilot study was cut short) to explore, post construction, where model assumptions deviate from the actual building that was constructed, and what can be learned from this for future predictions.
- A third approach is to study a set *S* of buildings, where each building has specific prediction and measurement data. This allows a generic quantification of the occurrence and magnitude of the performance gap. However, deeper contributions to understanding the gap can only be achieved via this route if it is possible to identify sets of buildings that either have very similar designs (and hence map back to approach one) or very similar performance, in fact establishing a set with a *m*:1 relation.

8. Discussion and conclusion

Bridging the energy performance gap is crucial if the building design/engineering stage is to provide serious input into the delivery of buildings that meet their (quantified) ambitions. This paper has presented an in-depth review of literature on the subject. It has then developed a typology that discerns three types of gap: (1) between first-principle predictions and measurements, (2) between machine learning and measurements, and (3) between predictions and display certificates in legislation. In types 1 and 2, uncertainty needs to be taken into account; in type 3 this must be excluded for legislative reasons.

Through a pilot study that attempts to account for uncertainty it is demonstrated that a direct PDF-based comparison of predicted versus measured annual energy use is difficult. This is partly caused by uncertainties in the design stage that are very difficult to model and propagate in energy simulations, and partly because of a lack of data points from actual measurements. Furthermore, the pilot study shows that the performance gap changes with external conditions (example given: outdoor temperature), and with the temporal resolution of the energy measure in use.

A coordinated approach is needed to better understand, and ultimately bridge, the energy performance gap. Further collection of evidence of the performance gap is needed, covering both breadth and depth, providing data that represents the status in the construction industry as well as providing deep insights into individual cases. Validation and verification need to be applied to energy prediction methods, increasing the adequacy for intended use of predictions, ensuring that the models used are state-of-the-art and continue to progress. At the same time the data used as input for predictions needs to be improved, through both solid monitoring work (capturing the current energy use and the related driving factors in today's buildings) as well as efforts aimed at forecasting complex issues such as technology developments, socio-economic trends and long-term weather forecasts. Yet the energy performance gap can only be addressed by changing the current practices in building engineering practice.

Three key approaches can be taken for further studies. The study of various buildings based on the same design is the most promising option, allowing the deepest exploration of the performance gap. An alternative is a deep case study, including model calibration, which explores and quantifies the factors that cause the performance gap in one specific instance. Furthermore, one could look at larger sets of buildings, hoping to identify homogenous sub-sets with very similar performance.

It is important to remember that energy efficiency is only one performance aspect of buildings. Once predicted and measured energy use are adequately matched, further work will be needed to address performance gaps in areas like thermal comfort, indoor air quality and others.

Acknowledgments

The work by described in this paper has been funded through the Royal Academy of Engineering and Leverhulme Senior Research Fellowship, Reference Number 10226/48. The author wishes to thank Godfried Augenbroe and Yuming Sun for deep discussions on the subject of this paper, which have significantly shaped the work reported, and Darren Pearson of C3Resources who was instrumental in obtaining AMR data for the pilot study. ModelCenter® has been used in the described work under academic license from Phoenix Integration.

References

[1] C. Turner, M. Frankel, *Energy performance of LEED for new construction buildings* (Final Report), New Buildings Institute, While Salmon (WA), 2008.
[2] Carbon Trust, Closing the Gap: Lessons Learned on Realising the Potential of Low Carbon Building Design, Carbon Trust, London, 2011.
[3] Zero Carbon Hub, A Review of the Modelling Tools and Assumptions: Topic 4, Closing the Gap between Designed and Built Performance, Zero Carbon Hub, London, 2010.
[4] C. Menezes, A. Cripps, D. Bouchlaghem, R. Buswell, Predicted vs. actual energy performance of non-domestic buildings: using post-occupancy evaluation data to reduce the performance gap, Appl. Energy 97 (2012) 355–364.
[5] N. Almeida, V. Souza, L. Alves Dias, F. Branco, A framework for combining risk management and performance-based building approaches, Build. Res. Inf. 38 (2) (2010) 157–174.
[6] X. Zhang, H. Gao, Optimal performance-based building facility management, Comput. Aided Civ. Infrastruct. Eng. 25 (2010) 269–284.
[7] W. Oberkampf, C. Roy, Verification and Validation in Scientific Computing, Cambridge University Press, Cambridge, 2010.
[8] J. Clarke, Energy Simulation in Building Design, Butterworth-Heinemann, Oxford, 2001.
[9] P. Strachan, G. Kokogiannakis, I. Macdonald, History and development of validation with the ESP-r simulation program, Build. Environ. 43 (2008) 601–609.
[10] H. Manz, P. Loutzenhiser, T. Frank, P. Strachan, R. Bundi, G. Maxwell, Series of experiments for empirical validation of solar gain modeling in building energy simulation codes - experimental setup, test cell characterization, specifications and uncertainty analysis, Build. Environ. 41 (2006) 1784–1797.
[11] T. Reddy, I. Maor, C. Panjapornporn, Calibrating detailed building energy simulation programs with measured data—Part I: general methodology, HVAC&R Res. 13 (2) (2007) 221–241.
[12] T. Reddy, I. Maor, C. Panjapornporn, Calibrating detailed building energy simulation programs with measured data—Part II: application to three case study buildings, HVAC&R Res. 13 (2) (2007) 243–265.
[13] T. Williamson, Predicting building performance: the ethics of computer simulation, Build. Res. Inf. 38 (4) (2010) 401–410.
[14] J. Shabi, Y. Reich, Developing an analytical model for planning systems verification, validation, Adv. Eng. Inform. 26 (2012) 429–438.
[15] Efficiency Valuation Organisation, , Website at www.evo-world.org .
[16] S. Ginestet, D. Marchio, Retro and on-going commissioning tool applied to an existing building: operability and results of IPMVP, Energy 35 (2010) 1717–1723.
[17] A. Ahmed, N. Korres, J. Ploennings, H. Elhadi, K. Menzel, Mining building performance data for energy-efficient operation, Adv. Eng. Inform. 25 (2011) (2011) 341–354.
[18] L. Norford, R. Socolow, E. Hsieh, G. Spadaro, Two-to-one discrepancy between measured and predicted performance of a 'low-energy' office building: insights from a reconciliation based on the DOE-2 model, Energy Build. 21 (2) (1994) 121–131.
[19] B. Bordass, R. Cohen, M. Standeven, A. Leaman, Assessing building performance in use 3: energy performance of probe buildings, Build. Res. Inf. 29 (2) (2001) 114–128.
[20] J. Scofield, Early performance of a green academic building, ASHRAE Trans. 108 (2002) 1214–1230.
[21] CarbonBuzz, Website at www.carbonbuzz.org (Last accessed 14 June 2013).

[22] M. Stoppel, F. Leite, Evaluating building energy model performance of LEED buildings: identifying potential sources of error through aggregate analysis, Energy Build. (2013), http://dx.doi.org/10.1016/j.enbuild.2013.04.016.
[23] P. Sanguinetti, Integrated performance framework to guide façade retrofit, (doctoral dissertation) Georgia Institute of Technology, College of Architecture, Atlanta, 2012.
[24] G. Newsham, S. Mancini, B. Birt, Do LEED-certified buildings save energy? Yes, but…, Energy Build. 41 (2009) 897–905.
[25] Morant, *The performance gap—non domestic buildings* (Report CEW1005), AECOM/Constructing Excellence Wales, Cardiff, 2012.
[26] A. Korjenic, T. Bednar, Validation and evaluation of total energy use in office buildings: a case study, Autom. Constr. 23 (2012) 64–70.
[27] A. Dasgupta, A. Prodromou, D. Mumovic, Operational versus designed performance of low carbon schools in England: bridging a credibility gap, HVAC&R Res. 18 (1–2) (2012) 37–50.
[28] G. Newsham, B. Birt, C. Arsenault, L. Thompson, J. Veitch, S. Mancini, A. Galasiu, I. Macdonald, G. Burns, Do green buildings outperform conventional buildings? *Indoor Environment and energy performance in North American offices* (Report RR-329), , Natural Research Council Canada, 2012.
[29] P. America, P. van de Laar, G. Muller, Experiences in evolvability research, Adv. Eng. Inform. 26 (2012) 478–486.
[30] T. Dwyer, Knowledge is power: benchmarking and prediction of building energy consumption, Build. Serv. Eng. Res. Technol. 34 (1) (2013) 5–7.
[31] S. Sequeira, R. Fay, J. Sargison, F. Soriano, A preliminary analysis of subfloor ventilation data: bridging the gap between theory and experiment, Archit. Sci. Rev. 53 (3) (2010) 315–322.
[32] M. Bell, J. Wingfield, D. Miles-Shenton, J. Seavers, Low Carbon Housing: Lessons from Elm Tree Mews, Joseph Rowntree Foundation, York, 2010.
[33] National Measurement Network, The Building Performance Gap—Closing it through Better Measurement, National Physical Laboratory, Teddington, 2012.
[34] R. Bunn, M. Way, Soft Landings, Building Services Research and Information Association and Usable Building Trust, Bracknell, 2010.
[35] F. Haldi, D. Robinson, On the behaviour and adaptation of office occupants, Build. Environ. 43 (2008) 2163–2177.
[36] E. Ryan, T. Sanquist, Validation of building energy modeling tools under idealized and realistic conditions, Energy Build. 47 (2012) 375–382.
[37] Z. Yu, F. Haghighat, B. Fung, E. Morofsky, H. Yoshino, A methodology for identifying and improving occupant behaviour in residential buildings, Energy 36 (2011) 6596–6608.
[38] M. Chung, H. Park, Development of a software package for community energy system assessment—Part I: building a load estimator, Energy 35 (2010) 2767–2776.
[39] A. Molin, P. Rohdin, B. Moshfegh, Investigation of energy performance of newly built low-energy buildings in Sweden, Energy Build. 43 (2011) 2822–2831.
[40] W. Lee, K. Lee, Benchmarking the performance of building energy management using data envelopment analysis, Appl. Therm. Eng. 29 (2009) 3269–3273.
[41] National Measurement Network, Building Performance Measurement, National Physical Laboratory, Teddington, 2012.
[42] National Measurement Network, Metrology Challenges for Infrastructure, National Physical Laboratory, Teddington, 2013.
[43] D. O'Sullivan, M. Keane, D. Kelliher, R. Hitchcock, Improving building operation by tracking performance metrics throughout the building lifecycle (BLC), Energy Build. 36 (2004) 1075–1090.
[44] B. Tofield, *Delivering a Low-Energy Building: Making Quality Common Place.* (Build with CaRe report), University of East Anglia, Norwich, 2012.
[45] B. Williamson, The Gap between Design and Build: Construction Compliance towards 2020 in Scotland, Edinburgh Napier University, Edinburgh, 2012.
[46] P. André, B. Georges, J. Lebrun, V. Lemort, From model validation to production of reference simulations: how to increase reliability and applicability of building and HVAC simulation models, Build. Serv. Eng. Res. Technol. 29 (1) (2008) 61–72.
[47] H. Koziolek, Performance evaluation of component-based software systems: a survey, Perform. Eval. 67 (2010) 634–658.
[48] P. Raftery, M. Keane, J. O'Donnell, Calibrating whole building energy models: an evidence-based methodology, Energy Build. 43 (2011) 2356–2364.
[49] W. Shen, Q. Hao, H. Mak, J. Neelamkavil, H. Xie, J. Dickinson, R. Thomas, A. Pardasani, H. Xue, Systems integration and collaboration in architecture, engineering, construction, and facilities management: a review, Adv. Eng. Inform. 24 (2010) 196–207.
[50] E. Curry, J. O'Donnell, E. Corry, S. Hasan, M. Keane, S. O'Riain, Linking building data in the cloud: integrating cross-domain building data using linked data, Adv. Eng. Inform. 27 (2013) 206–219.
[51] S. Wei, R. Jones, S. Goodhew, P. de Wilde, Occupant space heating behaviour in a simulation-intervention loop: a theoretical framework, *Building Simulation '13, 13th International IBPSA Conference*, Chambéry, France, August 25-30 2013, 2013.
[52] A. Mahdavi, People in building performance simulation, in: Roberts Hensen (Ed.), Building Performance Simulation for Design and Operation, Spon Press, London, 2011, pp. 56–83.
[53] E. Maldonado, P. Wouters, A. Panek, Concerted action: Supporting Transposition and Implementation of the Directive 2002/91/EC CA EPBD (2005-2007), Intelligent Energy Europe, Brussels, 2008.
[54] W. Tian, P. de Wilde, Uncertainty and sensitivity analysis of the performance of an air-conditioned campus building in the UK under probabilistic climate projections, Autom. Constr. 20 (2011) 1096–1109.
[55] C. Underwood, F. Yik, Modelling Methods for Energy in Buildings, Blackwell, Oxford, 2004.
[56] J. Hensen, R. Lamberts (Eds.), Building Performance Simulation for Design and Operation, Spon Press, London, 2011.
[57] A. Tsanas, A. Xifara, Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools, Energy Build. 49 (2012) 560–567.
[58] A. Kusiak, G. Xu, Modeling and optimization of HVAC systems using a dynamic neural network, Energy 42 (2012) 241–250.
[59] S. Cho, W. Kim, C. Tae, M. Zaheeruddin, Effect of length of measurement period on accuracy of predicted annual heating energy consumption of buildings, Energy Convers. Manag. 45 (2004) 2867–2878.
[60] J. Jin, J. Jeong, Thermal characteristic prediction models for a free-form building in various climate zones, Energy 50 (2013) 468–476.
[61] P. de Wilde, C. Martinez-Ortiz, D. Pearson, I. Beynon, M. Beck, N. Barlow, Building simulation approaches for the training of automated data analysis tools in building energy management, Adv. Eng. Inform. (2013), http://dx.doi.org/10.1016/j.aei.2013.05.001.
[62] H. Friedman, E. Crowe, E. Sibley, M. Effinger, The Building Performance Tracking Handbook, California Commissioning Collaborative, 2011.
[63] D. Harris, *Key Performance Indicators—Field Metering Study and Energy Performance Feedback* (Final Report), New Buildings Institute, While Salmon (WA), 2012.
[64] D. Djurdjanovic, J. Lee, J. Ni, Watchdog Agent—an infotronics-based prognostics approach for product performance degradation assessment and prediction, Adv. Eng. Inform. 17 (2003) 109–125.
[65] EU DIRECTIVE 2004/22/EC, Directive 2004/22/EC of the European Parliament and of the Council on measuring instruments, Off. J. Eur. Union 135 (2004) 1–80.
[66] W. Chung, Review of building energy-use performance benchmarking methodologies, Appl. Energy 88 (2011) 1470–1479.
[67] S. Kelly, D. Crawford-Brown, M. Pollitt, Building performance evaluation and certification in the UK: is SAP fit for purpose? Renew. Sustain. Energy Rev. 16 (2012) 6861–6878.
[68] CIBSE TM54, Evaluating operational energy performance of buildings at the design stage, Chartered Institution of Building Services Engineers, Lavenham, 2013.
[69] P. de Wilde, Y. Sun, G. Augenbroe, Quantifying the performance gap—a probabilistic attempt. *EG-ICE 2013*, *Conference on Intelligent Computing in Engineering*, Vienna, Austria, July 1-3 2013, 2013.
[70] B. Lee, Y. Sun, G. Augenbroe, C. Paredic, Towards better prediction of building performance: a workbench to propagate uncertainties through building simulation models, *Building Simulation '13, 13th International IBPSA Conference*, Chambéry, France, August 25-30 2013, 2013.
[71] V. Vesma, Energy Management Principles and Practice: A Companion to BS EN 16001:2009, British Standards Institution, London, 2009.
[72] J. Kissock, J. Haberl, D. Claridge, Development of a toolkit for calculating linear, change-point linear and multiple-linear inverse building energy analysis models, ASRAE Research Project 1050-RP, 2002.