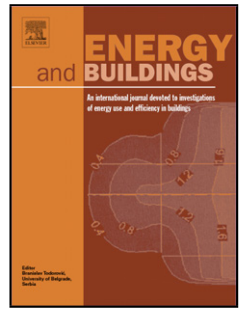
Accepted Manuscript



Title: Evaluating energy performance in non-domestic buildings: A review

Author: E.H. Borgstein R. Lamberts J.L.M. Hensen

PII: S0378-7788(16)30614-4
DOI: http://dx.doi.org/doi:10.1016/j.enbuild.2016.07.018
Reference: ENB 6849

To appear in: *ENB*

Received date: 21-5-2016
Revised date: 6-7-2016
Accepted date: 7-7-2016

Please cite this article as: E.H.Borgstein, R.Lamberts, J.L.M.Hensen, Evaluating energy performance in non-domestic buildings: A review, Energy and Buildings http://dx.doi.org/10.1016/j.enbuild.2016.07.018

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Evaluating energy performance in non-domestic buildings: a review

E.H. Borgstein[1*]edward@borgstein.org, R. Lamberts[1,2], J. L. M. Hensen[1]

[1] Laboratory for Energy Efficiency in Buildings (LabEEE), Department of Civil Engineering, Federal University of Santa Catarina, Florianópolis, SC 88040-900, Brazil

[2] Building Physics and Systems, Eindhoven University of Technology, P.O. Box 513, NL-5600 MB, Eindhoven, The Netherlands

Corresponding author. Mitsidi Projetos, Rua Pedro Taques 129, Consolação, São Paulo, SP, Brazil; Tel: +55 11 98359 4865.

Abstract

Evaluation methods can be used to determine what constitutes good energy performance in a building. With an increasing focus on energy use of buildings worldwide, this evaluation assumes a fundamental importance. This paper provides a comprehensive review of the available methods for analysing, classifying, benchmarking, rating and evaluating energy performance in non-domestic buildings.

Methodologies are grouped in five categories: engineering calculations, simulation, statistical methods, machine learning and other methods. Techniques for evaluating buildings are described, their principal applications are shown and limitations are identified. The use of performance evaluation in energy efficiency programmes and standards is mapped.

There is a need to further develop interactions between the main modelling techniques to produce simple, robust and validated models. Also, evaluation techniques must be developed to consider comfort or service provision in the buildings as a factor in energy performance.

ACCEPTED MANUSCRIPT

Introduction

Aims and scope

Buildings are responsible for a third of global final energy consumption and similar levels of CO_2 emissions [1]; as such the energy performance of buildings is an area that has assumed increasing importance worldwide. The International Energy Agency describes energy efficiency as the "first fuel" [2], with greater importance than any generation-side technology, and studies have shown that the greatest potential for energy efficiency is in the buildings sector [3].

However, as the built environment becomes increasingly complex, with multiple uses, larger buildings and higher demand for services such as thermal comfort and data processing [4], the analysis and evaluation of performance in buildings becomes a non-trivial problem. Since the introduction of the first building regulations aimed at energy conservation in the 1970s, and the increased focus on this area in the late 1990s and early 2000s [5], a wide range of methodologies has been developed for predicting, analysing and evaluating the energy performance of buildings. Several of these methodologies have been integrated into national or public evaluation programmes for new and existing buildings, while the past few years have seen a rapid growth in the development and application of new methodologies for evaluating and classifying energy performance. Public policies promoting energy efficiency in buildings have become increasingly common [6].

This paper provides a review of the prediction, evaluation, analysis and classification of building energy performance, identifying the most widely used methodologies and grouping them by analysis type, in order to propose standard terminology, identify gaps in the literature and inform future research. Only methods relevant to non-domestic buildings are included in this review, in the interest of simplicity and improving international comparisons (domestic energy loads vary more widely between countries and climate zones, and may require very different evaluation methodologies).

Energy design advice and methodologies for improving energy efficiency in buildings were reviewed extensively by Ruparathna et al [7], as well as being the subject of numerous publications and books (for example: [8], [9]). Whilst the aim of energy evaluation of buildings is generally linked to attempts to improve their performance, this paper explicitly avoids discussions of energy performance improvement through retrofit or new construction, focusing only on the evaluation of performance. In addition, this review is based only on energy performance, and does not directly discuss carbon emissions, which vary widely depending on grid carbon factors in different countries.

Defining building energy performance

The core question of this paper can be very simply phrased as: "What constitutes good energy performance in a building?"

ACCEPTED MANUSCRIPT

An extensive discussion of building rating performance [10] states clearly the difference between a low energy building, which may use energy inefficiently, and a high energy building which may use very efficient systems to provide a higher level of service. The paper proposes that a building performance rating scheme should include three elements:

The building must be energy efficient, through its design, systems and technologies;

The building must supply amenities and features typical for its typology; and

The building must be low energy, in other words it must be operated in such a manner as to be efficient.

The first and third elements of Olofsson's framework were developed in more detail in the EuroProsper and EPLabel projects [11], [12], which used the terminology Asset Rating to refer to the calculated, potential energy performance (based on the building design and systems installed) and Operational Rating to refer to the measured, in-use energy performance. This was formalised in ISO 16346, which defines Calculated Ratings and Measured Ratings [13]. Calculated ratings are further divided by their use, as shown in Table 1.

Rating type	Name	Input data			Utility or purpose
		Use	Climate	Building	
Calculated	Design	Standard	Standard	Design	Building permit, certificate under conditions
	Standard	Standard	Standard	Actual	Energy performance certificate, regulation
	Tailored	Depending on purpose		Actual	Optimisation, validation, retrofit planning
Measured	Operational	Actual	Actual	Actual	Energy performance certificate, regulation

Table 1 - Types of building energy ratings, from ISO 16346:2013 [13]

In order to evaluate energy performance, it is necessary to compare the calculated or measured performance of a building to some reference value or framework. These values may represent characteristics of the buildings (such as wall U-values) or the energy consumption of specific building systems, but it is increasingly common to carry out evaluations based on normalised whole-building energy consumption. In these cases, an Energy Performance Indicator (EPI) or Energy Use Intensity (EUI) is defined, usually as the energy consumption normalised by floor area, kWh/m²/annum (see for example the EnergyStar methodology [14]).

There are various ways to develop reference values (or baselines) for comparison with the building energy consumption, which are summarised in Table 2. Each type of baseline is dependent on a different technique for its development. These techniques, in turn, are based on the technical methodologies described in detail in the following sections of this paper.

Reference values for performance comparison (baseline)	Techniques for development of baselines	Technical methodologies

ACCEPTED MANUSCRIPT

comparison (baseline)	development of baselines	
Historical energy performance	Measurement and verification (M&V)	Statistical methods, Engineering calculations
Typical performance of similar buildings (empirical)	Stock models or market surveys	Statistical methods
Expected energy performance	Building-specific modelling at design stage	Simulation
Potential energy performance	Building-specific energy audits or reviews	Engineering calculations, Simulation
Required performance, norms or standards	Regulatory methodology	Engineering calculations, Simulation

Table 2 - Comparison baselines for energy performance evaluation

The IEA Annex 53 project recognised the discrepancy between designed and real energy use in buildings and aimed to develop analysis and evaluation methods for total energy use in buildings [1]. The analysis starts from the principle that six factors affect energy consumption in buildings:

Climate

Building envelope

Building systems

Operations & maintenance

Occupant behaviour

Indoor environmental conditions

It is worth noting that the environmental performance and comfort of a building are highly relevant to its overall performance. Ideally, a performance evaluation should clearly identify which of these factors it aims to evaluate. For example, a regulatory performance requirement based on a calculated rating will assume standardised (constant) climate, O&M, behaviour and indoor environmental conditions, in order to assess the efficiency of the building envelope and systems.

The ISO Standard 16436 discusses in more detail the different weightings that can be applied to energy ratings in order to standardise them. This is often important as buildings are likely to use more than one energy carrier; in these cases, it is necessary to use a common expression

ACCEPTED MANUSCRIPT

of energy carriers to aggregate the amounts, which may have different costs, units, exergy or environmental impact. Thus, the final energy consumed in a building may be weighted by a primary energy rating, by carbon dioxide emissions factors or even by energy cost. As this depends heavily on local conditions and the aims of specific policy, energy weightings are not discussed further in this paper, which considers the final, delivered energy to buildings.

Building energy benchmarking

In the evaluation of building energy performance, the term "benchmarking" is often used to refer to the comparison of building energy performance with that of similar buildings [15]–[17], although the term benchmark more generally applies to any standardised performance level that serves as a basis for evaluation or comparison [18]. The term "empirical benchmark" has been used to apply specifically to the comparison of actual (measured) building energy performance against the broader building market [19]. According to Wang and colleagues, benchmarking is "a simple method to inform decision makers with a relative energy performance level by comparing the whole- building energy performance index of the assessed building with pre-set benchmarks" [20].

In this context, each of the baselines defined in Table 2 can be considered to be a benchmark, when it is appropriately calculated. Thus, the process of building energy benchmarking involves three steps:

Identify the most appropriate baseline;

Calculate the energy performance of the building in accordance with these parameters;

Compare the building performance with the benchmark levels, to classify a performance level.

Classification of performance evaluation methodologies

Several sources have addressed the classification of building performance evaluation methodologies. A recent two-part study separates top-down (deductive) benchmarking methodologies involving statistical approaches [21] from bottom-up (inductive) methodologies which use aggregated end-use calculations or building simulation [22]. Other studies have categorised benchmarking through black, grey or white box methodologies [23], where these are defined as follows:

Black box models are purely statistical, with little information required on each building.

Grey box models mix limited building physics with statistical methodologies.

White box models are based on building physics, and are highly dependent on user inputs.

The ASHRAE Fundamentals Handbook [24] separates energy estimating and modelling methods into "forward modelling", used for building design and optimisation, and "data-driven modelling", used for modelling existing buildings and establishing baselines. The nineteen

ACCEPTED MANUSCRIPT

analysis methods identified include both dynamic and static methodologies, and are further divided into black box, grey box and calibrated simulation methodologies. For Wang and colleagues, energy quantification methods can be divided into calculation-based (including dynamic simulation and steady-state methods), measurement-based (based on monitoring or energy bills) and hybrid (calibrated simulation or dynamic inverse modelling). These energy quantification methods, in turn, are used to compare performance to benchmarks developed through statistical or calculated methodologies [20]. Another review on building benchmarking, rating and classification [15] discusses model-based and empirical benchmarking, identifying many of the principal methodologies adopted and discussing in detail the use of benchmarks as the foundation of rating systems and classification methodologies. In this review, as in previous papers [25], building evaluation methodologies are divided into four categories: "statistical analysis benchmarking (also known as regression model-based), points-based rating systems, simulation model-based benchmarking, and hierarchical and end-use metrics." The energy ratings discussed are explicitly linked to certification schemes by Pérez-Lombard [17], extending the concept of using an energy benchmark as a baseline for classification and certification systems.

It is clear that the field of building energy performance evaluation has adopted several different terminologies or classifications that essentially describe the same types of analysis. For the purposes of this paper, a simple but clear terminology has been adopted, dividing energy performance evaluation methodologies into engineering calculations, simulation, statistical methods and machine learning. A rough equivalence with the diverse terminology adopted in the literature is shown in Table 3.

Sections 2, 0, 0 and 0 of this paper will explore each of these methods in further detail, while section 0 will briefly mention further evaluation methodologies that do not fit within this framework and section 0 will describe the implementation of evaluations in public rating systems.

Engineering calculation	Simulation	Statistical methods	Machine learning	Sources
White box (sometimes grey box)	White-box	Black box (sometimes grey box)		[23], [26], [27]
Bottom-up		Top-down		[21], [22]
Forward modelling		Data-driven modelling or inverse modelling		[24], [28], [29]
Benchmarking				[15]
Simple calculation	Detailed calculation	n/a		[30]

ACCEPTED MANUSCRIPT

Expert knowledge	Simulation	Aggregated statistics	n/a	[10]
Calculated benchmark		Statistical benchmark		[20]
Simplified engineering methods	Elaborate engineering methods	Statistical methods	Neural Networks, Support Vector Machines	[31]

Table 3 - Comparison of different terminologies for describing building performance evaluation methodologies

The selection of the most appropriate analysis method will generally depend on the following factors: accuracy, sensitivity, versatility, speed, cost, reproducibility and ease of use [24].

ACCEPTED MANUSCRIPT

Engineering calculation methodologies

A wide range of simplified models is used for predicting building energy performance through rapid calculation. These methods are often used at early design stage of buildings, but are also useful for calculating estimated energy performance and predicting the impacts of energy conservation measures, for example during energy audits.

These simplified models have benefits over more complex simulation models, including ease of use, clear relation to physical parameters and speed of calculation [27]. A description by Raslan et. al. clearly makes the distinction between these methods and full building simulation:

"In general, a calculation involves the implementation of a simplified mathematical equation. Tools falling under this category generally use steady state models that average variables over a long period in which all building parameters are fixed. Quasi steady state models, which account for the effect of some transient parameters such as weather, may also be used in an aim to establish predictions about the building performance, which relate energy use to these input variables. Computer programs that use either of these simple methods do not aim to take all complex interactions of the building into account and therefore do not attempt to 'simulate' it. These are often referred to as calculation tools." [32]

In the context of this paper, engineering calculations are considered to be those that are fully based on building physics and are generally applied to single buildings (or complexes of buildings served by a single heating or cooling source). Hence, simplified models, which are partially dependent on machine learning, are discussed in the following sections.

International standards

The International Standard ISO 16346 [13] defines the processes for calculation methodologies, using the calculation methods in ISO 13790 [33] as a basis for calculating heating and cooling loads, before citing holistic and simplified calculation approaches for whole-building energy calculations (the holistic method uses iteration to arrive at more accurate results). European norms are cited for specific calculation methodologies for heating, cooling and ventilation systems (EN 15136, 15243 and 15241 respectively).

HVAC system calculations

The HVAC loads in a building are usually the most difficult to calculate with simple methodologies, as they have high dependence on external factors and generally are dynamically controlled. Their performance can be modelled as a function of the heating and cooling loads, combined with an estimate of system efficiency for the provision of heating or cooling.

ACCEPTED MANUSCRIPT

ISO 13790 identifies and describes the two principal simplified calculation methodologies for heating and cooling loads. The standard lays out calculations for a fully prescribed monthly quasi-steady state calculation method and for a fully prescribed simple hourly calculation method. These methods start from energy balances in the building and can effectively model heating loads but do not account for thermal mass, which has a significant impact on building cooling loads and a smaller impact on heating loads. Dynamic models or hourly methods consider thermal inertia by the use of a thermal capacitance in an equivalent Resistor-Capacitor (RC) model, while the quasi-steady state models use "utilisation factors" for heat losses or gains. The methods described in ISO 13790 have been shown to give similar results to full dynamic simulations [34], although in this case they required careful calibration of utilisation factors.

The equivalent RC method models building heat gains and losses through a network of equivalent resistors and capacitors, linking internal with external environmental conditions. Resistors represent barriers to heat exchange (insulation or thermal resistance), while capacitors represent thermal mass or thermal capacitance. This analogy is effective for directly modelling building physics, and has relatively high computational efficiency [26]. The simplest models only consider temperature, but an examination of input parameters reveals that a model for a complete year should also include relative humidity, global solar radiation and cloudiness [27].

Quasi-steady state models assume monthly average temperatures, and work best when indoor temperatures are held constant. A modelling tool based on monthly methods was selected as the most appropriate tool for energy auditing in the European iNSPiRe project, based on its ease of use and relative accuracy of the results [35]. The UK's National Calculation Methodology for Non-Domestic Buildings uses the Simplified Building Energy Model (SBEM), a steady-state monthly calculation methodology that considers building geometry, construction, use and HVAC and lighting equipment [36].

Simple models can be used to carry out annual energy consumption estimates by characterising climate impacts on heating and cooling through degree-days or bin methods. Degree-days are appropriate where building use and system efficiency are constant, and measure temperature variation from a balance point temperature. Different balance points are used for heating and cooling, and for areas with significant latent loads it may be appropriate to use wet bulb temperatures for cooling [24], [30], [37]. Where there is significant variation in heat loss coefficients, system efficiency or balance point temperatures, it is more appropriate to use bin methods. These involve the calculation of system consumption for several values of outdoor temperature, and the yearly consumption is estimated by multiplying the consumption in a certain condition by the annual number of hours in that bin or range of temperature. Modified bin methods can account for off-design conditions based on diversified load profiles [24], [30].

An alternative methodology for estimating heating and cooling loads which considers thermal mass is the "Admittance Method", which is a simple cyclical model, assuming that loads can be represented by the sum of a steady-state component and a sine wave. This model has been shown to compare well with other simple calculation methods and full simulation [38].

ACCEPTED MANUSCRIPT

As HVAC systems become increasingly complex, the usefulness of bin and degree-day methods is limited to simple first-stage analyses. The iSERV project [39] has started to gather and publish further information on system-level HVAC energy performance, based on measured data and a standardised data collection methodology. This should allow the development of specific HVAC benchmarks, which will in turn allow for more effective calculation and estimation of HVAC system energy performance [40].

Aggregated end-use

At the simplest level, total energy consumption can be calculated as the sum of the estimated consumption of all of the energy-consuming systems [41]. This aggregated method has been cited as one of the most accurate methods for predicting energy performance, and has the advantage of permitting the development and use of system level benchmarks [22]. In existing buildings, these calculations are compared to benchmark levels and total measured energy consumption in the building in order to give useful information and evaluations of efficiency levels, in accordance with the CIBSE TM22 methodology [42].

Once the HVAC system energy consumption has been estimated (as above), further calculation methods for other end-uses can be identified. These are often highly dependent on prior knowledge, studies and rules of thumb, but often give good performance estimates in areas such as plug loads [43] or elevators and escalators [44]. Energy auditing methodologies generally discuss the calculation of end-use energy consumption as a fundamental stage in an energy evaluation of an existing building [41], [45], [46], while a detailed research project by CIBSE developed a guide to estimating aggregated energy consumption levels for buildings at the design stage, including uncertainty estimates [47].

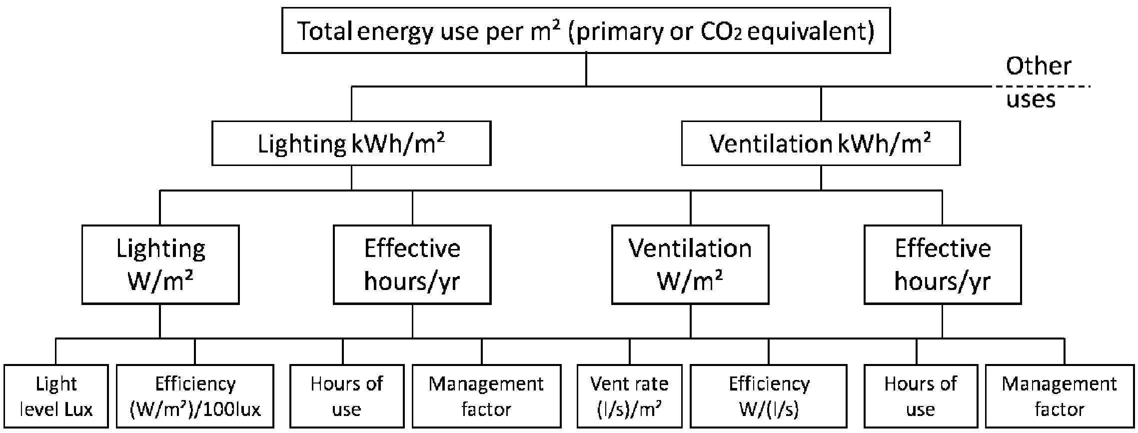


Figure 1 - Illustration of aggregated end-use calculation [41]

Although aggregated end-use calculations are clearly effective for carrying out energy audits and identifying potential energy saving measures in buildings, they show certain limitations when used for whole-building evaluations. The HVAC loads must generally be calculated separately (as above), and the calculations are most useful when applied in conjunction with system-level or empirical benchmarks [22].

ACCEPTED MANUSCRIPT

However, these methodologies have been used for developing sector benchmarks (known as "parametric benchmarks"), using the results of energy audits and aggregated end-use calculations in small samples of buildings. This methodology is demonstrated by the ECON19 office benchmarks in the UK [48], and was used to develop the majority of the empirical benchmarks referenced in CIBSE's TM46 document, which publishes benchmarks for the main non-domestic building typologies in the UK [49].

Performance parameters

A further step in performance evaluation is to use engineering calculations to define specific performance parameters, which can be used to evaluate the energy performance of the building. This has been done in several different ways by different authors.

Escrivá-Escrivá describes a methodology for collecting building energy consumption data through energy audits and in-situ measurements in order to calculate performance indices. Six separate performance indices are defined, based on measures of energy consumption and occupation, and an energy rating factor used to rank performance levels [50].

Empirical benchmarks (comparing to typical performance of similar buildings) are limited in their application, as there may be a very small sample of similar buildings to compare with, or all the buildings in the sample may have low levels of efficiency. As such, a bottom-up method based on engineering calculations and aggregated end-uses can be an important alternative method. Their development and application to laboratory buildings has been demonstrated [51]. Fumo used simulation to develop predetermined coefficients for specific building typologies, which could be used alongside monthly electricity and fuel bills to carry out energy analysis and evaluations [52]. Pérez-Lombard developed specific indicators for energy efficiency in HVAC systems, based on measured performance levels, and linking energy demand to physical systems and services provided [53]. Another study has applied decomposition methods to evaluate the impacts of different factors on HVAC efficiency levels, developing methods that can be used to quickly evaluate energy intensity levels and potential improvements [54].

Yan et. al. take this type of analysis further by identifying the lack of information available for many buildings, and the need for different levels of energy analysis for different requirements (as shown in Figure 2). They overcome this difficulty by proposing and demonstrating a novel methodology for developing energy performance parameters for buildings using monthly energy bills and a few in-situ measurements. This allows the development of a "relative performance factor", which demonstrates the difference between actual and potential energy performance, and estimate potential energy performance improvement [55].

ACCEPTED MANUSCRIPT

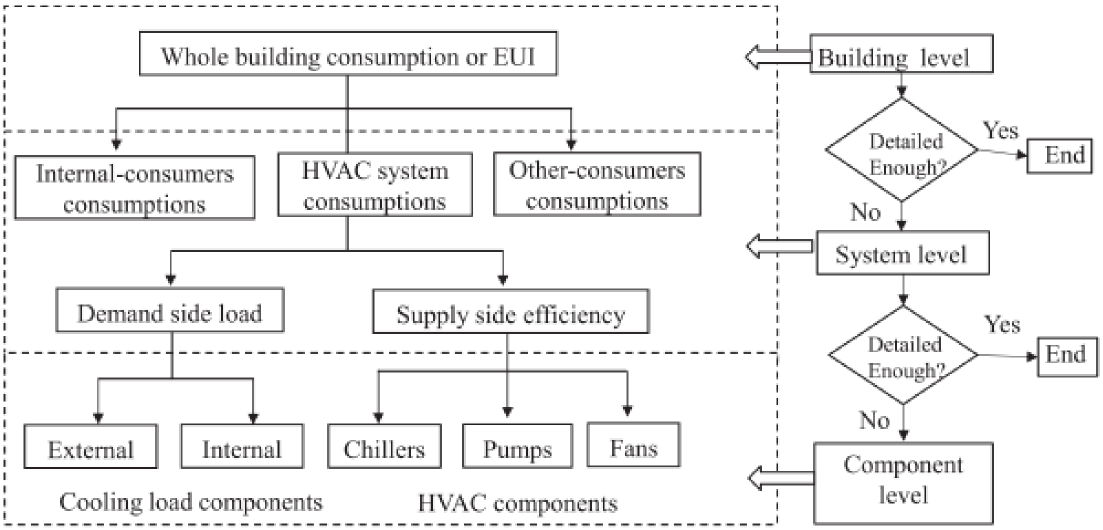


Figure 2 – Top-down approach for multi-level diagnosis (from [55])

A final approach uses data on the building and pre-calculated indices to calculate bottom-up energy consumption levels, developing performance predictions which could then be compared to real energy consumption to evaluate performance. This has been carried out on the basis of space usage type, considering room-scale profiles [56] and on the basis of occupancy levels and time-varying occupancy rates [57].

Limitations and applications

The calculation tools described in this section do not replicate the dynamic processes of full simulation methodologies, and as such have lower levels of accuracy, and may not be precise in predicting building performance. In order to use these methods for the evaluation of performance, it is often necessary to define "typical" parameters or performance levels, which requires some sort of complementary data collection effort or benchmark publication.

However, they are faster and easier to apply and have formed the bases of many national calculation systems. In addition, they may be used during energy auditing procedures, where brief calculations and rough performance estimates are often sufficient for evaluating performance and estimating improvement potential. When performed properly, an energy audit will provide a detailed evaluation of energy performance by identifying end-use energy consumption and comparing the performance of a building to its own potential performance, following the application of energy efficiency measures.

ACCEPTED MANUSCRIPT

Simulation

Dynamic simulation, also known as dynamic thermal modelling or building energy performance simulation, involves the use of computer models to simulate the performance of a building in determined conditions.

According to Hensen and Lamberts, "simulation... is multi-disciplinary, problem-oriented and wide in scope. It assumes dynamic... boundary conditions and... aims to provide an approximate solution of a realistic model of complexity in the real world" [58].

Simulation can be used for many purposes, including lighting design, airflow (through CFD), comfort and HVAC system design. The design of low energy and efficient buildings is increasingly reliant on the use of building simulation tools for performance prediction, often linked to the design calculations for HVAC systems. This paper focuses on the integrated models that are used to predict or evaluate energy performance.

Building energy simulation software is generally applied to new buildings for compliance evaluation, assessment tools and design tools [59]. In addition, simulation is increasingly used at the operational stage, for monitoring and verification or evaluation of performance through calibrated simulation [60]. Figure 3, Figure 4 and Figure 5 demonstrate the application of simulation for existing buildings, compliance and design respectively. It should be noted that these uses are applied using different criteria and often give very different performance predictions. This section focuses on the technical aspect of modelling, and how this can be used for design and performance prediction, while section 6 of this paper discusses the application of compliance frameworks through thermal modelling and other assessment methods.

The evaluation of building performance through simulation generally involves the following steps:

Construction of a model to represent geometry, material properties, building specifications and external environment (orientation and shading);

Addition of data regarding the building systems and performance;

Application of proposed or estimated usage patterns, schedules and operational parameters;

Dynamic performance simulation for periods of up to a year, applying external conditions using measured climatic data or calculated typical meteorological years (TMYs);

Evaluation of results and energy performance estimates; and

Comparison of results to measured data or performance criteria, such as benchmarks or regulatory requirements.

ACCEPTED MANUSCRIPT

In practice, many simulations require several stages of iteration to reach desired levels of confidence, and are highly dependent on the skill and technical background of the user [22]. There is often a clear difference between the predicted performance in simulation models and the observed or measured performance; this is known as the Performance Gap and is described in more detail below.

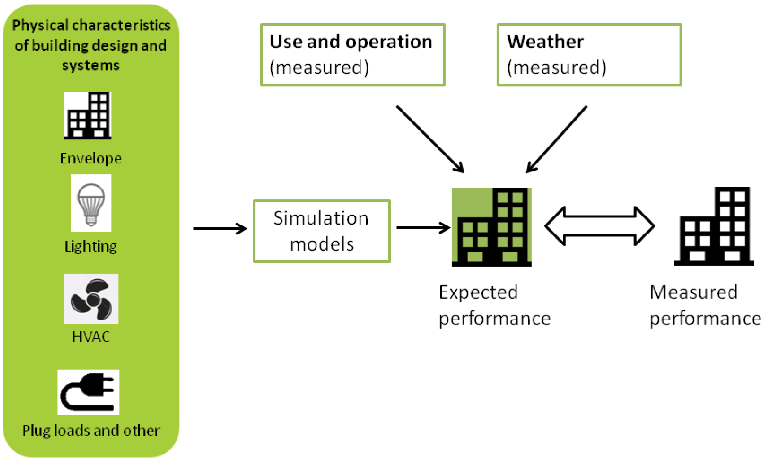


Figure 3 - Evaluation of energy performance in existing buildings by calibrated simulation

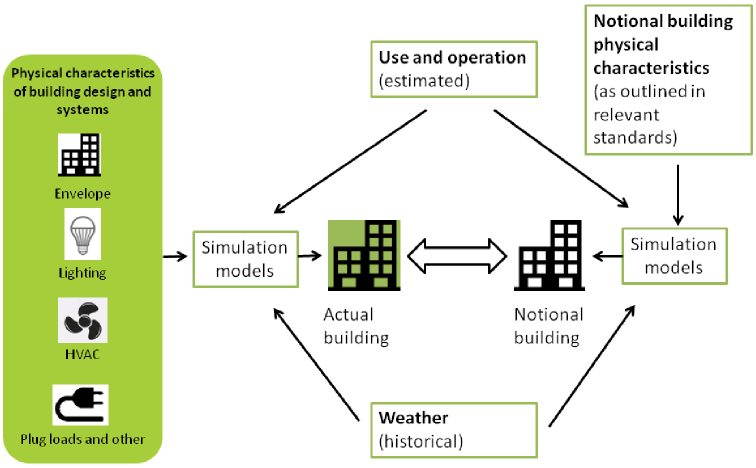


Figure 4 - Evaluation of energy performance by simulation for compliance

ACCEPTED MANUSCRIPT

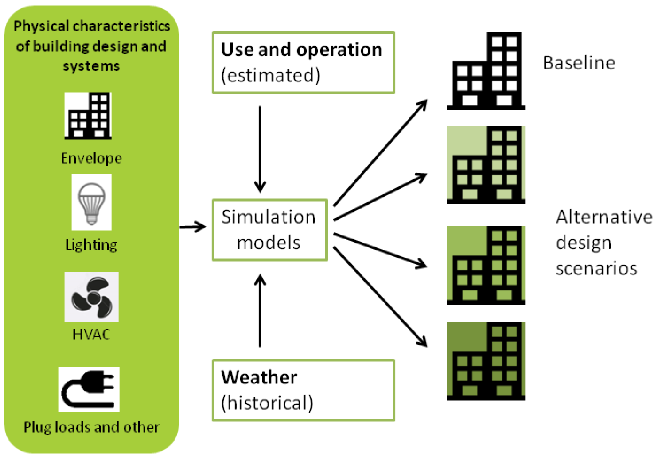


Figure 5 - Evaluation of energy performance by simulation at design stage

For simulation to be effective as a tool for building performance prediction, it must prove itself to be robust, limit uncertainty and identify sensitivity [58]. Current problems include the lack of standards, expense of creating accurate models and poor integration between 3D modelling packages and simulation software [28].

A series of building simulation software packages are available commercially or freely distributed; these were described extensively by Crawley and colleagues [61], and the major software systems currently available are mostly updated versions of those reviewed in 2008. In addition to those reviewed, some software packages are available that provide graphical user interfaces, programming and even online severs for existing calculation software, typically using EnergyPlus [62]–[64]. The US Department of Energy provides an online directory, which lists 42 "Whole Building Energy Simulation" programmes, although several of these are specifically aimed at residential buildings [65]. Another review compares 23 separate building simulation programmes, and maps the capabilities and characteristics of nine principal simulation engines used by these programmes [66]. All have geometric descriptions and consider thermal mass and most have time-step approaches and consider simultaneous radiation and convection. Beyond this they show a large variation in the detail of treatment of simulation parameters, such as comfort, solar shading and the solution method for heat transfer. In large buildings, the application of these tools to HVAC simulations is often the most complex aspect and produces varied results, as described by Trcka and Hensen [67]. It is clear that the selection of the most appropriate programme depends on the use, type of modelling, experience of the user and computing power available.

A detailed Applications Manual published by CIBSE covers the technical implementation and quality assurance for simulation of building performance [59]. In addition, several publications cover technical issues and methodologies for simulation [58], [68]. This section does not attempt to repeat these studies, but reviews the state of the art around the use of simulation as a tool for the evaluation and prediction of building energy performance, raising the principal issues and challenges as well as the applications for which building simulation is most suitable.

ACCEPTED MANUSCRIPT

Performance gap between predicted and observed performance

Building energy simulations use highly detailed building physics models and extensive parameter inputs by skilled professionals and as such should expect to provide highly accurate performance predictions. However, the PROBE studies carried out in the late 1990s by CIBSE first pointed out a "Performance Gap" between predicted and observed performance in supposedly low-energy buildings [69], and since then this has been the topic of extensive further study.

In a review of the performance gap, de Wilde separates the performance gap into three types: the gap between machine learning and measurements, the gap between predictions and display certificates in legislation and the gap observed between first-principle predictions and measurements (discussed here) [70]. The final type of performance gap, between modelled and measured data, can be sub-divided into static (comparing predictions from modelling to measured energy use) and dynamic (using calibrated predictions and measured energy use in a longitudinal approach to diagnose underlying issues) [71].

The review by van Dronkelaar and colleagues finds the magnitude of the performance gap to be +34%, with a standard deviation of 55% based on 62 buildings. This paper finds the dominant causes to be:

Specification uncertainty in modelling;

Occupant behaviour; and

Poor operational practices [71].

Other issues raised include poor commissioning, onsite workmanship, model simplification, inter-model variability, weather data and operational schedules.

Daly and colleagues clearly state that "building energy modellers typically rely on a range of simulation assumptions and default values for certain ‘hard-to-measure’ building and behavioural inputs to building performance simulations", and that variations in these values can cause variations of more than 50% from baseline archetypes commonly used in Australian cities [72].

Tuohy and Murphy propose the use of BIM as a framework to address the performance gap, based on improved building data [73]. Heidarinejad and colleagues study LEED certified buildings to map the variation and uncertainty in three main variables: days of operation, process loads and occupancy rates. Analysing these variables can give a measure of risk for levels of performance gap [74].

Two separate studies have discussed in detail the performance gap in schools in the UK, identifying handover and user education [75] as well as policy, design, and commissioning [76] as major issues which must be addressed in order for buildings to meet their potential performance. Fedoruk and colleagues find the primary causes of the performance gap (and barriers to adequate performance) to be non-technical, discussing institutional regimes and

ACCEPTED MANUSCRIPT

the way that different stages in the building lifecycle were specified, contracted and implemented [77].

Thus model calibration, uncertainty analysis and the impacts of user behaviour must all be considered in order to minimise the impacts of the performance gap.

Model validation and calibration

In order for a modelling approach to be applied with a reasonable degree of confidence, it must be validated. This validation is carried out by comparing model outputs to results developed by three methods [24], [78]:

Empirical validation – test of model and solution process, compare results to real data;

Analytical validation – test of solution process, compare results to known solutions; and

Comparative testing – relative test of model and solution process, compare results to outputs of other, validated software tools.

A methodology for building simulation validation called BESTEST (Building Energy Simulation Test) was developed by the IEA and has been widely applied [79]–[81]. It formed the basis an ASHRAE standard for the validation of software tools [82], which applies the three types of test identified above separately to the thermal fabric of the building and the mechanical equipment. It aims to identify and diagnose prediction differences caused by "algorithmic differences, modelling limitations, faulty code or input errors" [24].

A range of studies have shown individual case studies or evaluations of the results of many building energy simulations. While some carefully endorse simulation as being able to predict performance when models are accurately built [83], others show that uncalibrated simulations are often very far from representing reality and may not be useful tools for predicting performance [84], [85].

Simulations of existing buildings can be calibrated to provide higher confidence in their results. This can validate the energy simulation methodology used (empirical validation, as discussed above), or can be used to develop highly reliable models for evaluating and optimising the operation and retrofit of buildings.

ASHRAE Guideline 14 provides a set of technical procedures for carrying out and evaluating building simulation calibration [60], while Raftery and colleagues propose a formal methodology [86]. A detailed discussion of calibration procedures is presented by Fabrizio and Monetti, who identify five levels of calibration, based on available information. Level 1 requires only utility bills and as-built data, Level 2 includes a site visit, Level 3 adds a detailed audit and Levels 4 and 5 add short- and long-term monitoring respectively [87]. A model can be considered to be calibrated when it meets a set of statistical criteria in monthly or hourly evaluations and comparisons, generally having Mean Bias Error (MBE) and Coefficient of Variation of the Root Mean Squared Error (CvRMSE) within certain intervals [60], [88].

ACCEPTED MANUSCRIPT

Calibration can be carried out by both manual and automatic methods, using iteration, graphical techniques, special tests and analysis procedures, mathematical procedures and combinations of the above [28], [87]. A series of studies have published simulation calibration results and processes in recent years, using a range of tools and procedures and showing successful results [89]–[93] The review papers mentioned above both cite more complete lists of calibration studies [28], [87]. However, there are still a range of problems with the calibration of building energy models. As Coakley and colleagues conclude, these issues can be resolved into seven principal areas [28]:

Standards – lack of consensus

Expense – significant effort required

Simplification – models are over-specified and under-determined

Inputs – difficulty of accurate measurements

Uncertainty – should be stated, and is often disregarded

Automation – this would aid the process.

On-going research aims to further standardise procedures and simplify the implementation of building energy model calibration. For example, the AUTOTUNE project is further developing automated calibration procedures, to apply to retrofit projects [94]. This work should make calibration more reliable and accessible for non-academic projects.

As one major example of the application of calibrated simulation, the International Protocol for Measurement and Verification of Performance (IPMVP) identifies calibrated simulation as one of the four options for measurement and verification, to be used in complex evaluations involving several interlinking efficiency measures [95].

Characterisation and quantification of uncertainty

Building energy simulation clearly involves a series of uncertainties, regarding physical parameters, weather, occupancy profiles and calculation algorithms. However, the "explicit appraisal of uncertainty is the exception rather than the rule and most decisions are based on single-valued estimates" [96], which is understandable in the context of commercial building simulations that have limitations on time and resource. As Figure 6 shows, the level of uncertainty can be reduced only with increases in model complexity and improved input factors.

Macdonald and Strachan use sensitivity analysis to address uncertainty associated with the following issues:

Model realism;

Input parameters, where measured data are not available;

ACCEPTED MANUSCRIPT

Stochastic procedures, including variations in occupancy levels and operational factors;

Simulation program capabilities; and

Design variations [97].

Many uncertainty analyses are carried out using sensitivity analyses, and methods for applying sensitivity analysis in buildings are reviewed in detail by Tian [98]. As Hopfe and Hensen point out, uncertainty and sensitivity analysis brings additional benefits and may help to simplify models, analyse their robustness, identify unexpected sensitivities and allow decision support simulations [99].

Different studies have focussed on uncertainty evaluations in specific areas, such as microclimate and indoor temperature distributions [100], [101], set points and schedules [72], material properties and physical characteristics [97], [99] and model uncertainty [102]. All of these studies show that there is a clear need to take into account the different types of uncertainty, and a quantitative appraisal of these can contribute to more rational design decisions.

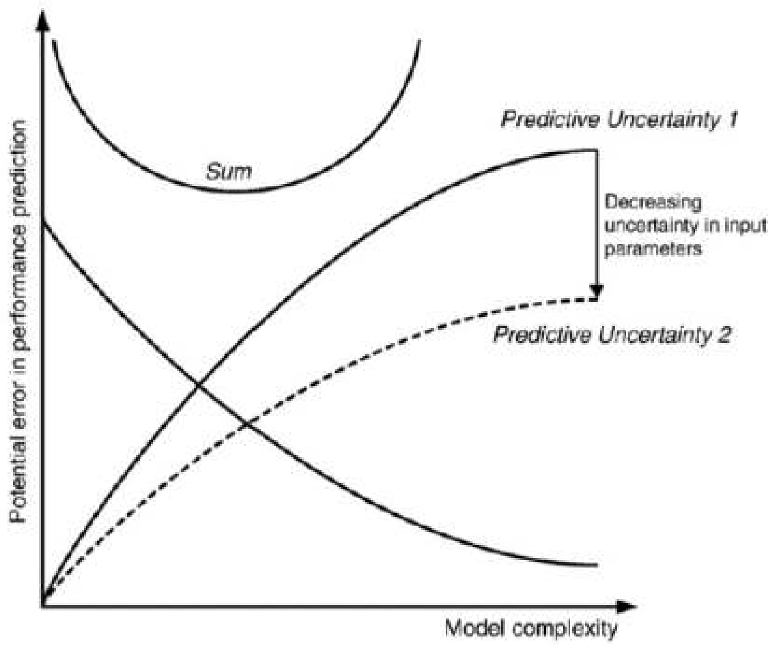


Figure 6 - Model uncertainty versus complexity [67]

Impacts of user behaviour

The behaviour of users clearly has a large impact on the total building energy performance and is less well studied than other aspects such as modelling algorithms. In many cases, building energy models do not accurately model user behaviour. Ryan and Sanquist suggest that despite the improvement in building energy models over the years, "there have been no major

ACCEPTED MANUSCRIPT

improvements in the methods used to model occupant's behaviour". They suggest the use of stochastic and other detailed behaviour models, although this would require much larger data collection efforts [78]. Hong and colleagues propose ways for gathering and organising this data in order to provide more effective data on users for simulation [103].

The sensitivity of total building energy consumption to assumptions on user behaviour has been shown in several studies, considering thermal comfort, set points, control strategy and energy management [104]–[107]. As Hoes and colleagues note, the increased focus on sustainable buildings has the potential to increase the impact of user behaviour on total performance, and user behaviour should be analysed in more detail for individual buildings [108].

The behaviour of occupants in buildings is being extensively studied by Annex 66 of the IEA's Energy in Buildings and Communities Programme, which aims to produce improved and consistent models for simulation of occupant behaviour in buildings [109].

Stock and district models

In some cases, building simulation is used to develop stock models or archetype buildings that can be used to represent large sets of buildings or particular typologies. These models can be used for supporting policy development or implementation of retrofit programmes, by simulating the impact of changes to design and extrapolating to understand their impacts on a large set of buildings. The US Department of Energy publishes a set of reference models for common commercial building typologies [110]. Heo and colleagues present a probabilistic methodology based on Bayesian calibration to model large stocks of buildings efficiently while considering uncertainty; this work is aimed at understanding the impacts of large-scale retrofit programmes [111]. Nikolaou simulates a virtual building dataset with 30,000 buildings to represent the building stock and develop national benchmarks in the absence of detailed stock data [112]. Katafygiotou and Serghides built archetype models of schools in Cyprus to evaluate the benefits of retrofit projects in this typology [113], Salat maps the effect of urban typologies on energy performance in France [114], while two of the authors of this paper used simulated archetype models as a basis for the development of national benchmarks [37].

An emerging field of research is that of urban city or district models, which use modelling or simulation techniques to understand the energy performance of entire urban regions. Choudhary and Tian map gas consumption in London to separate the extrinsic and intrinsic factors responsible for energy performance [115], while Davila and colleagues have recently published an urban energy demand model for Boston [116]. A review paper by Reinhart and Davila describes the techniques of urban energy modelling and maps the principal work in this area [117].

It should be noted that although these models can identify significant energy saving potential, the practical implementation of these energy saving measures may be difficult; the literature identifies a "rebound effect", in which energy efficiency improvements are translated into

ACCEPTED MANUSCRIPT

increased comfort or levels of service, rather than energy saving. This is principally documented in the residential sector, but is likely to be relevant in other areas as well [118].

Limitations and application

Building energy simulation is a highly effective tool for modelling individual buildings, whether existing or at design stage, provided sufficient information is available regarding the building characteristics and usage profiles. Although the sector is advancing rapidly, there is a need to develop further understanding of the key factors that reduce accuracy and confidence, in order to be able to apply simulation effectively and rapidly at scale and depend less on the skill of individual modellers. ASHRAE's certification programme for building energy modelling helps to recognise professionals who have the necessary levels of experience to implement simulation tools effectively [119]. Annex 53 and Annex 66 of the IEA's Energy in Buildings and Communities Programme both go into more detail on how proper calibration and knowledge of uncertain boundary conditions and input values (in particular those relating to occupancy behaviour effects) can substantially reduce the uncertainty of simulation [120], [109].

Harish and Kumar note that: "A systematically developed simplified building model, dealing with temperature, heat and relative humidity is lacking" and that "models which are accurate to a fair degree require a large computer memory and processing or computation time" [66], while Daly and colleagues show that changes in input variable assumptions for simulation can vary payback for lighting upgrades from 2.4 to 10.3 years [72]. A discussion on the ethics of building performance simulation calls for a clearer discussion around presentation, interpretation and use of the results of simulations if they are to continue to be used to inform and direct "responsible decision making" [121], and all results should certainly be presented with accompanying estimates of uncertainty.

ACCEPTED MANUSCRIPT

Statistical methods

The increase in availability of large volumes of data on building energy performance has allowed the development of top-down methodologies for the analysis of building energy performance. These methodologies use statistical techniques to predict and evaluate energy performance based on existing datasets of multiple buildings. Most statistical models use some sort of regression – often Ordinary Least Squares (OLS) – to model and explain the energy performance of buildings. The statistical models are often used to form the basis of benchmarks and evaluation systems, by calculating an expected (or target) energy consumption level for a specific building.

Generally, these evaluations are based on the Energy Use Intensity (EUI), measured in energy consumption per unit of floor area, as this is often the easiest information to obtain for large quantities of buildings. By adding information on further variables and statistically mapping the relationship between these variables and energy consumption, these models can then be used to predict performance levels for buildings based on their characteristics (as in Figure 7). Simple statistical evaluations such as average or median performance within a dataset are used in some cases [122], [123], but may prove to be insufficient for evaluating performance, showing a requirement for more detailed models and correction factors to normalise for variables impacting energy performance [124].

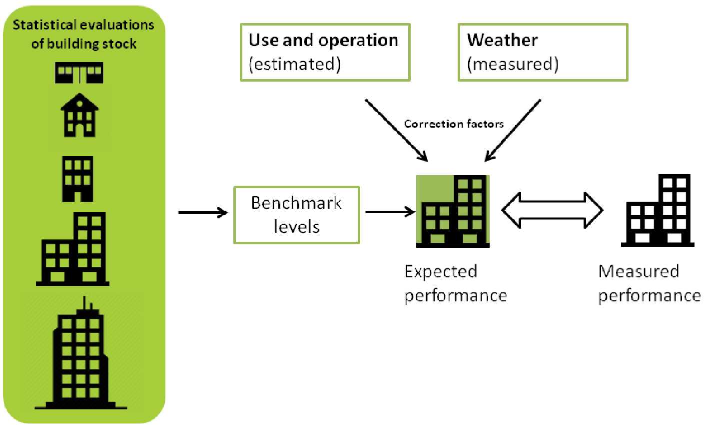


Figure 7 - Evaluation of building performance by comparison with statistical benchmarks

The principal statistical methods for benchmark development and evaluation of building energy performance are described below, while Table 4 briefly summarises the principal references and examples of the application of these methodologies.

ACCEPTED MANUSCRIPT

Algorithm	Applications	References
Simple and multivariate linear regression	Simple models for building performance based on a few characteristics	[60], [124]
Change-point regression	Model the non-linear effects of external conditions, e.g. below a certain external temperature, heating systems are switched on.	[29], [60]
Gaussian process and Gaussian Mixture regression	Prediction of dynamic performance, with an understanding of uncertainty. Flexible models, but more complex.	[29]
Stochastic Frontier Analysis	Effective when there are a large number of efficient buildings and a few that inefficient. Outliers may make the method ineffectual, as residuals will be large.	[16], [125]
TOPSIS	Can be used to develop effective benchmarks, based on regressions.	[126], [127]
Data Envelopment Analysis	Evaluates the technical efficiency and improvement potential of buildings. Can only be applied to buildings within the original dataset.	[16], [128]
Correction factors	Relate building performance to physical parameters, useful for benchmarking.	[129]

Table 4 - Sumary of principal statistical and machine learning methods applied to buildings

Simple and multivariate regression models

The most widely used statistical models for building performance evaluation are regression models, which relate energy consumption to one or more variables, generally using Ordinary Least Squares (OLS) regressions. ASHRAE Guideline 14 includes a section that describes regression models in detail, and presents eight sample models for whole-building energy use, principally relating performance to climatic variables, as shown in Table 5 [60]. Simple regressions relate building energy performance to a single variable, while multivariate regressions link performance to several input variables.

Name	Independent variables	Form	Examples
No adjustment/ constant model	None	$E = E_b$	Weather-independent use
Day-adjusted model	None	$E = E_b \times day_b/day_c$	Weather-independent use

ACCEPTED MANUSCRIPT

Two-parameter model	Temperature	$E = C + B_1(T)$	
Three-parameter model	Degree-days/ temperature	$E = C + B_1(B_2 - T)^+$ $E = C + B_1(T - B_2)^+$	Seasonal weather-sensitive use (fuel in winter, electricity in summer for cooling), weather-sensitive use
Four-parameter change-point model	Temperature	$E = C + B_1(B_3 - T)^+ - B_2(T - B_3)^+$ $E = C - B_1(B_3 - T)^+ + B_2(T - B_3)^+$	
Five-parameter model	Degree-days/ temperature	$E = C - B_1(DD_{TH}) + B_2(DD_{TC})$ $E = C + B_1(B_3 - T)^+ - B_2(T - B_4)^+$	Heating and cooling supplied by the same meter
Multivariate model	Degree-days/ temperature, other independent variables	Combination form	Energy use dependent on non-temperature-based variables (occupancy, production, etc.)
Variable-base degree-day model	Heating degree-days, Cooling degree-days	$E = C + B_1(DD_{BT})$	

Table 5 - Sample regression models for the whole-building energy use approach, from [60]

Change-point regression models are used to model the non-linear impact of variables, and are generally used when buildings display strong on-off schedule-dependent loads. For example, a heating system may only be switched on below a certain external temperature, changing the coefficient relating temperature to energy consumption.

An influential study by Sharp [124] used stepwise linear regression on buildings in the US Commercial Building Energy Consumption Survey (CBECS) to identify the variables with greatest impact on energy performance of buildings. The two dominant variables identified are the logarithm of the population density and the number of personal computers. Along with four other variables (operating hours, whether the building was owner-occupied and the presence of economizers and chillers) , these explain almost all of the variation in electrical EUIs that can be explained by the highly detailed CBECS data. A model was developed from these parameters that served at the basis for the US EnergyStar Portfolio Manager models. A more recent update of this methodology states that:

"The technical review concluded that [other techniques including SFA and DEA] do not offer an advantage over ordinary least squares regression. Ordinary least squares regression provides a technically rigorous approach and yields descriptive linear equations that are statistically valid and easily replicable." [14].

Bloomfield and Bannister use multi-dimensional regression to develop a simple model for shopping centre energy benchmarking [129], while Chung develops benchmarks for supermarkets with central air conditioning systems considering nine variables in a regression

ACCEPTED MANUSCRIPT

analysis [130]. Further studies apply regressions models to benchmark and characterise performance in different building types, including hotels [131], banks [132], office buildings [133], IT facilities [134] and cooling loads [135]. In the latter reference, accuracy is improved by using principal component analysis (PCA), dynamic two-step analysis and consideration of the continuous effect of high temperature (thermal inertia). So and Richman apply multiple regression models to map the energy consumption of groups of buildings that do not have individual metering [136].

Hsu applies regression models to evaluate dependence of measured energy performance on different variables in New York buildings, and shows that historical benchmarking data explains performance better than energy audit data or EnergyStar benchmark models, indicating that even regression models which include multiple variables are likely to leave much of the variation in energy performance unexplained [137].

A Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) is proposed for developing benchmarks from regression models [126], although Wang notes that multi-collinearity of variables in a linear regression may be a challenge and suggests the application of TOPSIS using both regression and Principal Component Analysis [127].

Gaussian regression models

Gaussian process regressions assume that data can be represented as a sample from a multivariate Gaussian distribution, and construct a model by specifying the structure of a covariance matrix composed of explanatory variables. The Gaussian Mixture Regression Model is a parametric probability density function represented as the weighted sum of probability densities. Both are presented in detail by Zhang, who carries out a case study and finds Gaussian Mixture Regression Models to give better fits to test data than Gaussian Process, ANN or change-point regression, although the differences are small and the case study is simple [29]. Olofsson applies Gaussian Mixture Models to evaluate the energy performance of a small set of buildings and evaluate energy efficiency parameters [138].

Stochastic Frontier Analysis

Stochastic Frontier Analysis (SFA) is a corrected OLS regression that includes two separate error components: a random error and a measure of inefficiency. Thus the regression line becomes an efficiency frontier, in which the inefficiency is measured as the distance from the frontier, once the random variation is subtracted. This type of analysis is described in detail by Chung [16]. Another study applies Stochastic Frontier Analysis to a dataset of Canadian buildings to estimate savings potential [125].

Data Envelopment Analysis

ACCEPTED MANUSCRIPT

Data Envelopment Analysis (DEA) is a mathematical procedure that is applied to homogenous sets of data to carry out multi-factor productivity analysis. In the context of buildings, each building is defined as a Decision-Making Unit (DMU), and comparative efficiency levels are defined. A non-parametric piecewise frontier is constructed from the DMUs, and used to define maximum efficiency levels. The buildings efficiency levels are assessed based on their distance from the frontier. The Variable Returns to Scale model separates the overall efficiency into scale efficiency and technical efficiency, and buildings are ranked or evaluated according to their pure technical efficiency [128].

DEA is a nonparametric approach and relies on mathematical programming, rather than regression lines. This brings the advantage of not assuming any particular functional form, but it has been criticised for its "deterministic and nonstatistical nature" [16]. In other words, there are no representations of real physical phenomena in DEA, making it harder to validate the models. In addition, a small dataset or one where all buildings are inefficient will give misleading results. A single outlier in DEA can skew the results significantly.

DEA has been applied by Önüt for evaluating the energy performance of hotels [139], as well as being applied by Lee for benchmarking government buildings [140], evaluating the effectiveness of building energy management [141], benchmarking cooling energy consumption [142] and classifying the performance of clusters of office buildings [128]. Ruiz extends DEA with a consideration of "expert preferences" to help prevent the selection of inappropriate benchmarks [143].

Overall, DEA is not a widely used technique. In part, this is because the results can only be applied for internal benchmarking; the results from DEA cannot be applied to buildings that were not within the original dataset [16].

Correction factors and tailored benchmarking

An alternative to the use of statistical methodologies for direct comparisons of buildings is to use a statistical database or engineering methodology to develop "correction factors" which directly represent physical phenomena and can be applied to the EUI of a building to make it comparable to other buildings or a standard performance level. (Alternatively, the correction can be applied to the standard performance level or benchmark to make it comparable to the building's EUI).

CIBSE publishes energy benchmarks in TM46, which includes descriptions of the three types of performance correction that are permitted in order to allow comparison with benchmarks [49]:

Weather adjustment, using a degree-days methodology.

Occupancy adjustment, considering only periods of the year in which the building is operational.

ACCEPTED MANUSCRIPT

Separable energy uses, removing specific end-use process loads which may distort performance (this generally requires sub-metering).

The regression analysis used on shopping centre energy performance by Bloomfield and Bannister produces a benchmarking model that corrects for cooling and heating requirements, car parking area and serviced area [129]. Lam develops a climatic index based on dry bulb temperatures, wet bulb temperatures and solar radiation; this index was validated through long-term simulation and showed that it could be effectively used as a climate correction factor [144]. Finally, Hong applies statistical analysis and artificial neural networks (ANN) to evaluate the CIBSE benchmarks for schools, and finds that the static benchmarks should be updated and reviewed regularly in order to follow trends in building performance. In addition, there are several factors that are shown to have significant impacts on performance but are not considered in benchmark corrections. These include the number of pupils (impact on electrical consumption) and the compactness of the school (impact on fossil fuel consumption) [145].

This concept is taken further by the development of tailored benchmarks, as described by Bordass and colleagues [146]. The concept of tailored benchmarks developed from the ECON19 guide for energy performance in offices [48] and from the TM22 methodology for energy auditing and end-use energy consumption breakdown [42], and can be expanded to develop specific types of benchmark depending on amount of detail in the inputs provided. Thus, a tree diagram with bottom-up energy performance calculations allows different levels of energy evaluation depending on the building data supplied. More recently, a study by Yan and colleagues reinforced the need for customised benchmarks in order to consider energy saving potential [55].

Limitations and application

Hsu notes that although OLS regression is still a popular method, it can "result in biased coefficients that are inflated in magnitude, have the wrong signs, or which radically change depending on the model and variables that are selected"; the solution provided is the use of regularisation, or penalised regression, for selection of variables and interactions [147].

The principal limitation of statistical models would seem to be their lack of a relation to physical phenomena, meaning that it can be difficult to interpret the results and identify errors. Kramer concludes that additional work is needed to develop a simplified model with physical meaning [27].

ACCEPTED MANUSCRIPT

Machine learning

With the rapid growth in the quantity and availability of data on buildings, through more advanced BMS, remote metering and regulations requiring buildings to publish energy data, there is an increased interest in the use of big data techniques such as data mining, artificial intelligence and machine learning.

Machine learning is a term used to describe algorithms that can "learn" from data, using large amounts of data and comparatively few additional inputs. In buildings, several machine learning techniques have been applied over the past decade to predict and evaluate energy performance in different situations. The most common are artificial neural networks and clustering analyses.

Artificial neural networks

Artificial Neural Networks (ANN) are a machine-learning technique often used to solve ill-defined or complex problems. They operate on a "black box" principle, requiring no detailed system information, but instead learn the relationships between input variables and outputs using historical data. They can handle complex systems and learn key information patterns, as well as being "fault tolerant, robust, and noise immune" [148]. As such, they are increasingly used for modelling building energy consumption for evaluation and prediction of performance, including both short-term performance and total consumption (EUI).

Hong applies ANNs to schools in the UK, with mean absolute percentage errors of 22,0% and 20,6% for prediction of heating and electrical consumption respectively, showing it to be more effective at assessing operational efficiency than simple comparison with sector benchmarks, but less accurate than bottom-up methods (simulation and engineering calculations) [21].

Zhang's comparison of ANN with different regression models notes that ANN is less effective than the simpler models at predicting energy performance, probably because of the requirement for large volumes of training data [29]. However, Neto and Fiorelli show satisfactory performance comparing the results of ANN with EnergyPlus simulation results [91].

Ekici and Aksoy show satisfactory results when ANN is applied to simple heated buildings [149], while Yalcintas develops effective energy performance benchmarks using ANN in tropical climates [150]. However, Li's comparison of ANN with other methods (such as multiple linear regression and RC Networks) shows ANN to be less accurate than the other models.

A problem with ANN is that it can be hard to extract rules and physical relationships from datasets. Hou and colleagues apply rough set theory to identify the relevant parameters for energy performance; this pre-processing makes the subsequent ANN evaluation more effective [151].

ACCEPTED MANUSCRIPT

In Hong's application of ANN to school building performance, the authors extract several factors that show themselves to be prominent determinants of energy consumption, and notes the limitations of static benchmarks; the datasets should be reviewed regularly to update evaluations [145].

Li and colleagues aim to optimise the performance of ANN; PCA is used to reduce input dimensions and simplify the model structure, then an improved particle swarm optimisation algorithm is used to optimise the internal parameters of the model. When compared to ANN using genetic algorithms or typical methods, the particle swarm optimisation hybrid model shows a better ability to predict performance [152].

Melo and colleagues use ANN for surrogate buildings as a baseline for development of performance benchmarks for energy labelling [153].

Clustering analyses

Clustering is a machine learning technique that is considered to be one of the most effective data mining processes, and is widely used in energy analysis. It addresses the problem of finding a structure in a collection of unlabelled data, and can map distributions and patterns of the underlying datasets. Buildings are typically classified by use type or typology, but clustering is a more effective way of developing these categorisations. The process of clustering involves the following steps [154]:

Data collection;

Separation of variables that might affect building energy consumption;

Application of an appropriate clustering algorithm to classify the building samples; and

Benchmarking of buildings within their clusters .

The most commonly used algorithm for building energy analysis is K-means clustering, which classifies a dataset by defining centroids, each of which will form the midpoint of a cluster. The clustering analysis should be followed by a validation stage, for example ensuring that the optimal number of clusters has been identified. Finally the results must be interpreted. This may involve a comparison of the performance of each building to its expected energy performance, defined by the centroid of the cluster. Clustering is often used in often conjunction with Principal Component Analysis (PCA) to reduce the dimensionality of the problem and improve the understanding of the correlation between explained variables [155].

A complete description of the K-means clustering process is given by Gao and Malkawi, who show that clustering is more effective than the EnergyStar methodology for benchmarking energy performance, as it is capable of incorporating "all the statistically significant building characteristics affecting energy use" [154].

However, Hsu critiques commonly used clustering techniques, noting that although K-means clustering generally gives stable results, choices such as the algorithm used, variables selected

ACCEPTED MANUSCRIPT

and initial assumptions for the analysis can cause it to give widely different values. This could lead to misleading results in some cases. Thus, Hsu shows that carrying out clusterwise regression, simultaneously fitting the model for the clustering and linear regressions of the clusters, can give higher predictive accuracy, at the expense of model stability [156].

Cluster analyses have been widely used in recent years. Santamouris introduces clustering for building energy efficiency analyses, using fuzzy techniques to classify school buildings in Greece [157]. Clustering with PCA has been used by Gaitani to evaluate heating in schools [155] and Lam to evaluate chillers' energy performance [158]. Yu and Chan also evaluate chiller performance through clustering and multivariate analysis [159].

Gao shows K-means clustering to be more effective than EnergyStar in developing comprehensive benchmarks for building performance [154]. Petcharat applies clustering analysis to evaluate energy efficiency potential from lighting upgrades [160], while Heidarinejad applies clustering techniques to simulated energy consumption of LEED certified buildings, to show that unregulated and internal loads are the principal contributor to total energy consumption [74]. Arambula-Lara uses clustering analysis by MLR to evaluate school energy performance and define reference buildings for each cluster [161]. Pieri ranks hotel energy performance based on clustering analysis and identifies improvement targets [162]. Famuyibo develops archetype building models through clustering techniques [163]. Yu uses clusters to develop insights into occupant behaviour and energy performance [164].

Other machine learning methods

Several other machine learning methods are referenced in the literature, although their implementation is not yet found to be widespread.

Support Vector Machines (SVM) are a machine learning technique that can be used for regression estimation and prediction. They were applied by Dong to predict energy consumption in tropical buildings [165] and by Li to predict hourly cooling loads [166]. Zhao and Magoulès introduce a parallel implementation of SVM to speed up model training on large datasets [167].

Yu and colleagues present decision trees as an accurate method for evaluating building energy performance that is relatively simple to apply [168], while Tooke and colleagues develop an innovative approach applying random forests to remotely collected LiDAR data to predict building characteristics.

Li and colleagues develop a hybrid genetic algorithm adaptive network-based fuzzy interference system (GA-ANFIS) and use a case study to show that it can have superior or equivalent performance to ANN in some cases [169].

Limitations and application

ACCEPTED MANUSCRIPT

In general, machine learning technique are applied to "big data", as increasing levels of monitoring and metering make large quantities of data available for analysis. This makes them powerful computational techniques when sufficient data are available, but leaves them dependent on the collection of this information.

As with statistical models, it would seem that one of the main drawbacks of machine learning methods is that their "black-box" approach makes it difficult to show real, physical interpretations for the findings of the models.

Li and Huang compare ANN, MLR and R-C networks and find that none of the models give prediction errors that follow normal distribution, indicating that further development is needed [170].

However, this is an emerging area that clearly shows great potential for improving the evaluation and prediction of building energy performance where appropriate datasets and monitoring information are available.

ACCEPTED MANUSCRIPT

Other building evaluation techniques

There are several types of evaluation that are not considered by the division of building energy performance evaluation into the four categories above (engineering calculations, simulation, statistics and machine learning). This section briefly describes these evaluation techniques.

Dynamic methods and real-time analysis

Increasingly, dynamic methods are being used to evaluate and predict building performance in real-time. These techniques are used to optimise building operation through the control systems, and use various techniques. Li and Wen provide a review of the application of modelling for operation and control, comparing the use of real-time simulation models (white-box), R-C models (grey-box) and statistical regression models known as "Autoregressive with Exogenous Outputs" or ARX (black-box). They conclude that simulation methods are too computationally complex for online use (although they may be used to define an operation strategy), statistical methods are rapid and effective but require large amounts of training data, while the rapid computation time of simple, grey-box engineering models makes them the most effective for practical application [26]. Thus the simplicity of engineering calculations, which reduces their usefulness in whole-building energy evaluations, is actually beneficial for optimising operation.

Pang investigates the use of simulation for real-time operational control [171]. Yun uses ARX for hourly predictive control of building systems [172], Karatasou shows that ANN can predict energy performance 24h ahead [173], while Kusiak uses ANN to save up to 30% of HVAC energy consumption with predictive control [174]. In the absence of model training data for ANN, de Wilde uses building simulation models to train the ANN models for building energy management [175]

Load-curve analysis and energy bill disaggregation

Several studies discuss the evaluation of building performance using easily available information from energy bills or simple monitoring data in the form of typical load-curves.

Yan shows an energy bill disaggregation method for simplified energy performance analysis in mechanically cooled buildings [176], while Iyer uses facility energy models to disaggregate loads in supermarkets for energy analysis [177].

Rasanen groups energy users into clusters based on their load profiles, with data-based load curves defined for each group [178], while Panapakidis uses clustering to evaluate load curves and identify energy efficiency opportunities [179]. Gul highlights the importance of analysing load profiles in performance evaluation, alongside the impacts of occupants [180]. Jota synthesizes load curves using clustering analysis in order to use them for energy management programmes [181].

ACCEPTED MANUSCRIPT

Energy Audits

Energy auditing is a technique for evaluating the energy performance of a building and identifying its improvement potential, generally with a visit by a qualified professional who will measure and evaluate the performance of the principal building systems. The end result of the audit includes a list of energy saving measures, based on the real physical characteristics, operation and usage profiles of the building. These energy saving measures will often show estimated energy savings (developed using the methodologies discussed in section 0), cost savings and payback for installation of the energy saving measures. These measures will include:

No-cost and low cost options, involving changes in operational or maintenance procedures;

Medium cost measures, requiring the installation of new sensors, controls or small investments; and

High cost measures, involving capital investments in new equipment.

Energy audits are more expensive than the benchmarking methodologies reviewed above, and are often carried out in buildings where benchmarks have indicated a significant energy saving potential. In a sense, they provide a customised benchmark for the building, comparing its current performance with the performance that could be achieved with the implementation of feasible measures. Both ASHRAE and CIBSE publish methodologies and guides for energy auditing, giving guidance for contracting and carrying out these procedures [45], [182].

Post-occupancy analysis, comfort and environmental quality

As noted in section 0, there are six principal factors which define building energy consumption, of which three are related to people: indoor environmental conditions, user behaviour and operation & maintenance procedures [1]. Thus a clear limitation of the building performance evaluation methods listed above is their inability to receive and process data on the actual conditions and behaviour inside the buildings. It has been recognised for some time that ideally, building performance evaluations would link user satisfaction and energy performance [46], but currently no commonly used methodologies make this link explicit or use it as a weighting factor for evaluating performance.

However, the use of post-occupancy evaluation (POE) and evaluations of user satisfaction have become increasingly common, and it are starting to be linked to the performance evaluation of buildings [183]. An EU-funded project made the link between occupant satisfaction and performance clear [184], making it even more important to guarantee that indoor environmental quality indicators are met before evaluating efficiency indicators. Goçer and colleagues describe the use of POE to improve building performance feedback [185], while Yang and colleagues review the links between occupant satisfaction and energy performance and identify several areas for improvement and further research [186]. Menezes and

ACCEPTED MANUSCRIPT

colleagues use the results of POE directly, to improve the prediction accuracy of other energy performance models and evaluations [187].

Both energy audits and post-occupancy evaluations can identify performance issues, maintenance problems and other aspects that may have negative impacts on indoor environmental quality, and may be key to making improvements and maintaining performance.

Overall, it seems to be clear that evaluations of environmental quality and user satisfaction should be a key part of building performance evaluation, but have been largely marginalised, perhaps because of the difficulty of detailed study. Instead, sets of assumptions on user behaviour, environmental conditions and operational schedules are used as the basis of all of the evaluation methodologies shown above.

ACCEPTED MANUSCRIPT

Public evaluation systems and standards

The evaluation techniques described in this paper are applied in order to classify, rate and rank building energy performance. Usually, this means that benchmark performance levels are defined so that buildings can be shown to have good or poor energy performance, and actions taken accordingly. In most cases, the final aim is to guarantee performance levels and energy consumption reductions in new and existing buildings. This chapter describes the practical application of the technical benchmarking methodologies described above, through voluntary and mandatory evaluation systems and standards.

The aim of the rating might be to drive user preference when buying or renting a building, to assess whether a building's energy use is within range (ie. to see whether something is "wrong"), to identify outliers amongst a group of buildings with similar function, or to detect and diagnose saving potential. Public evaluation systems for energy performance may be used in new and existing buildings in the following ways:

Enforced compliance with minimum standards (such as building regulations or building codes)

Policies to penalise poor energy performance or reward good energy performance

Mandatory labelling of building energy performance

Transparency and disclosure programmes

Voluntary sustainability or energy certification schemes

Voluntary internal benchmarking and evaluation of improvement opportunities

Sometimes the same tool or evaluation system can be used in several different ways. For example, Energy Star Portfolio Manager is used for several transparency and disclosure programmes in the USA [188], as a baseline for the LEED certification of existing buildings [189] and as a tool for voluntary benchmarking [190].

It is important that the benchmarks or calculation methodologies for public evaluation systems are transparent, reliable and simple enough to be used effectively. There will generally be a tension between the accuracy of the prediction (requiring more data collection and complex calculations) and the usability of the system, ensuring that it is as simple as possible (but no simpler). Within the context of a wide range of policies, standards and efforts to improve building energy performance, different calculation methodologies will be applicable in different situations.

Usually, the development of national benchmarks is based on the use of a large database, such as the Commercial Building Energy Consumption Survey (CBECS) in the USA [191], and similar energy consumption surveys in California [192] and Canada [193]. The UK has recently started development of a National Energy Efficiency Database [194], while Nikolaou and colleagues have proposed the use of simulated virtual building datasets to overcome the lack of such a database in Greece [112].

ACCEPTED MANUSCRIPT

In theory, a performance-based rating approach should be based (and is in almost all other industries) on “requirement setting” and “compliance” checking by measurements. Most performance based approaches in buildings still rely on compliance checking by calculations, although this is slowly changing through updates in regulations such as the EU-EPBD.

Benchmarking tools

A series of benchmarking tools and methods have been developed and publicised. Usually these are available through online portals, which may be accessed by registered users, by qualified professionals or by any user with building energy information.

These systems often use a statistical or empirical benchmark for comparison and evaluation of building energy performance. The tools may be used as a way of demonstrating compliance with certification schemes, but if they are recognised to have a high degree of reliability, they may also form the basis of other policies.

The most widely used system is Energy Star Portfolio Manager, which uses benchmarks developed by multiple linear regressions to compare buildings of several typologies, based on different characteristics [14]. The benchmarks for comparison are developed from CBECS data, and give a score from 0-100, based on the relative position of the building compared to its peer group. Weather corrections are carried out with heating and cooling degree-days. The benchmarks are specifically designed to include variables explaining how a building operates, and to exclude technology factors and market conditions.

Matson and Piette compare Energy Star to the California-specific programme Cal-Arch and find both to be adequate for benchmarking buildings in California [195]. Additional benchmarking tools are described in Table 6.

National standards and labelling

National standards for energy performance in buildings were first developed in the form of building codes, with prescriptive requirements for certain building characteristics, aimed at maintaining minimum energy performance levels. However, as buildings become more complex these standards often use simulation to demonstrate compliance [196]. Regulations such as the European Energy Performance of Buildings Directive (EU-EPBD) have required countries to publish more detailed energy performance requirements and start to label energy performance levels in both new and existing buildings. The national standards and labelling schemes are often seen as a key part of public policy on energy efficiency and carbon dioxide emission reductions.

Corgnati and Corrado give a detailed introduction to the concepts of energy monitoring and labelling [197]. Research projects such as EPLabel developed the technical bases for labelling systems in Europe [11], which use both Asset Ratings (or calculated ratings) and Operational Ratings (or measured ratings), as defined in Table 1 and in the ISO standard [13]. Generally,

ACCEPTED MANUSCRIPT

Asset Ratings are developed by engineering calculations following prescribed methodologies such as the Simplified Building Energy Model (SBEM) or by simulation, following specific procedures. Operational Ratings compare measured energy consumption to previously developed statistical benchmarks, allowing for some correction factors. (An exception to this is the framework developed by Melo and colleagues for the application of ANN for building energy labelling [153].) The final step of energy labelling involves packaging the information from the asset or operational ratings into a clearly displayed certificate which is visibly displayed or otherwise made accessible and communicates the essential energy performance information to users.

Cohen and Bordass provide a detailed history of the development of operational rating systems, comparing policy frameworks with a specific focus on the UK, and noting the relative success of the Australian NABERS system [198]. Bannister discusses NABERS in more detail, and charts its development over 12 years [199].

A review on labelling systems by Rajagopalan and Leung Tony details the progress in building energy labelling in the EU and seven other countries [200]. International studies by the International Partnership for Energy Efficiency Collaboration (IPEEC) and the Institute for Market Transformation (IMT) provide detailed evaluations of the energy rating systems in 14 and 10 countries respectively, comparing the type of assessment, buildings affected and baselines used [6], [201]. Generally, asset ratings and operational ratings are calculated separately. Although the Chinese MOHURD rating system incorporates both, it has yet to create a full integration between the two [202].

Increasingly, transparency and disclosure programmes are implemented within specific jurisdictions. One of the most successful is Local Law 84, a part of the Greater Greener Buildings Plan, which obliges all buildings over 5,000m^2 in the city of New York to benchmark their energy performance using Energy Star Portfolio Manager and publicly publish the results [203]. As the benefits become clearer, several other cities and states across the USA have followed suit, and the Department of Energy has published guidance for such programmes [188]. More information on the development of building labelling and rating is published by the IMT on the BuildingRating.org portal [204].

Certification and points-based rating systems

Voluntary certifications have become increasingly popular as a way of driving sustainability in the built environment. The BREEAM programme was launched in the UK in the 1990 and popularised a points-based rating system, with different categories of sustainability given different weightings and opportunities for gaining points; the total number of points indicates a sustainability rating level (there are also mandatory, prescriptive requirements) [205]. BREEAM is widely used in the UK and in Europe. The US Green Building Council developed the Leadership in Energy and Environmental Design (LEED) certification, which uses some of the same principles as BREEAM but developed criteria through a collaborative approach and has very successfully marketed worldwide as a tool for standardising sustainability evaluations in

ACCEPTED MANUSCRIPT

portfolios [189]. In addition to these international programmes, CIBSE identifies 35 country-specific certification schemes in 25 countries [59].

Generally, the points-based rating systems give a high priority to energy performance in the weighting scheme. New buildings are often rated with simulation methodologies, using notional buildings simulated under the same usage conditions to compare energy performance (see for example, Appendix G of ASHRAE 90.1 [196]). Existing buildings are often rated using the benchmarking tools described above; for example, the LEED certification for operation and maintenance of existing buildings uses Energy Star Portfolio Manager to define the performance levels of the buildings under evaluation.

Raslan and Davies note that modelling for compliance is still very dependent on the reliability and accuracy of data, user skill, applicability of the tool and calculation method used, and that more work is needed for this method to produce credible results [32]. These findings are echoed by Schwartz and Raslan, who show that the same building modelled for LEED and BREEAM in different simulation software gives significantly different results [206].

More alarmingly, analyses of in-use data on LEED-certified buildings do not always show clear, significant performance improvements compared to normal buildings [207], [208]. (These results echo other findings on the performance gap, which were examined in detail in the Simulation chapter.) As LEED certification has been increasingly adopted as a requirement for construction permits or tax credits in different areas, failure to meet energy performance levels has actually led to litigation in the USA regarding the energy performance of buildings (see for example [209]). As LEED was designed as a voluntary programme, many of the requirements are self-reported so it may be difficult to certify some key aspects of design or construction. Accordingly, the International Green Construction Code (IgCC) has been developed as an alternative to LEED certification that can be enforced in the same way as a building code [210].

Public databases and review

The availability of large amounts of data (often public) on building energy performance made available through labelling schemes, certifications and accredited benchmarking programmes has made it possible to carry out reviews of performance and even update benchmarks using this data.

The data from tens of thousands of UK Display Energy Certificates (DECs) has been made available for some research projects. Thus Bruhns and colleagues were able to perform a review of the CIBSE benchmarks and identify trends in building energy performance: although many benchmarks were still reasonably accurate, in several typologies, since the benchmarks were published, thermal energy consumption has reduced but electrical energy consumption has increased [211]. Armitage and colleagues use DEC data to map energy consumption in public sector office buildings, and note that a focus on unregulated (electrical) loads will be crucial in ensuring performance improvement to meet national emissions targets [212].

ACCEPTED MANUSCRIPT

The data released on New York's buildings has been used by Hsu to evaluate the effectiveness and impact of disclosure laws [137] and by Scofield to further evaluate the effectiveness of LEED certification in reducing energy consumption [213].

Summary of rating systems

This section briefly summarises (in Table 6) the principal public evaluation systems described above, separates them by type and briefly describes their application and the methodology used for benchmarking. It is not an exhaustive review, and serves principally to illustrate the diversity of schemes and programmes in existence and the interaction between them.

Evaluation system	Type	Application	Benchmarking methodology	Country	Website
ASHRAE 90.1	National regulation, minimum standard	New buildings in the USA. Defines notional building which is baseline for LEED and some other certification schemes.	Prescriptive or simulation with comparison to notional building.	USA	www.ashrae.org
New York LL84	Mandatory transparency scheme	Required annual public benchmarking for large buildings in New York City	Energy Star Portfolio Manager (CBECS database and correction factors)	USA	http://www.nyc.gov/html/gbee/html/plan/plan.shtml
Display Energy Certificates	Labelling (asset ratings)	Evaluation of efficiency for existing buildings, annually renewed label	Comparison with CIBSE TM46 benchmarks	UK	https://www.gov.uk/government/publications/display-energy-certificates-and-advisory-reports-for-public-buildings
Energy Performance Certificates	Labelling (operational ratings)	New buildings or buildings that are to be sold or let	Engineering calculation (SBEM) or simulation, comparing to a notional building in both cases	UK	https://www.ndepcregister.com/

PBE Edifica	Labelling (asset ratings)	New buildings and major renovations	Prescriptive or simulation with comparison to notional buildings	Brazil	http://pbeedifica.com.br/
Energy Star Portfolio Manager	Benchmarking tool	Voluntary benchmarking and local disclosure programmes	CBECS database and correction factors	USA	http://portfoliomanager.energystar.gov/
EnergyIQ	Benchmarking tool	Voluntary benchmarking and identification of energy saving actions	CBECS database and correction factors	USA	http://energyiq.lbl.gov/
ASHRAE bEQ	Labelling (asset and operational ratings)	Voluntary labelling with detailed information to inform and drive performance improvements	ASHRAE Standard 100, using CBECS data	USA	http://buildingenergyquotient.org/
CBCS DEO	Benchmarking tool	Voluntary benchmarking	Comparison with national benchmarks	Brazil	http://benchmarkingenergia.cbcs.org.br/
CarbonBuzz	Benchmarking tool	Voluntary comparison with other buildings, mapping the performance gap between design expectations and in-use performance	Comparison with national benchmarks, as-designed levels and an internal database	UK	http://www.carbonbuzz.org/
US DOE Building Performance Database	Benchmarking tool	Voluntary mapping of building performance and retrofit potential, based on other buildings of similar characteristics	Comparison with large datasets of building performance information, filtering by desired characteristics	USA	https://bpd.lbl.gov/

China 3 Star Building Energy Efficiency Evaluation	Labelling (asset and operational ratings)	Labelling scheme that incorporates both asset and operational ratings, mandatory in some types of buildings	Comparison to notional regulated energy use by simulation or measurement, as well as required items and optional items (prescriptive)	China	http://www.buildingrating.org/jurisdiction/China
NABERS	Labelling (operational ratings)	Office buildings in Australia, particularly with respect to Landlords' services	Comparison with national benchmarks	Australia	http://www.nabers.gov.au

Table 6 – Examples of building energy performance rating systems and schemes

ACCEPTED MANUSCRIPT

Discussion

Selection of evaluation method

This paper has discussed a wide range of methodologies for energy performance evaluation in buildings. The methodologies have been grouped by type, a standardised terminology has been applied and their principal benefits and drawbacks have been observed. Finally, the practical application of benchmarking in buildings has been discussed.

In general, statistical benchmarking is effective for identifying the energy performance level of a building, but does not give a detailed understanding of the underlying mechanisms and reasons for the performance. Machine learning methods also operate without understanding of the physical characteristics of buildings, which may limit their effectiveness in identifying energy performance improvements. Engineering calculation methodologies are effective for identifying the improvement potential from specific retrofit measures, but do not have high levels of reliability. Simulation models can be highly accurate, but this requires extensive calibration and large amounts of building detail, which limits their application. Table 7 compares these methodologies.

Method	Inputs needed	Accuracy	Applications	Restrictions
Engineering calculations	Simplified building information	Variable	Design. End-use evaluations. Highly flexible.	Limited accuracy.
Simulation	Detailed building information	High	Design. Compliance. Complex buildings. Cases where high accuracy is necessary.	Dependent on user skill and significant data collection.
Statistical	Dataset of existing buildings	Average	Benchmarking systems. Simple evaluations.	Dependent on statistical data. Limited accuracy.
Machine Learning	Large dataset	Average to high	Buildings with highly detailed data collection. Complex problems with many parameters.	Models construction is complicated. Do not consider direct physical characteristics.

Table 7 - Comparison of principal building energy performance evaluation methodologies

As Hong notes [21], there is generally no attempt to calibrate top-down and bottom-up benchmarking methods, which weakens both. In addition, there is generally no attempt made to reconcile predicted and measured energy performance, despite much research on the subject. An effective benchmarking framework would be able to use the insights from statistical or machine learning evaluations to rank overall building energy efficiency, while

ACCEPTED MANUSCRIPT

using simple engineering calculations to map end-uses and identify the performance of individual systems and improvement potential.

Leaman and colleagues discuss the practical application of building performance evaluation, based on many years of experience, and criticise the purely academic approach to evaluation,. They discuss the use of a variety of methods in order to create robust evaluation methodologies, in a manner that is clearly different to the purely scientific, laboratory-based methodologies that are cited in many of the papers in this review [183]. Buildings are real-world problems, with highly complex interactions between systems, users and the environment.

Thus the key consideration of building evaluation techniques is fitness for purpose. The available data and aim of the evaluation must be clear in order to define the appropriate allocation of resources and hence modelling approach for developing benchmarks and evaluating performance. This means the use of rapid, simplistic models in some cases, and detailed simulation or energy auditing in other situations. An ideal benchmarking model would be as simple as possible in order to evaluate performance, but no simpler.

In the development of national evaluation and benchmarking systems, it is important to evaluate the cost (in data collection and man-hours of work) of the development of the evaluation system, as well as the cost of reliably evaluating building performance using the system.

Accuracy and confidence intervals should be discussed more in the results that are presented, as simpler methods would be likely to show greater uncertainty bands. It should be possible to define fit-for-purpose accuracy demands, depending on the evaluation needs, for example:

A comparison of design options does not need to estimate total real energy consumption, just to compare the relative consumption under different scenarios.

A simple compliance framework requires high accuracy under standard conditions.

An energy performance contract or implementation of high-cost energy saving measures require high levels of accuracy in identifying real energy consumption.

Explanations for poor building performance

Although the performance gap between predicted and measured energy performance has been discussed in some detail, there is still a lack of understanding of some reasons for poor energy performance in buildings.

Increased complexity in buildings systems (paradoxically, often driven by sustainability certifications and rating schemes) has increased the challenges of managing energy performance and may lead to higher maintenance costs, rapid obsolescence of systems, constant need for re-training of operations staff and deterioration of performance. The consistently high financial returns of building commissioning and retro-commissioning [214] show that buildings are not operating at their full potential. De Wilde and colleagues carried

ACCEPTED MANUSCRIPT

out one of the only studies identified during this review that considers long-term degradation of building energy performance, explicitly considering some effects of poor maintenance [215], but not the operational inefficiencies of the building (the methodology is computationally intensive).

There are several examples of more complex buildings showing inferior energy performance. Escrivá-Escrivá notes that "...buildings with centralised air-conditioning systems have poorer energy behaviour than buildings with split-systems. This is because these centralised systems require a higher energy input, do not adapt easily to external temperatures, and users do not efficiently manage these systems." This is despite the fact that central air conditioning systems are generally considered to be more efficient than split systems, and are consistently shown to be so in simulation studies. Bordass and colleagues also noted more complex systems generally ran "more liberally and wastefully", as well as noting a factor of six variation of carbon dioxide emissions between buildings with similar end-uses [69]. This is consistent with the practical experience of the authors of this paper.

Overall, it has been recognised since the PROBE studies in the 1990s that there is a need to explain poor performance and evaluate it in real terms [216], but this review shows that there is still much work to be done in this area.

Comfort conditions and energy consumption

As Corgnati notes, "It is useless to express energy consumption for microclimate control in a building without relating such consumption to the microclimatic quality assessed for the environment" [197]. Olofsson notes that energy efficiency and user comfort might conflict, and resolves this by suggesting that in order to be evaluated by a rating system, a building must "supply the amenities and features typical for that kind of building. Thus, an office must provide around 60 hours a week of suitably conditioned air, lighting and equipment" [10].

However, none of the energy performance methods evaluated proposed a methodology for linking environmental quality to energy performance. Thus, all of the building evaluation methodologies (with the exception of the user satisfaction questionnaires) must assume that buildings work properly and provide comfort conditions for users. This is generally not the case, and makes it extremely difficult to compare different types of building with the same occupancy. For example, in Brazil's climate it is common to find office buildings that are fully air-conditioned with central air conditioning systems, as well as mixed-mode office buildings (with split air conditioning units in some or all areas) and some naturally ventilated offices. In this context, energy performance comparisons between offices become meaningless unless they are first separated by the type of air conditioning system.

There is clearly a gap in the literature relating energy performance evaluations to observed environmental quality and user satisfaction.

Building energy use as service provision

ACCEPTED MANUSCRIPT

In many areas of energy performance, it is possible to define a theoretical optimal energy requirement for carrying out a type of work; the real energy efficiency level is then defined as the ratio between the actual energy consumption and the theoretical minimum energy requirement (for example, heat pumps, car engines and wind turbines can all use theoretical maximum performance levels to evaluate their real-world efficiency). In buildings, most energy is consumed in order to provide a few basic services: thermal comfort, fresh air, lighting, information & communication technology, transport and hot and cold water.

Thus in theory, it should be possible to consider building energy use as a service provision, and hence define maximum or typical efficiency levels. This possibility has been investigated by Pérez-Lombard in some papers, by defining levels of service and considering HVAC efficiency indicators based on theoretical performance [53], [217], as well as being considered in Federspiel's evaluation of laboratory energy efficiency [51]. However, aside from these authors, the possibility of linking energy services and real, physical systems with simple benchmarking methodologies has largely been ignored.

It would seem that there is great potential for the development of benchmarking and evaluation systems that are based on theoretical minimum performance levels of well-understood physical systems for the provision of levels of energy service. This could be a way of developing methods for rapid performance evaluation that take into account the differing comfort levels mentioned above.

Further work

The authors cited in this review consistently agree on the need for wider debate, transparency and disclosure regarding building energy information, especially given the importance of the buildings sector in global energy consumption and carbon dioxide emissions. As noted by Zhao and Magoulès, there is still a need for "effective, robust, reliable and efficient prediction models" [31] in order to be able to effectively evaluate the energy performance of buildings.

Based on the review and mapping of the sector carried out in this paper, there should be further development of interactions between the main modelling techniques, combining the strengths of different approaches in order to develop efficient and useful evaluation systems.

A focus of future work should be the effective evaluation of energy performance in tropical or developing countries, where many buildings operate in mixed-mode and do not meet the comfort criteria laid out in international standards.

ACCEPTED MANUSCRIPT

References

[1] International Energy Agency, “Total energy use in buildings Analysis and evaluation methods,” 2013.

[2] International Energy Agency, “Capturing the Multiple Benefits of Energy Efficiency,” 2014.

[3] J. M. Cullen, J. M. Allwood, and E. H. Borgstein, “Reducing energy demand - what are the practical limits,” *Environ. Sci. Technol.*, vol. 45, pp. 1711–1718, 2011.

[4] L. Pérez-Lombard, J. Ortiz, and C. Pout, “A review on buildings energy consumption information,” *Energy Build.*, vol. 40, no. 3, pp. 394–398, 2008.

[5] H. Davies, “Tracing the continuing development of Part L,” *Modern Building Services*, 2013. [Online]. Available: http://www.modbs.co.uk/news/fullstory.php/aid/12062/Tracing_the_continuing_development_of_Part_L.html. [Accessed: 20-Mar-2016].

[6] IPEEC Building Energy Efficiency Tasktroup, “Building Energy Rating Schemes: Assessing Issues and Impacts,” Paris, 2014.

[7] R. Ruparathna, K. Hewage, and R. Sadiq, “Improving the energy efficiency of the existing building stock: A critical review of commercial and institutional buildings,” *Renew. Sustain. Energy Rev.*, vol. 53, pp. 1032–1045, 2016.

[8] R. Thomas, *Environmental design : an introduction for architects and engineers*, 3rd editio. Ta, 2006.

[9] D. Mumovic and M. Santamouris, Eds., *A Handbook of Sustainable Building Design and Engineering*. Earthscan, 2009.

[10] T. Olofsson, A. Meier, and R. Lamberts, “Rating the Energy Performance of Buildings,” *Int. J. Low Energy Sustain. Build.*, vol. 3, 2004.

[11] R. Cohen, “EPLabel: A programme to deliver energy certificates for display in Public buildings across Europe within a harmonising framework,” 2007.

[12] R. Cohen, W. Bordass, and J. Field, “Fixed and Customised Benchmarks for Building Energy Performance Certificates based on Operational Ratings,” *Energy Sustain. Dev.*, pp. 3–8, 2007.

[13] ISO, “Energy performance of buildings - Assessment of overall energy performance - ISO 16346:2013,” vol. Energy per, 2013.

[14] US Environmental Protection Agency, “ENERGY STAR ® Performance Ratings Technical Methodology,” 2010.

ACCEPTED MANUSCRIPT

[15] T. Nikolaou, D. Kolokotsa, and G. Stavrakakis, “Review on methodologies for energy benchmarking, rating and classification of buildings,” *Adv. Build. Energy Res.*, vol. 5, no. 1, pp. 53–70, 2011.

[16] W. Chung, “Review of building energy-use performance benchmarking methodologies,” *Appl. Energy*, vol. 88, no. 5, pp. 1470–1479, 2011.

[17] L. Pérez-Lombard, J. Ortiz, R. González, and I. R. Maestre, “A review of benchmarking, rating and labelling concepts within the framework of building energy certification schemes,” *Energy Build.*, vol. 41, no. 3, pp. 272–278, 2009.

[18] U.S. Department of Energy’s Commercial Building Initiative, “Buildings Performance Metrics Terminology,” 2009. [Online]. Available: http://energy.gov/sites/prod/files/2013/12/f5/metrics_terminology_20090203.pdf. [Accessed: 08-Mar-2016].

[19] P. Bannister and A. Hinge, “Empirical Benchmarking of Building Performance,” *ACEEE Summer Study Energy Effic. Build.*, pp. 15–26, 2006.

[20] S. Wang, C. Yan, and F. Xiao, “Quantitative energy performance assessment methods for existing buildings,” *Energy Build.*, vol. 55, pp. 873–888, 2012.

[21] S.-M. Hong, G. Paterson, E. Burman, P. Steadman, and D. Mumovic, “A comparative study of benchmarking approaches for non-domestic buildings: Part 1 – Top-down approach,” *Int. J. Sustain. Built Environ.*, vol. 2, no. 2, pp. 119–130, 2013.

[22] E. Burman, S.-M. Hong, G. Paterson, J. Kimpian, and D. Mumovic, “A comparative study of benchmarking approaches for non-domestic buildings: Part 2 - Bottom-up approach,” *Int. J. Sustain. Built Environ.*, vol. 2, no. 2, pp. 119–130, 2013.

[23] Z. Li, Y. Han, and P. Xu, “Methods for benchmarking building energy consumption against its past or intended performance: An overview,” *Appl. Energy*, vol. 124, pp. 325–334, 2014.

[24] ASHRAE, “ENERGY ESTIMATING AND MODELING METHODS,” in *ASHRAE Handbook - Fundamentals*, ASHRAE, 2013.

[25] S. Kinney and M. A. Piette, “Development of a California Commercial Building Benchmarking Database,” *ACEEE Summer Study Energy Effic. Build.*, 2002.

[26] X. Li and J. Wen, “Review of building energy modeling for control and operation,” *Renew. Sustain. Energy Rev.*, vol. 37, pp. 517–537, 2014.

[27] R. Kramer, J. van Schijndel, and H. Schellen, “Simplified thermal and hygric building models: A literature review,” *Front. Archit. Res.*, vol. 1, no. 4, pp. 318–325, 2012.

[28] D. Coakley, P. Raftery, and M. Keane, “A review of methods to match building energy simulation models to measured data,” *Renew. Sustain. Energy Rev.*, vol. 37, pp. 123–141, 2014.

ACCEPTED MANUSCRIPT

[29] Y. Zhang, Z. O'Neill, B. Dong, and G. Augenbroe, "Comparisons of inverse modeling approaches for predicting building energy performance," *Build. Environ.*, vol. 86, pp. 177–190, 2015.

[30] M. S. Al-Homoud, "Computer-aided building energy analysis techniques," *Build. Environ.*, vol. 36, no. 4, pp. 421–433, 2001.

[31] H. Zhao and F. Magoulès, "A review on the prediction of building energy consumption," *Renew. Sustain. Energy Rev.*, vol. 16, no. 6, pp. 3586–3592, 2012.

[32] R. Raslan and M. Davies, "Results variability in accredited building energy performance compliance demonstration software in the UK: An inter-model comparative study," *J. Build. Perform. Simul.*, vol. 3, no. 1, pp. 63–85, 2010.

[33] ISO, "Thermal performance of buildings — Calculation of energy use for space heating and cooling - ISO 13790:2005," 2005.

[34] Y.-J. Kim, S.-H. Yoon, and C.-S. Park, "Stochastic comparison between simplified energy calculation and dynamic simulation," *Energy Build.*, vol. 64, pp. 332–342, 2013.

[35] M. Gustafsson, F. Ochs, S. Birchall, G. Dermentzis, C. Bales, C. Dipasquale, and T. Calabrese, "iNSPiRe: D2.2 Report on Auditing tool for assessment of building needs," 2015.

[36] BRE, "UK National Calculation Methodology for Non-Domestic Buildings." [Online]. Available: http://www.uk-ncm.org.uk/. [Accessed: 20-Mar-2016].

[37] E. H. Borgstein and R. Lamberts, "Developing energy consumption benchmarks for buildings: Bank branches in Brazil," *Energy Build.*, vol. 82, pp. 82–91, 2014.

[38] CIBSE, "CIBSE Guide A: Environmental Design," 2015.

[39] iSERV, "http://www.iservcmb.info/," 2016. .

[40] I. Knight, "Assessing electrical energy use in HVAC systems," *REHVA J.*, no. January, 2012.

[41] J. W. Field, J. Soper, P. G. Jones, and W. T. Bordass, "Energy Performance of Occupied Non Domestic Buildings: Assessment by analysing end-use consumptions," *Build. Serv. Eng. Res. Technol.*, vol. 18, no. 1, pp. 39–46, 1997.

[42] CIBSE, "TM22 : Energy assessment and reporting method," 2006.

[43] A. C. Menezes, N. Nkonge, A. Cripps, D. Bouchlaghem, and R. Buswell, "Review of Benchmarks for Small Power Consumption in Office Buildings," *CIBSE ASHRAE Tech. Symp. 2012*, no. April, pp. 1–19, 2012.

[44] CIBSE, "CIBSE Guide D: Transportation systems in buildings," 2010.

[45] ASHRAE, "Procedures for Commercial Building Energy Energy Audits," 2011.

ACCEPTED MANUSCRIPT

[46] R. Cohen, M. Standeven, B. Bordass, and A. Leaman, “Assessing building performance in use 1: the Probe process,” *Build. Res. Inf.*, vol. 29, no. 2, pp. 85–102, 2001.

[47] CIBSE, “TM54: Evaluating operational energy performance of buildings at the design stage,” 2013.

[48] B. Bordass, “Energy Consumption Guide 19: Energy Use in Offices,” 2000.

[49] CIBSE, “TM46: Energy benchmarks,” 2008.

[50] G. Escrivá-Escrivá, C. Álvarez-Bel, and E. Peñalvo-López, “New indices to assess building energy efficiency at the use stage,” *Energy Build.*, vol. 43, no. 2–3, pp. 476–484, 2011.

[51] C. Federspiel, Q. Zhang, and E. Arens, “Model-based benchmarking with application to laboratory buildings,” *Energy Build.*, vol. 34, no. 3, pp. 203–214, 2002.

[52] N. Fumo, P. Mago, and R. Luck, “Methodology to estimate building energy consumption using EnergyPlus Benchmark Models,” *Energy Build.*, vol. 42, no. 12, pp. 2331–2337, 2010.

[53] L. Pérez-Lombard, J. Ortiz, I. R. Maestre, and J. F. Coronel, “Constructing HVAC energy efficiency indicators,” *Energy Build.*, vol. 47, pp. 619–629, 2012.

[54] L. R. Liu, J. J. Gu, and J. Liu, “Analysis of operational energy intensity for central air conditioning systems with water-cooled chiller by decomposition method,” *Energy Build.*, vol. 93, pp. 154–159, 2015.

[55] C. Yan, S. Wang, F. Xiao, and D. Gao, “A multi-level energy performance diagnosis method for energy information poor buildings,” *Energy*, vol. 83, pp. 189–203, 2015.

[56] R. Liddiard, “Room-scale profiles of space use and electricity consumption in non-domestic buildings,” *Build. Res. Inf.*, vol. 42, no. 1, pp. 72–94, 2014.

[57] W. Zhaoxia and D. Yan, “An occupant-based energy consumption prediction model for office equipment,” *Energy Build.*, vol. 109, pp. 12–22, 2015.

[58] J. L. M. Hensen and R. Lamberts, Eds., *Building performance simulation for design and operation*. Routledge, 2011.

[59] CIBSE, “AM11: Building Performance Modelling,” 2015.

[60] ASHRAE, “Guideline 14 - Measurement of Energy, Demand, and Water Savings,” vol. 2014. 2014.

[61] D. B. Crawley, J. W. Hand, M. Kummert, and B. Griffith, “Contrasting the capabilities of building energy performance simulation programs,” *Build. Environ.*, vol. 43, pp. 661–673, 2008.

[62] Sefaira Ltd, “Sefaira,” 2016. [Online]. Available: http://sefaira.com/. [Accessed: 20-Mar-2016].

[63] UFSC, “S3E,” 2016. [Online]. Available: http://s3e.ufsc.br/. [Accessed: 20-Mar-2016].

ACCEPTED MANUSCRIPT

[64] DesignBuilder Software Ltd, "DesignBuilder," 2016. [Online]. Available: http://designbuilder.co.uk/. [Accessed: 20-Mar-2016].

[65] DOE, "Building Energy Software Tools directory," 2016. [Online]. Available: http://www.buildingenergysoftwaretools.com. [Accessed: 23-Mar-2016].

[66] V. S. K. V. Harish and A. Kumar, "A review on modeling and simulation of building energy systems," *Renew. Sustain. Energy Rev.*, vol. 56, pp. 1272–1292, 2016.

[67] M. Trčka and J. L. M. Hensen, "Overview of HVAC system simulation," *Autom. Constr.*, vol. 19, no. 2, pp. 93–99, 2010.

[68] A. Malkawi and G. Augenbroe, Eds., *Advanced Building Simulation*. Spon Press, 2003.

[69] B. Bordass, R. Cohen, M. Standeven, and A. Leaman, "Assessing building performance in use 3: energy performance of the Probe buildings," *Build. Res. Inf.*, vol. 29, no. 2, pp. 114–128, 2001.

[70] P. De Wilde, "The gap between predicted and measured energy performance of buildings: A framework for investigation," *Autom. Constr.*, vol. 41, pp. 40–49, 2014.

[71] C. van Dronkelaar, M. Dowson, C. Spataru, and D. Mumovic, "A Review of the Regulatory Energy Performance Gap and Its Underlying Causes in Non-domestic Buildings," *Front. Mech. Eng.*, vol. 1, no. January, pp. 1–14, 2016.

[72] D. Daly, P. Cooper, and Z. Ma, "Understanding the risks and uncertainties introduced by common assumptions in energy simulations for Australian commercial buildings," *Energy Build.*, vol. 75, pp. 382–393, 2014.

[73] P. G. Tuohy and G. B. Murphy, "Closing the gap in building performance: learning from BIM benchmark industries," *Archit. Sci. Rev.*, vol. 58, no. 1, pp. 47–56, 2015.

[74] M. Heidarinejad, M. Dahlhausen, S. Mcmahon, C. Pyke, J. Srebric, T. U. S. Green, and B. Council, "Building classification based on simulated annual results: towards realistic building performance expectations," *13th Conf. Int. Build. Perform. Simul. Assoc.*, pp. 1706–1713, 2013.

[75] C. Demanuele, T. Tweddell, and M. Davies, "Bridging the gap between predicted and actual energy performance in schools," *World Renew. Energy Congr. XI*, no. September, pp. 1–6, 2010.

[76] a. Dasgupta, a. Prodromou, and D. Mumovic, "Operational versus designed performance of low carbon schools in England: Bridging a credibility gap," *HVAC&R Res.*, vol. 18, no. 1–2, pp. 37–41, 2012.

[77] L. E. Fedoruk, R. J. Cole, J. B. Robinson, and A. Cayuela, "Learning from failure: understanding the anticipated–achieved building energy performance gap," *Build. Res. Inf.*, vol. IN PRESS, no. August, pp. 1–15, 2015.

[78] E. M. Ryan and T. F. Sanquist, "Validation of building energy modeling tools under idealized and realistic conditions," *Energy Build.*, vol. 47, pp. 375–382, 2012.

ACCEPTED MANUSCRIPT

[79] S. O. Jensen, “Validation of building energy simulation programs" a methodology *,” *Energy Build.*, vol. 22, pp. 133–144, 1995.

[80] J. Neymark, R. Judkoff, G. Knabe, H. T. Le, M. Dürig, A. Glass, and G. Zweifel, “Applying the building energy simulation test (BESTEST) diagnostic method to verification of space conditioning equipment models used in whole-building energy simulation programs,” *Energy Build.*, vol. 34, no. 9, pp. 917–931, 2002.

[81] R. H. Henninger, M. J. Witte, and D. B. Crawley, “Analytical and comparative testing of EnergyPlus using IEA HVAC BESTEST E100 – E200 test suite,” vol. 36, no. January 2002, pp. 855–863, 2004.

[82] ASHRAE, “Standard 140-2011 - Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs.” ASHRAE, 2011.

[83] I. Knight, S. Stravoravdis, and S. Lasvaux, “Predicting operational energy consumption profiles - Findings from detailed surveys and modelling in a UK educational building compared to measured consumption,” *Int. J. Vent.*, vol. 7, no. 1, pp. 49–57, 2008.

[84] M. Ahmad and C. H. Culp, “Uncalibrated Building Energy Simulation Modeling Results,” *HVAC&R Res.*, vol. 12, no. 4, pp. 1141–1155, 2006.

[85] N. Li, Z. Yang, B. Becerik-Gerber, C. Tang, and N. Chen, “Why is the reliability of building simulation limited as a tool for evaluating energy conservation measures?,” *Appl. Energy*, vol. 159, pp. 196–205, 2015.

[86] P. Raftery, M. Keane, and J. O’Donnell, “Calibrating whole building energy models: An evidence-based methodology,” *Energy Build.*, vol. 43, no. 9, pp. 2356–2364, 2011.

[87] E. Fabrizio and V. Monetti, “Methodologies and Advancements in the Calibration of Building Energy Models,” *Energies*, vol. 8, no. 4, pp. 2548–2574, 2015.

[88] T. A. Reddy, I. Maor, and C. Panjapornpon, “Calibrating Detailed Building Energy Simulation Programs with Measured Data-- Part I: General Methodology (RP-1051).,” *HVAC&R Res.*, vol. 13, no. 2, pp. 221–241, 2007.

[89] A. Pedrini, F. S. Westphal, and R. Lamberts, “A methodology for building energy modelling and calibration in warm climates,” *Build. Environ.*, vol. 37, no. 8–9, pp. 903–912, 2002.

[90] T. A. Reddy, I. Maor, and C. Panjapornpon, “Calibrating Detailed Building Energy Simulation Programs with Measured Data—Part II: Application to Three Case Study Office Buildings (RP-1051),” *HVAC&R Res.*, vol. 13, no. 2, pp. 243–265, 2007.

[91] A. H. Neto and F. A. S. Fiorelli, “Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption,” *Energy Build.*, vol. 40, no. 12, pp. 2169–2176, 2008.

[92] G. Mustafaraj, D. Marini, A. Costa, and M. Keane, “Model calibration for building energy efficiency simulation,” *Appl. Energy*, vol. 130, pp. 72–85, 2014.

ACCEPTED MANUSCRIPT

[93] M. Royapoor and T. Roskilly, “Building model calibration using energy and environmental data,” *Energy Build.*, vol. 94, pp. 109–120, 2015.

[94] Oak Ridge National Laboratory, “AUTOTUNE,” 2016. [Online]. Available: http://rsc.ornl.gov/autotune/. [Accessed: 23-Mar-2016].

[95] Efficiency Valuation Organisation, “Concepts and Options for determining Energy and Water Savings,” in *International Performance Measurement and Verification Protocol (IPMVP)*, vol. 1, no. January, EVO, 2012.

[96] S. de Wit, “Uncertainty in building simulation,” in *Advanced Building Simulation*, A. Malkawi and G. Augenbroe, Eds. Spon Press, 2003.

[97] I. MacDonald and P. Strachan, “Practical application of uncertainty analysis,” *Energy Build.*, vol. 33, no. 3, pp. 219–227, 2001.

[98] W. Tian, “A review of sensitivity analysis methods in building energy analysis,” *Renew. Sustain. Energy Rev.*, vol. 20, pp. 411–419, 2013.

[99] C. J. Hopfe and J. L. M. Hensen, “Uncertainty analysis in building performance simulation for design support,” *Energy Build.*, vol. 43, no. 10, pp. 2798–2805, 2011.

[100] S. De Wit and G. Augenbroe, “Analysis of uncertainty in building design evaluations and its implications,” *Energy Build.*, vol. 34, no. 9, pp. 951–958, 2002.

[101] Y. Sun, Y. Heo, M. Tan, H. Xie, C. F. Jeff Wu, and G. Augenbroe, “Uncertainty quantification of microclimate variables in building energy models,” *J. Build. Perform. Simul.*, vol. 7, no. 1, pp. 17–32, 2014.

[102] A. S. Silva and E. Ghisi, “Uncertainty analysis of the computer model in building performance simulation,” *Energy Build.*, vol. 76, pp. 258–269, 2014.

[103] T. Hong, S. C. Taylor-Lange, S. D’Oca, D. Yan, and S. P. Corgnati, “Advances in research and applications of energy-related occupant behavior in buildings,” *Energy Build.*, 2015.

[104] S. P. Corgnati, E. Fabrizio, and M. Filippi, “The impact of indoor thermal conditions, system controls and building types on the building energy demand,” *Energy Build.*, vol. 40, no. 4, pp. 627–636, 2008.

[105] J. Heller, M. Heater, and M. Frankel, “Sensitivity Analysis: Comparing the Impact of Design, Operation, and Tenant Behaviour on Building Energy Performance,” 2011.

[106] T. Zhang, P. O. Siebers, and U. Aickelin, “Modelling electricity consumption in office buildings: An agent based approach,” *Energy Build.*, vol. 43, no. 10, pp. 2882–2892, 2011.

[107] E. Azar and C. C. Menassa, “A comprehensive analysis of the impact of occupancy parameters in energy simulation of office buildings,” *Energy Build.*, vol. 55, pp. 841–853, 2012.

[108] P. Hoes, J. L. M. Hensen, M. G. L. C. Loomans, B. de Vries, and D. Bourgeois, “User behavior in whole building simulation,” *Energy Build.*, vol. 41, no. 3, pp. 295–302, 2009.

ACCEPTED MANUSCRIPT

[109] IEA, “IEA EBC Annex 66,” *Energy in Buildings and Communities Programme*. [Online]. Available: http://annex66.org/. [Accessed: 20-Mar-2016].

[110] US DOE, “Commercial Reference Buildings.” [Online]. Available: http://energy.gov/eere/buildings/commercial-reference-buildings. [Accessed: 23-Mar-2016].

[111] Y. Heo, R. Choudhary, and G. A. Augenbroe, “Calibration of building energy models for retrofit analysis under uncertainty,” *Energy Build.*, vol. 47, pp. 550–560, 2012.

[112] T. Nikolaou, I. Skias, D. Kolokotsa, and G. Stavrakakis, “Virtual Building Dataset for energy and indoor thermal comfort benchmarking of office buildings in Greece,” *Energy Build.*, vol. 41, no. 12, pp. 1409–1416, 2009.

[113] M. C. Katafygiotou and D. K. Serghides, “Analysis of structural elements and energy consumption of school building stock in Cyprus: Energy simulations and upgrade scenarios of a typical school,” *Energy Build.*, vol. 72, pp. 8–16, 2014.

[114] S. Salat, “Energy loads , CO2 emissions and building stocks : morphologies , typologies , energy systems and behaviour,” *Build. Res. Inf.*, vol. 37, no. 5–6, pp. 598–609, 2009.

[115] R. Choudhary and W. Tian, “Influence of district features on energy consumption in non-domestic buildings,” *Build. Res. Inf.*, vol. 42, no. 1, pp. 32–46, 2014.

[116] C. C. Davila, C. Reinhart, and J. Bemis, “Modeling Boston : A workflow for the generation of complete urban building energy demand models from existing urban geospatial datasets,” 2014.

[117] C. F. Reinhart and C. C. Davila, “Urban building energy modeling e A review of a nascent fi eld,” *Build. Environ.*, vol. 97, pp. 196–202, 2016.

[118] H. Hens, W. Parijs, and M. Deurinck, “Energy consumption for heating and rebound effects,” *Energy Build.*, vol. 42, no. 1, pp. 105–110, 2010.

[119] ASHRAE, “Building Energy Modelling Professional Certification.” [Online]. Available: https://www.ashrae.org/education--certification/certification/bemp-building-energy-modeling-professional-certification. [Accessed: 20-Mar-2016].

[120] IEA, “IEA EBC Annex 53,” *Energy in Buildings and Communities Programme*. [Online]. Available: http://www.iea-ebc.org/index.php?id=141. [Accessed: 20-Mar-2016].

[121] T. J. Williamson, “Predicting building performance: the ethics of computer simulation,” *Build. Res. Inf.*, vol. 38, no. 4, pp. 401–410, 2010.

[122] C. Filippín, “Benchmarking the energy efficiency and greenhouse gases emissions of school buildings in central Argentina,” *Build. Environ.*, vol. 35, no. 5, pp. 407–414, 2000.

[123] C. Martin, “Generating low-cost national energy benchmarks: A case study in commercial buildings in Cape Town, South Africa,” *Energy Build.*, vol. 64, pp. 26–31, 2013.

ACCEPTED MANUSCRIPT

[124] T. Sharp, “Energy Benchmarking In Commercial Office Buildings,” *ACEEE Summer Study Energy Effic. Build.*, vol. 4, pp. 321–329, 1995.

[125] J. Buck and D. Young, “The potential for energy efficiency gains in the Canadian commercial building sector: A stochastic frontier study,” *Energy*, vol. 32, no. 9, pp. 1769–1780, 2007.

[126] W. S. Lee and L. C. Lin, “Evaluating and ranking the energy performance of office building using technique for order preference by similarity to ideal solution,” *Appl. Therm. Eng.*, vol. 31, no. 16, pp. 3521–3525, 2011.

[127] E. Wang, “Benchmarking whole-building energy performance with multi-criteria technique for order preference by similarity to ideal solution using a selective objective-weighting approach,” *Appl. Energy*, vol. 146, pp. 92–103, 2015.

[128] W. S. Lee and C. K. Kung, “Using climate classification to evaluate building energy performance,” *Energy*, vol. 36, no. 3, pp. 1797–1801, 2011.

[129] C. Bloomfield and P. Bannister, “Energy and Water Performance Benchmarking in the Retail Sector: NABERS Shopping Centres,” pp. 1–13, 2010.

[130] W. Chung, Y. V. Hui, and Y. M. Lam, “Benchmarking the energy efficiency of commercial buildings,” *Appl. Energy*, vol. 83, no. 1, pp. 1–14, 2006.

[131] P. Bohdanowicz and I. Martinac, “Determinants and benchmarking of resource consumption in hotels—Case study of Hilton International and Scandic in Europe,” *Energy Build.*, vol. 39, no. 1, pp. 82–95, 2007.

[132] A. Aranda, G. Ferreira, M. D. Mainar-Toledo, S. Scarpellini, and E. Llera Sastresa, “Multiple regression models to predict the annual energy consumption in the Spanish banking sector,” *Energy Build.*, vol. 49, pp. 380–387, 2012.

[133] C. Ghiaus, “Experimental estimation of building energy performance by robust regression,” *Energy Build.*, vol. 38, no. 6, pp. 582–587, 2006.

[134] A. Sabapathy, S. K. V Ragavan, M. Vijendra, and A. G. Nataraja, “Energy efficiency benchmarks and the performance of LEED rated buildings for Information Technology facilities in Bangalore, India,” *Energy Build.*, vol. 42, no. 11, pp. 2206–2212, 2010.

[135] Q. Guo, Z. Tian, Y. Ding, and N. Zhu, “An improved office building cooling load prediction model based on multivariable linear regression,” *Energy Build.*, vol. 107, pp. 445–455, 2015.

[136] N. So and R. Richman, “A high level method to disaggregate electricity for cluster-metered buildings,” *Energy Build.*, vol. 111, pp. 351–368, 2016.

[137] D. Hsu, “How much information disclosure of building energy performance is necessary?,” *Energy Policy*, vol. 64, pp. 263–272, 2014.

ACCEPTED MANUSCRIPT

[138] T. Olofsson, S. Andersson, and J. U. Sj??gren, “Building energy parameter investigations based on multivariate analysis,” *Energy Build.*, vol. 41, no. 1, pp. 71–80, 2009.

[139] S. Önüt and S. Soner, “Energy efficiency assessment for the Antalya Region hotels in Turkey,” *Energy Build.*, vol. 38, no. 8, pp. 964–971, 2006.

[140] W.-S. Lee, “Benchmarking the energy efficiency of government buildings with data envelopment analysis,” *Energy Build.*, vol. 40, no. 5, pp. 891–895, 2008.

[141] W.-S. Lee and K.-P. Lee, “Benchmarking the performance of building energy management using data envelopment analysis,” *Appl. Therm. Eng.*, vol. 29, no. 16, pp. 3269–3273, 2009.

[142] W.-S. Lee, “Benchmarking the energy performance for cooling purposes in buildings using a novel index-total performance of energy for cooling purposes,” *Energy*, vol. 35, no. 1, pp. 50–54, 2010.

[143] J. L. Ruiz, J. V. Segura, and I. Sirvent, “Benchmarking and target setting with expert preferences: An application to the evaluation of educational performance of Spanish universities,” *Eur. J. Oper. Res.*, vol. 242, no. 2, pp. 594–605, 2015.

[144] J. C. Lam, K. K. W. Wan, S. L. Wong, and T. N. T. Lam, “Principal component analysis and long-term building energy simulation correlation,” *Energy Convers. Manag.*, vol. 51, no. 1, pp. 135–139, 2010.

[145] S.-M. Hong, G. Paterson, D. Mumovic, and P. Steadman, “Improved benchmarking comparability for energy consumption in schools,” *Build. Res. Inf.*, vol. 42, no. 1, pp. 47–61, 2014.

[146] B. Bordass, R. Cohen, R. Spevack, E. Burman, S.-M. Hong, P. Ruyssevelt, and J. Field, “Tailored energy benchmarks for offices and schools, and their wider potential,” *CIBSE ASHRAE Tech. Symp. Dublin, Ireland, 3-4 April 2014*, no. April, pp. 1–15, 2014.

[147] D. Hsu, “Identifying key variables and interactions in statistical models of building energy consumption using regularization,” *Energy*, vol. 83, pp. 144–155, 2015.

[148] S. a. Kalogirou, “Artificial neural networks in energy applications in buildings,” *Int. J. Low-Carbon …*, 2006.

[149] B. B. Ekici and U. T. Aksoy, “Prediction of building energy consumption by using artificial neural networks,” *Adv. Eng. Softw.*, vol. 40, no. 5, pp. 356–362, 2009.

[150] M. Yalcintas, “An energy benchmarking model based on artificial neural network method with a case example for tropical climates,” *Int. J. Energy Res.*, vol. 30, no. 14, pp. 1158–1174, 2006.

[151] Z. Hou, Z. Lian, Y. Yao, and X. Yuan, “Cooling-load prediction by the combination of rough set theory and an artificial neural-network based on data-fusion technique,” *Appl. Energy*, vol. 83, no. 9, pp. 1033–1046, 2006.

ACCEPTED MANUSCRIPT

[152] K. Li, C. Hu, G. Liu, and W. Xue, "Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis," *Energy Build.*, vol. 108, pp. 106–113, 2015.

[153] A. P. Melo, D. Cóstola, R. Lamberts, and J. L. M. Hensen, "Development of surrogate models using arti fi cial neural network for building shell energy labelling," *Energy Policy*, vol. 69, pp. 457–466, 2014.

[154] X. Gao and A. Malkawi, "A new methodology for building energy performance benchmarking: An approach based on intelligent clustering algorithm," *Energy Build.*, vol. 84, pp. 607–616, 2014.

[155] N. Gaitani, C. Lehmann, M. Santamouris, G. Mihalakakou, and P. Patargias, "Using principal component and cluster analysis in the heating evaluation of the school building sector," *Appl. Energy*, vol. 87, no. 6, pp. 2079–2086, 2010.

[156] D. Hsu, "Comparison of integrated clustering methods for accurate and stable prediction of building energy consumption data," *Appl. Energy*, vol. 160, pp. 153–163, 2015.

[157] M. Santamouris, G. Mihalakakou, P. Patargias, N. Gaitani, K. Sfakianaki, M. Papaglastra, C. Pavlou, P. Doukas, E. Primikiri, V. Geros, M. N. Assimakopoulos, R. Mitoula, and S. Zerefos, "Using intelligent clustering techniques to classify the energy performance of school buildings," *Energy Build.*, vol. 39, no. 1, pp. 45–51, 2007.

[158] J. C. Lam, K. K. W. Wan, and K. L. Cheung, "An analysis of climatic influences on chiller plant electricity consumption," *Appl. Energy*, vol. 86, no. 6, pp. 933–940, 2009.

[159] F. W. Yu and K. T. Chan, "Using cluster and multivariate analyses to appraise the operating performance of a chiller system serving an institutional building," *Energy Build.*, vol. 44, no. 1, pp. 104–113, 2012.

[160] S. Petcharat, S. Chungpaibulpatana, and P. Rakkwamsuk, "Assessment of potential energy saving using cluster analysis: A case study of lighting systems in buildings," *Energy Build.*, vol. 52, pp. 145–152, 2012.

[161] R. Arambula Lara, G. Pernigotto, F. Cappelletti, P. Romagnoni, and A. Gasparella, "Energy audit of schools by means of cluster analysis," *Energy Build.*, vol. 95, pp. 160–171, 2015.

[162] S. P. Pieri, I. Tzouvadakis, and M. Santamouris, "Identifying energy consumption patterns in the Attica hotel sector using cluster analysis techniques with the aim of reducing hotels' CO2 footprint," *Energy Build.*, vol. 94, pp. 252–262, 2015.

[163] A. A. Famuyibo, A. Duffy, and P. Strachan, "Developing archetypes for domestic dwellings - An Irish case study," *Energy Build.*, vol. 50, pp. 150–157, 2012.

[164] Z. Yu, B. C. M. Fung, F. Haghighat, H. Yoshino, and E. Morofsky, "A systematic procedure to study the influence of occupant behavior on building energy consumption," *Energy Build.*, vol. 43, no. 6, pp. 1409–1417, 2011.

ACCEPTED MANUSCRIPT

[165] B. Dong, C. Cao, and S. E. Lee, “Applying support vector machines to predict building energy consumption in tropical region,” *Energy Build.*, vol. 37, no. 5, pp. 545–553, 2005.

[166] Q. Li, Q. Meng, J. Cai, H. Yoshino, and A. Mochida, “Applying support vector machine to predict hourly cooling load in the building,” *Appl. Energy*, vol. 86, no. 10, pp. 2249–2256, 2009.

[167] H. Zhao and F. Magoulès, “Parallel Support Vector Machines Applied to the Prediction of Multiple Buildings Energy Consumption,” *J. Algorithm. Comput. Technol.*, vol. 4, no. 2, pp. 231–249, 2010.

[168] Z. Yu, F. Haghighat, B. C. M. Fung, and H. Yoshino, “A decision tree method for building energy demand modeling,” *Energy Build.*, vol. 42, no. 10, pp. 1637–1646, 2010.

[169] K. Li, H. Su, and J. Chu, “Forecasting building energy consumption using neural networks and hybrid neuro-fuzzy system: A comparative study,” *Energy Build.*, vol. 43, no. 10, pp. 2893–2899, 2011.

[170] Z. Li and G. Huang, “Re-evaluation of building cooling load prediction models for use in humid subtropical area,” *Energy Build.*, vol. 62, pp. 442–449, 2013.

[171] X. Pang, M. Wetter, P. Bhattacharya, and P. Haves, “A framework for simulation-based real-time whole building performance assessment,” *Build. Environ.*, vol. 54, pp. 100–108, 2012.

[172] K. Yun, R. Luck, P. J. Mago, and H. Cho, “Building hourly thermal load prediction using an indexed ARX model,” *Energy Build.*, vol. 54, pp. 225–233, 2012.

[173] S. Karatasou, M. Santamouris, and V. Geros, “Modeling and predicting building’s energy use with artificial neural networks: Methods and results,” *Energy Build.*, vol. 38, no. 8, pp. 949–958, 2006.

[174] A. Kusiak and G. Xu, “Modeling and optimization of HVAC systems using a dynamic neural network,” *Energy*, vol. 42, no. 1, pp. 241–250, 2012.

[175] P. De Wilde, C. Martinez-Ortiz, D. Pearson, I. Beynon, M. Beck, and N. Barlow, “Building simulation approaches for the training of automated data analysis tools in building energy management,” *Adv. Eng. Informatics*, vol. 27, no. 4, pp. 457–465, 2013.

[176] C. Yan, S. Wang, and F. Xiao, “A simplified energy performance assessment method for existing buildings based on energy bill disaggregation,” *Energy Build.*, vol. 55, pp. 563–574, 2012.

[177] S. R. Iyer, M. Sankar, P. Venkata Ramakrishna, V. Sarangan, A. Vasan, and A. Sivasubramaniam, “Energy disaggregation analysis of a supermarket chain using a facility-model,” *Energy Build.*, vol. 97, pp. 65–76, 2015.

[178] T. Rasanen, D. Voukantsis, H. Niska, K. Karatzas, and M. Kolehmainen, “Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data,” *Appl. Energy*, vol. 87, no. 11, pp. 3538–3545, 2010.

ACCEPTED MANUSCRIPT

[179] I. P. Panapakidis, T. a. Papadopoulos, G. C. Christoforidis, and G. K. Papagiannis, “Pattern recognition algorithms for electricity load curve analysis of buildings,” *Energy Build.*, vol. 73, pp. 137–145, 2014.

[180] M. S. Gul and S. Patidar, “Understanding the energy consumption and occupancy of a multi-purpose academic building,” *Energy Build.*, vol. 87, pp. 155–165, 2015.

[181] P. R. S. Jota, V. R. B. Silva, and F. G. Jota, “Building load management using cluster and statistical analyses,” *Int. J. Electr. Power Energy Syst.*, vol. 33, no. 8, pp. 1498–1505, 2011.

[182] CIBSE, *CIBSE Guide F: Energy efficiency in Buildings*. Chartered Institute of Building Services Engineers, 2012.

[183] A. Leaman, F. Stevenson, and B. Bordass, “Building evaluation: Practice and principles,” *Build. Res. Inf.*, vol. 38, no. 5, pp. 564–577, 2010.

[184] P. Steskens and M. Loomans, “PERFECTION: T1.3 - Performance Indicators for Health and Comfort,” 2010.

[185] O. Goçer, Y. Hua, and K. Goçer, “Completing the missing link in building design process: Enhancing post-occupancy evaluation method for effective feedback for building performance,” *Build. Environ.*, vol. 89, pp. 14–27, 2015.

[186] L. Yang, H. Yan, and J. C. Lam, “Thermal comfort and building energy consumption implications - A review,” *Appl. Energy*, vol. 115, pp. 164–173, 2014.

[187] A. C. Menezes, A. Cripps, D. Bouchlaghem, and R. Buswell, “Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap,” *Applied Energy*, vol. 97. pp. 355–364, 2012.

[188] US Department of Energy, “Benchmarking & Transparency Policy and Program Impact Evaluation Handbook,” 2015.

[189] US Green Building Council, “LEED - Leadership in Energy and Environmental Design.” [Online]. Available: http://www.usgbc.org/LEED/. [Accessed: 20-Mar-2016].

[190] US EPA, “Energy Star Portfolio Manager.” [Online]. Available: https://www.energystar.gov/buildings/facility-owners-and-managers/existing-buildings/use-portfolio-manager.

[191] US Energy Information Administration, “Commercial Building Energy Consumption Survey.” [Online]. Available: http://www.eia.gov/consumption/commercial/. [Accessed: 20-Mar-2016].

[192] California Energy Commission, “California Commercial End-Use Survey (CEUS),” *Analysis*, no. March, 2006.

[193] Natural Resources Canada, “Survey of Commercial and Institutional Energy Use: Buildings 2009,” 2012.

ACCEPTED MANUSCRIPT

[194] C. Nicholls, "Energy use in non-domestic buildings: the UK government's new evidence base," *Build. Res. Inf.*, vol. 42, no. 1, pp. 109–117, 2014.

[195] N. E. Matson and M. A. Piette, "Review of California and National Methods for Energy-Performance Benchmarking of Commercial Buildings," no. 57364, p. 74, 2005.

[196] ANSI/ASHRAE/IES, "Standard 90.1 - Energy Standard for Buildings Except Low-Rise Residential Buildings." ASHRAE, 2013.

[197] S. P. Corgnati and V. Corrado, "Energy Monitoring and Labelling," in *A Handbook of Sustainable Building Design and Engineering*, D. Mumovic and M. Santamouris, Eds. Earthscan, 2009.

[198] R. Cohen and B. Bordass, "Mandating transparency about building energy performance in use," *Build. Res. Inf.*, vol. 43, no. 4, pp. 534–552, 2015.

[199] P. Bannister, "NABERS: Lessons from 12 years of performance-based ratings in Australia," *Exergy*, no. Eu 2002, 2012.

[200] P. Rajagopalan and C. Y. Leung Tony, "Progress on building energy labelling techniques," *Adv. Build. Energy Res.*, vol. 6, no. 1, pp. 61–80, 2012.

[201] D. Leipziger, "Comparing Building Energy Performance Measurement. A framework for international energy efficiency assessment systems," p. 39, 2013.

[202] K. Mo, L. Burt, B. Hao, J. Cheng, A. Burr, and S. Kemkar, "Comparative Analysis of U . S . and China Building Energy Rating and Labeling Systems China ' s MOHURD Rating and Labeling Program," *ACEEE Summer Study Energy Effic. Build.*, pp. 256–269, 2010.

[203] L. Kerr, H. Beber, and D. Hope, "New York City Local Law 84 Benchmarking Report," New York, 2012.

[204] Institute for Market Transformation, "BuildingRating.org." [Online]. Available: http://www.buildingrating.org/. [Accessed: 20-Mar-2016].

[205] Building Research Establishment, "BREEAM - BRE Environmental Assessment Methodology." [Online]. Available: http://www.breeam.com/. [Accessed: 20-Mar-2016].

[206] Y. Schwartz and R. Raslan, "Variations in results of building energy simulation tools, and their impact on BREEAM and LEED ratings: A case study," *Energy Build.*, vol. 62, pp. 350–359, 2013.

[207] G. R. Newsham, S. Mancini, and B. J. Birt, "Do LEED-certified buildings save energy? Yes, but…," *Energy Build.*, vol. 41, no. 8, pp. 897–905, 2009.

[208] J. H. Scofield, "Do LEED-certified buildings save energy? Not really…," *Energy Build.*, vol. 41, no. 12, pp. 1386–1390, 2009.

[209] *Henry Gifford, et al. v. United States Green Building Council, et al.*, vol. 10 CIV 774. .

[210] International Code Council, "International Green Construction Code," 2012.

ACCEPTED MANUSCRIPT

[211] H. Bruhns, P. Jones, R. Cohen, B. Bordass, and Hywel Davies, “CIBSE Review of Energy Benchmarks for Display Energy Certificates - Analysis of DEC results to date,” no. May, p. 66, 2011.

[212] P. Armitage, D. Godoy-Shimizu, K. Steemers, and T. Chenvidyakarn, “Using Display Energy Certificates to quantify public sector office energy consumption,” *Build. Res. Inf.*, vol. 43, no. 6, pp. 691–709, 2014.

[213] J. H. Scofield, “Efficacy of LEED-certification in reducing energy consumption and greenhouse gas emission for large New York City office buildings,” *Energy Build.*, vol. 67, pp. 517–524, 2013.

[214] E. Mills, H. Friedman, T. Powell, N. Bourassa, Da. Claridge, T. Haasl, and M. A. Piette, “The Cost-Effectiveness of Commercial-Buildings Commissioning,” *Lawrence Berkeley Natl. Lab.*, 2004.

[215] P. de Wilde, W. Tian, and G. Augenbroe, “Longitudinal prediction of the operational energy use of buildings,” *Build. Environ.*, vol. 46, no. 8, pp. 1670–1680, 2011.

[216] B. Bordass, A. Leaman, and P. Ruyssevelt, “Assessing building performance in use 5: conclusions and implications,” *Build. Res. Inf.*, vol. 29, no. 2, pp. 37–41, 2001.

[217] L. Pérez-Lombard, J. Ortiz, and D. Velázquez, “Revisiting energy efficiency fundamentals,” *Energy Effic.*, vol. 6, no. 2, pp. 239–254, 2013.