Energy and Buildings 158 (2018) 1662–1682

Contents lists available at ScienceDirect

Energy and Buildings

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

# Suitability analysis of modeling and assessment approaches in energy efficiency in buildings

C. Koulamas[a], A.P. Kalogeras[a], R. Pacheco-Torres[b,*], J. Casillas[c], L. Ferrarini[d]

[a] Industrial Systems Institute, ATHENA Research and Innovation Center, Platani-Patras, 26504, Greece
[b] Departamento de Ingeniería Civil, Construcción, Infraestructura y Transporte, ETS de Ingeniería Civil, Universidad Politécnica de Madrid, Alfonso XII, 3, 28014, Madrid, Spain
[c] Department of Computer Science and Artificial Intelligence, University of Granada, CITIC-UGR, E-18071, Granada, Spain
[d] Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Via Ponzio 34/5, 20133, Milano, Italy

ARTICLE INFO

*Article history:*
Received 21 February 2017
Received in revised form 8 November 2017
Accepted 2 December 2017
Available online 5 December 2017

*Keywords:*
Building energy model
Building energy consumption
Energy performance assessment
Modeling comparison

ABSTRACT

The widely accepted importance of energy efficiency in the building sector is continuously acknowledged by the engineering and research community, as proven by the quantity and diversification of relevant modeling proposals in literature. It is often difficult to collect and assess this plethora of approaches and sometimes the diversity of the features of the available options makes it hard to decide what is the most convenient for the purpose required. This work presents a comprehensive analysis of the most important results today, along with their various classification and assessment approaches for modeling energy building consumption. A critical review of the limitations of the different existing approaches is conducted, and open research challenges are also highlighted. Finally, a horizontal and selective assessment of their suitability according to a descriptive set of qualitative comparison contexts and parameters is provided.



**Contents**

1. Introduction ....... 1662
2. Classification of modeling approaches ....... 1663
   2.1. Classification according to the relative hierarchical position of data inputs and building sector ....... 1663
   2.2. Classification according to the details of required information ....... 1664
      2.2.1. White box or physical models ....... 1664
      2.2.2. Black/grey box or statistical models ....... 1664
   2.3. Classification according to the energy data acquisition approach ....... 1665
3. Combined insight on classification and methodologies ....... 1665
   3.1. White box/Physical/Forward models ....... 1666
   3.2. Black box/Statistical/Inverse models ....... 1677
      3.2.1. Conventional statistics/Regression-based models ....... 1677
      3.2.2. Machine learning ....... 1677
   3.3. Grey box/Hybrid models ....... 1678
4. Discussion on the suitability of approaches ....... 1679
5. Concluding remarks ....... 1680
   Acknowledgments ....... 1680
   References ....... 1680

## 1. Introduction

The residential and commercial building sectors account for about 20% of the total energy consumption in the industrialized

* Corresponding author.
*E-mail addresses:* koulamas@isi.gr (C. Koulamas), kalogeras@isi.gr (A.P. Kalogeras), rosalia.pacheco@upm.es (R. Pacheco-Torres), casillas@decsai.ugr.es (J. Casillas), luca.ferrarini@polimi.it (L. Ferrarini).

https://doi.org/10.1016/j.enbuild.2017.12.002
0378-7788/

world [1]. The sector expansion drives its energy consumption increase. More specifically world delivered energy consumption grows by an average 1.4% per year in the residential building sector and 1.6% per year in the commercial building sector from 2012 to 2040 [1]. Nevertheless, there is growing interest in the reduction of building energy consumption and the associated greenhouse gas emissions. In Europe, the European Union has especially addressed the issue of building energy consumption and efficiency [2], in order to reduce its energy dependency, and greenhouse gas emissions.

With reference to residential buildings, most of the energy goes towards space conditioning. Top four energy end-uses in US residential buildings in 2005 included space heating, space cooling, water heating, and lighting with 30.7%, 12.3%, 12.2%, and 11% of total energy consumed in buildings respectively. Refrigeration, electronics, wet cleaning, cooking, and computers supplement the list of most important residential energy end-uses. On the other hand, when it comes to commercial buildings, space conditioning remains the primary target for energy end-uses. Top four energy end-uses in US commercial buildings in 2005 included lighting, space heating, space cooling, and water heating with 25.5%, 14.2%, 13.1%, and 6.8% of total energy consumed in buildings respectively. Electronics, ventilation, refrigeration, computers, and cooking supplement the list of the most important commercial energy end-uses [3].

Different parameters affect the degree to which energy end-uses affect overall energy consumption, including climate and meteorological conditions, occupancy and occupant behavior, building characteristics, building systems and appliances. Furthermore, when it comes to energy consumption end-uses may affect one another, as is the case of space heating attributed to appliances. Depending on the building site, the source of energy may be diverse, e.g. electricity, natural gas, or oil, and it may include secondary sources, such as generation (e.g. Renewable Energy Sources – RES), co-generation, and passive solar gains.

The modeling of energy consumption and efficiency in buildings is a useful tool that allows the quantification of building energy consumption and sharing of end-uses. In this context, it can provide a useful prediction of consumed energy that, accumulated to a regional or national scale, can determine energy supply requirements. Furthermore, it can provide useful feedback on decision support with reference to building retrofits, application of new technologies and materials, so that return of investment is calculated for different types of building interventions.

The focus of this paper is to review approaches for modeling energy consumption and efficiency in buildings, and propose an assessment methodology of existing approaches, based on a qualitative comparison. The rest of the paper is structured as follows. Section 2 provides a classification of modeling approaches. Section 3 presents selected state-of-the-art implementations and results. Section 4 presents a discussion on the relative suitability of modeling approaches. Finally, Section 5 summarizes the conclusions.

## 2. Classification of modeling approaches

Different modeling approaches appear in literature. They utilize input data to calculate or simulate energy consumption. Modeling approaches can vary significantly depending on the availability and details of the data. Different criteria have been defined for their classification, including the relative hierarchical position of data inputs as compared to the building sector, the details of the required information, and the energy data acquisition approach. A brief description of the main categories in each criterion and their strengths and weakness is given below. The limitations and open challenges of the existing approaches are also highlighted.

### *2.1. Classification according to the relative hierarchical position of data inputs and building sector*

Two general categories may be discerned: top-down and bottom-up. Bottom-up models calculate the energy consumption of an individual building or groups of buildings, and then extrapolate to a regional or national scale. Top-down models utilize total building sector energy consumption estimation, and map energy consumption to building sector global characteristics. Macroeconomic indicators, such as Gross Domestic Product (GDP), price indices, and employment rates are used to perform regression analysis and obtain the energy consumption. A subsequent microscale approach may provide individual consumption. Following this approach, a classification of modeling techniques is presented in [4]. According to this, the top-down category includes econometric and technological models. Economic indexes, such as energy price, are the main input of the former, while technological models attribute energy consumption to broad characteristics of building stock.

The main limitation of this approach is the primary need of massive data that in some cases is not available or supplied by building managers. Furthermore, sensitive information such as housing surveys may be needed, which is not always accessible. An existing gap in methodological resources to explain energy consumption of singular buildings or buildings under very specific energy use conditions is evident. Due to the fact that this approach does not distinguish energy consumption due to individual end-uses, it is not the most appropriate to identify massive energy consumers in buildings. In general, the approach output does not provide detailed information in order to design specific energy saving strategies in buildings oriented to reduce energy needs by end-use.

Bottom-up models estimate separately the energy consumption of a building, and then extrapolate to regional or national level. Two different methodologies may be used: statistical or engineering. Statistical methods exploit established relations between end-uses and energy consumption. Relevant models can be applied to predict the energy consumption of representative buildings. Historical information is used to establish relations between building energy consumption and end-uses. Regression, conditional demand analysis, and neural networks are classified under statistical methods. Regression and conditional demand analysis use regression analysis to determine model coefficients, while the latter takes into account the existence of end-use appliances. Neural networks rely on simplified mathematical models seeking to minimize errors.

Engineering methods estimate final energy consumption based on building characteristics and uses. Historical consumption data are used for the calibration of derived models ensuring compliance with the building Measuring and Verification guidelines [5]. Distribution, archetypes and samplings are classified under engineering methods. Distribution technique relies on the distribution and use of end-use appliances aggregating to end-use energy consumption, missing though end-use interactions. Archetypes classify buildings to representative building classes. Energy consumption is an estimate of modeled archetypes, allowing extrapolation to a larger scale. Sampling technique utilizes energy consumption data from a sample of buildings or energy consuming units. Providing a wide range of buildings and making the sample representative of the building stock can lead to wider energy consumption estimation.

One of the main limitations of bottom-up models is the need for detailed data on energy consumption, frequently acquired by advanced metering systems. A considerable amount of historical data is also necessary, in order to have enough base data to train the predictive model. They also present a clear limitation with reference to the need for detailed building constructive information. When project documents are not available or accessible, such modeling techniques have to rely to a large extent on user/engineer experience and previous knowledge. Engineering models are not suitable for ancient or historical buildings, which were built in the absence of technical guides, making it almost impossible to know such constructive details about the material composition and the real status of external walls.

The top-down modeling's main strength is the need for aggregated data only, which is generally widely available. Top-down approaches rely on historical data and allow the forecast of energy consumption on a larger scale, without going into detail on the specific end-uses. Thus, the approach is quite suitable for the purpose of decision making on energy policies at regional, national or international level. Nevertheless, the reliance on historical data is a drawback, since there is no possibility to model discontinuous technological advances. Furthermore, when it comes to large buildings, the historical data acquisition can be a complex process, while in the case of modeling of several buildings there is a need for historical data harmonization. Finally, the lack of end-use details makes it difficult to identify key consuming areas in the case of modeling for energy retrofitting purposes.

Meanwhile, the bottom-up methodology allows a closer approximation to consumer areas. Also, it is related to a range of parameters that affect final energy demand. However, it requires a great level of detailed data and may be subject to a number of difficulties in order to choose a sufficiently representative sample of the building portfolio.

### 2.2. Classification according to the details of required information

According to the details of their required information, modeling techniques can be classified as white box, black box, and grey box [6]. White box techniques, or otherwise called physical models, use sets of equations to solve building physical phenomena, such as heat transfer. A deep level of detail about building geometry and description of material properties is required, presenting one of the main limitations of these techniques. Yet, there is no need for model training data. White box methods are widely used to model the building thermal behavior and their results may be interpreted in physical terms. Another limitation of these techniques is the need for an expert to build the model and interpret results, a role not suitable for the common energy managers. Furthermore, resulting models have difficulties in extracting conclusions or being adapted to different buildings bearing different physical behaviors. Despite the high impact of building user behavior on final energy consumption, the use in these models is usually misleading.

On the other hand, black box approaches do not require such detailed physical information on buildings. Such models utilize samples of training data, describing the behavior of specific systems. Black box approaches can predict energy consumption, when given a large amount of training data over an exhaustive period of time. Trends may be found across different buildings, yet data mining techniques are building-specific, leading to needs for new modeling, when a new building is treated. Difficulties exist with reference to the interpretation of results in physical terms. The main limitation of these models is the difficulty to adapt the model to individual buildings, given that their internal calculation engine is not accessible to users or it does not provide a friendly user interface.

Grey box models combine physical and statistical approaches. A rough description of building geometry and parameters is supplemented by a smaller amount of training data over a short period of time. Grey box models use the mathematical structure of the physics-based white box models and measured data to estimate their parameters. Results can be interpreted in physical terms; yet, this hybrid approach that covers two distinct scientific domains may be more difficult to grasp. Grey box models represent a balance between the good generalization capability of white box models and high accuracy of black box models. Compared to the black-box models, grey box models require more effort during the definition stage, having good generalization capabilities, whereas obtaining higher accuracy compared to the white-box models.

#### 2.2.1. White box or physical models

Physical models are based on solving mathematical equations, derived from physical laws, such as the energy conservation law. Numerical software is usually used for this purpose. A wide range of mechanisms can be taken into account, such as conditioning systems, renewables, hydrothermal plants, and occupant behavior. There are three main calculations [6]: the Computational Fluid Dynamics method, the Zonal method, and the Nodal method.

The Computational Fluid Dynamics (CFD) method is the most complete approach, since it is three-dimensional. Each building zone is divided into cells; each cell is a control volume. Thus, quite complex building geometries can be studied. The method's main drawback is its large computation time and model complexity. The application fields of this method are HVAC systems, indoor air quality, and pollutant distribution.

The Zonal method represents a two-dimensional simplification of the CFD method. Each building zone is divided again into cells; each cell is the division of a room. It permits the determination of local state variables, such as temperature, concentration, pressure and air velocity in a large volume. Despite being simpler than the CFD method, it remains quite time-consuming, while requiring detailed descriptions of factors affecting indoor flow profiles. Application fields of this method include indoor thermal comfort, artificial and natural ventilation.

The nodal method represents the simplest of the physical methods. Each building zone is regarded as a homogeneous cell, a node, with uniform distribution of physical quantities (e.g. ambient temperature) modeled as state variables. Equations are solved per node, offering a one-dimensional approach. The implementation is easier and the calculation times are reasonable. Yet, it is difficult to study large volume systems and it is impossible to address local effects like heat or source of pollutant. The application fields of this method are the determination and time evolution of total energy consumption, average room temperature and cooling and heating loads.

#### 2.2.2. Black/grey box or statistical models

Statistical methods do not require any physical information about the building, yet, they rely on training data to extract system functions. Multiple regression, Artificial Neural Networks (ANN) and decision trees represent three statistical techniques used for predicting electrical energy consumption [7]. Regression models are commonly used due to the interpretability of model parameters and ease of use; yet they can only ascertain the relationship among the selected variables about the underlying causal mechanism, but there might be uncertainties, when a relevant variable is missing or badly measured. Neural networks are useful for energy prediction, when mathematical formulas and prior knowledge on the relationship among inputs and outputs are not known, yet they have difficulty in testing parameter significance. Despite solving this problem, decision tree models are complex, as they are susceptible to noise.

When comparing the ANN-based model and the physical simulation model (based on the EnergyPlus® software for example), as predicting tools for energy consumption forecasting of a non-residential building, models based on physical principles typically offer the possibility to evaluate new strategies for reducing energy consumption, while ANN models appeared more limited in this sense [8].

The performance of grey-box models and black-box models focusing on residential heating, ventilation and air conditioning systems (HVAC) is compared in [9]. Grey box models consist of a combination of energy balance equations and parameter estimation based on sensor measurements of subsystem inputs and outputs. Black box models that were compared were based on Multiple-Input and Multiple-Output (MIMO) ANN, transfer function process, state-space and autoregressive exogenous model. ANN models performed best among compared models.

Two aspects to consider with reference to statistical methods are data dimensionality and obtained model interpretability versus accuracy balance [6]. An important amount of data is required by statistical techniques. The preferable measurement resolution is the hour or days; the resolution of months is hardly useful. With reference to the number of variables, there is a tendency to use as many variables as possible, without considering the redundancy or correlation, since current machine learning techniques can deal with large numbers of variables and variable selection, so that processing can be applied. When it comes to the aspect of interpretability versus accuracy, techniques like Support Vector Machines (SVM) or ANN produce models that are not understandable by humans, thus being useful for behavior simulation, but not for reasoning explanation. On the other hand, decision trees or rule sets (such as greedy or genetic algorithms) are easily understood and can help better analyze variables and relationship causalities. In between, regression or graphical models can be interpreted in a general way. On the one hand, applying different statistical techniques to the same problem and data can generate more accurate but more illegible models for prediction, while on the other hand, more easily interpretable but less accurate models for description.

### 2.3. Classification according to the energy data acquisition approach

The energy performance assessment method is based on a relevant energy quantification process, which in turn depends on an energy data acquisition approach. Energy quantification methods may be classified into three categories: calculation-based, measurement-based, and hybrid methods [10].

Many of the physical and statistical methods are categorized under the broader Calculation-based class of energy quantification methods [6,10]. Measurement-based methods focus mainly on the Building Management System (BMS)/Sub-metering utilization and on energy disaggregation. Energy disaggregation can be achieved either by means of pattern recognition setups trained by available sub-metering information, or through approximations summing up to the total energy consumption known from the energy bills.

Calculation-based methods are diverse with reference to their consideration of building and system dynamic effects. They are classified into dynamic and steady-state methods. Dynamic methods capture building and system dynamics resulting in calculation complexity often implemented through detailed simulation. Dynamic simulations usually use a forward modeling approach, although dynamic inverse modeling is also reported (classified under hybrid methods). Typical input parameters include building, system and component descriptions along with weather conditions. The details of the mathematical simulation algorithms are described in the simulation engine and involve thermal load calculations, based either on heat balance or weighting factor methods, various air-handling and control systems simulation according to their schedules and calculation of final electricity and fuel use based on system component characteristics. Different simulation tools include e-Quest® (DOE-2), EnergyPlus® (DOE), ESP-r, and TRNSYS®.

On the other hand, steady-state methods ignore or simplify dynamic effects thus decreasing complexity and achieving high computation speeds. They may adopt forward or inverse modeling approaches. The Simplified Building Energy Model (SBEM), adopted from the current Energy Performance Building Directive (EPBD) related standards such as the EN ISO-13790, is a typical steady-state forward model, which follows a quasi-steady state method for the monthly heating and cooling demands, taking into account dynamic effects through correlation factors called utilization factors. Modeling examples using an inverse modeling approach relate energy performance indicators to energy influential factors and can be applied either to a whole building level or to a HVAC system level. Thus, the whole building energy consumption can be regressed in various ways against influencing parameters. Examples of steady-state inverse models are included in the ASHRAE Inverse Modeling Toolkit [11]. Such models include constant or mean models, two-parameter, three-parameter, four-parameter, five-parameter and multivariate models. Typical examples of other methods for building load calculation are the degree-day method, variable base degree-day method, BIN and modified BIN methods and the equivalent full load hour method.

Measurement-based quantification is based on measured data that represent actual building energy consumption, ranging from energy bills to more detailed energy sub-metering and monitoring. Energy bills represent a source of high quality energy measurement data that need to be disaggregated into end-uses, in order to develop a better understanding on energy use. Different disaggregation methods have been proposed such as the bottom-up calculation method, bottom-up short-term measurement method, and top-down disaggregation algorithm. Different methods have been proposed to increase disaggregation accuracy and detail, investigating sources of inaccuracy and introducing metrics for performance quantification [12]. The monitoring-based methods provide such accuracy and detail through metering and monitoring systems and platforms. Examples of such approaches include end-use sub-metering, installing separate energy meters on relevant circuit branches, the Non-Intrusive Load Monitoring (NILM) method, which is a pattern recognition-based method capable of firstly determining end-use operating characteristics and secondly distributing monitored energy into end-uses, and Building Management System (BMS) based methods.

Hybrid quantification methods are actually a combination of calculation-based methods and measurement-based methods, where measurements are used to reduce calculation discrepancies or identify model parameters. Usually, calibration procedures are based on hybrid methods using a building simulation program to tune input values, so that the program energy predictions follow closely energy data measurements and the Dynamic inverse modeling, being capable of capturing building dynamic effects, yet introducing a higher degree of complexity and needing measurements for model tuning. Typical examples of Dynamic inverse modeling include AutoRegressive Moving Average (ARMA) models, Fourier series models, ANN models and grey models.

## 3. Combined insight on classification and methodologies

With the exception of a first level classification of top-down vs. bottom-up [4], all other surveys focus mainly on the bottom-up sub-tree. All recent works tend to agree on a second-level classification, although with partial overlaps, and the use of slightly different terminology for the same underlying principles: physical/statistical/hybrid or white/black/grey, close to calculation/measurement/hybrid, close to engineering/statistical/hybrid approaches; all aligned with the classical (forward) and data-driven (inverse) relevant classification [13], with their dynamic or steady-state variances. The large picture relationships among current classification approaches, having combined the dimensions of pure modeling methods and quantification methods appear in Fig. 1. This presentation indicates on the one hand the clear sub-classes of distinct methodologies at the two ends of the spectrum

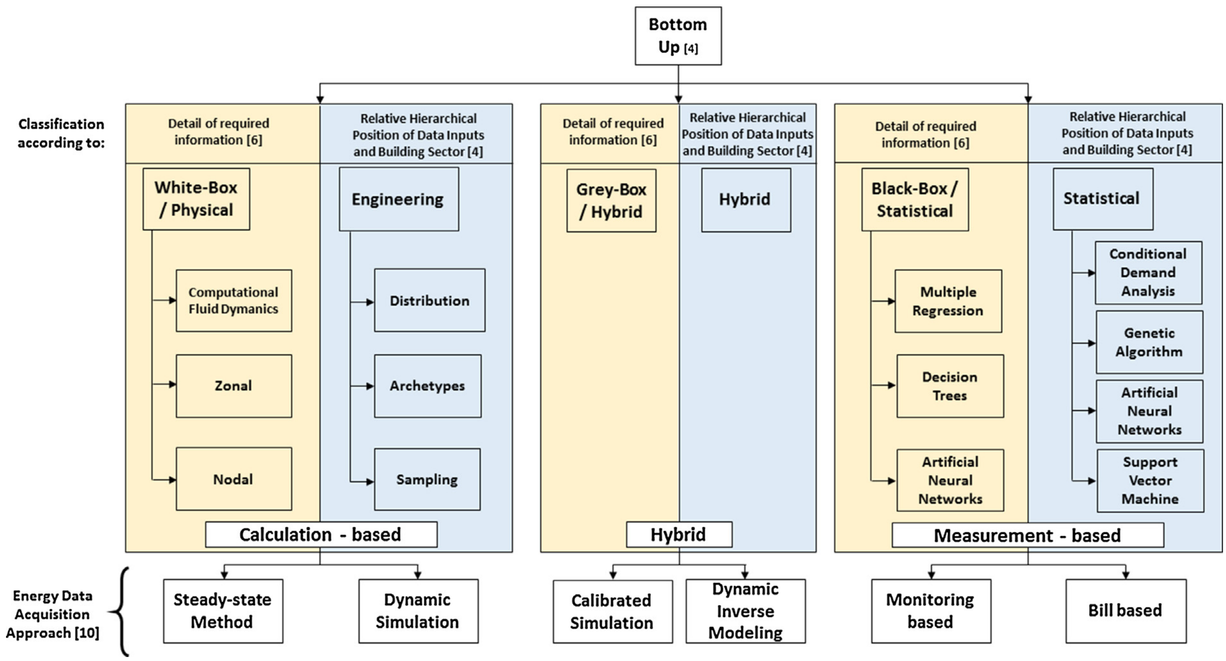


**Fig. 1.** Building modeling approaches classification.

and the lack of a similar analysis as we approach the middle point (hybrid methods). On the other hand, it clarifies the fact that the quantification methods cannot be assumed as orthogonal to or disconnected from the modeling approaches. For instance, there is evident dependency between a white-box model (i.e. physics) and its usage in a calculation method to quantify the energy consumption of an element. Another example is related to dependencies between a monitoring-based data collection method (e.g. a BMS) and the exploitation of measurement data in order to develop a statistical model (e.g. train an ANN or extract a regression function).

The following sections shed light into representative approaches from the literature in order to: a) make clear what are the current state-of-the art and quality limits of the forward and inverse modeling methodologies, and b) point out the benefits of hybrid methods that combine elements from the far ends of the modeling and quantification spectrum limits, as well as demonstrate their heterogeneity and multi-disciplinarity, explaining why current classifications do not typically provide generic subclasses of them. Table 1 summarizes the revised studies grouped by modeling approach.

### 3.1. White box/Physical/Forward models

A comprehensive introduction to important physical properties, processes and respective improvements related to important building envelope components appears in [14], including energy-efficient walls, fenestration technologies, advances in energy-efficient roofs and effects of thermal mass and phase change material on building air tightness, infiltration and cooling/heating loads and peak loads. The effects of microclimate around a building are discussed in [15], presenting a building energy performance quantitative analysis method, linking a microclimate model to the EnergyPlus® simulation program, in order to study effects of solar and long wave radiation, temperature, humidity, and wind speed. One limitation of [15] is the assumption that the surface temperatures of the ground and the obstructions are the same as the outdoor air temperature, assuming that the obstruction materials have no impact on the microclimate model. In this context, the positive effect of passive energy-saving techniques, such as façades shading by trees, cannot be simulated entirely, minimising the effect of urban contexts on the simulation process.

A methodology based on dynamic simulations analysing the parameters that mostly affect the cooling energy performance of the building space is discussed in [16], showing the secondary role of thermal and solar parameters of the opaque surface in contrast to the glazed surface, as well as the weak influence of the office building envelope compared to the more significant influence of internal heat sources in contrast to residential buildings. The effects of thermal insulation and in particular the usage of Phase Change Materials (PCM) are studied in [17] through an experimental validation of a semi-empirical model for the simulation of the phase change process.

The unbalanced study of summer versus winter performance indicators is the driving force behind the work [18], presenting results of comparison between actual and normalized energy performance of a cooling plant that equipped a Milan residential building. Real energy requirements, estimated via monitoring, were lower than those calculated with the Lombardy standard energy certification procedure, yet consistent with EnergyPlus® building simulation calculations. The real performance calculated for the winter case is consistent with the certification procedure calculation [19]. User behavior may lead to output differences.

Similar results appear in [20], where the simple hourly dynamic calculation method of EN ISO 13790 standard is applied using Matlab/Simulink® for an indicative building in 5 climate zones in Poland. The normative monthly method calculations show significant differences to EnergyPlus simulated values. The Dynamic method and steady-state monthly method of Italian standards are compared in [21], showing dynamic method adequacy to deal with structure inertial properties, with the model being validated by a measurement campaign. A methodology for heating and cooling energy consumption estimations, simplifying dynamic simulation methods, is presented in [22], implemented in Excel and validated against actual hospital measurements, as well as against the EnergyPlus simulation and the EN ISO13790 implementation.

Besides specialized civil, mechanical and electrical engineering sources, physical/forward models, at a building, system or plant level can be also found in literature of building control systems and

**Table 1**
Summary of the studies reviewed.

| Publication/Reference | Building description | | | | Aim of the work | Key contribution | Algorithm/Method | Data variable | Software used |
|---|---|---|---|---|---|---|---|---|---|
| | Type of building | Size [m$^2$] | Number of floors | Location | | | | | |
| Category: White box or physical models | | | | | | | | | |
| Yoshida, Ito, and Yokoyama [14] | Hospital | 25,000 | 10 | Japan | To investigate what energy supply system structure is suitable for a hospital for the purpose of saving cost. | Testing 25 different structures of alternative energy supply systems | Optimization approach – mixed-integer linear programming | Utility rates, the utility maximum contract demands, the numbers and capacities of equipment, the energy flow rates | GAMS/CPLEX solver |
| Yang et al. [15] | | 1000 | 5 | cooling: Guangzhou/heating: Frankfurt | Proposing an integrated simulation system for building energy assessment in urban environments | Coupling ENVI-met with EnergyPlus | Open loop analysis (one-way coupling simulation and no data feedback to the ENVI-met model) | ENVI-met: climatic variables, E+: convective heat transfer coefficient for each linking unit of the building, weather condition | The urban microclimate model ENVI-met/the building energy software Energy-Plus/coupling platform Building Controls Virtual Test Bed (BCTVB). |
| Ballarini and Corrado [16] | Residential/office | 192/928 | 2/intermediate | Rome (Italy) | Examines the relationship between the effects of the thermal insulation on the building energy need for cooling and all the aspects that have the most effect on the energy performance of buildings. | Analysing the different contributions to the internal air convective heat balance and their interrelations with different driving forces | Dynamic driving forces | Convective and radiative heat transfer, thermo-physical parameters, | Energy Plus |
| Gowreesunker, Tassou, and Kolokotroni [17] | Case study: environmental chamber, with 4 T-type thermocouples | 100 mm × 70 mm × 80 mm | It's a box of a new material | UK | Semi-empirical model for the simulation of the phase change process in phase change materials (PCM). | The contribution of this approach is to differentiate between melting and freezing, so that the solver uses the appropriate heat source function | CFD | Conductive heat transfer variables | Computational fluid dynamics (CFD) simulation environments/FLUENT |
| Dall'O', Sarto, Sanna, et al. [18] | Residential buildings with a total of 196 flats | Ground floor:82.3/4th floor: 94.36/top floor: 126.12 | 4 | Milan, Lombardy region-Italy | Comparison between the predicted and actual energy performance for summer cooling in high-energy performance residential buildings | The paper proposes an extensive comparative evaluation between the actual and normalized energy performance of a high-performance residential building equipped with a cooling plant. | Thermal dynamic modeling | Energy consumption/environmental conditions such as: indoor and outdoor air temperature and humidity | Software HOBOware of Onset/EnergyPlus |

Table 1 (*Continued*)

| Publication/Reference | Building description | | | | Aim of the work | Key contribution | Algorithm/Method | Data variable | Software used |
|---|---|---|---|---|---|---|---|---|---|
| | Type of building | Size [$m^2$] | Number of floors | Location | | | | | |
| Dall'O', Sarto, Galante, et al. [19] | 3 flats located in: Ground floor, 4th floor, top floor) | Ground floor:82.3/4th floor: 94.36/top floor: 126.12 | 4 | Lombardy region-Italy | Comparison between predicted and actual energy performance for winter heating in high-performance residential buildings | A detailed analysis of 3 flats with different ways of energy computation: theoretical (normative) estimation and real measurements. | Thermal dynamic modeling | Energy consumption/environmental conditions such as: indoor and outdoor air temperature and humidity | Software HOBOware of Onset/EnergyPlus |
| Michalak [20] | Typical building (house) | 130.8 | 2 | Poland | To present in detail the application of the simple hourly dynamic calculation method from EN ISO 13790 standard to calculate the annual demand of heating and cooling energy. | Ability to test different control strategies, to determine the optimal power value of heating/cooling sources, etc | Dynamic modeling: state space model | ISO 13790 standard variables | Matlab/Energy plus/Audytor OZC (based on EN ISO 13790 standard) |
| De Lieto Vollaro et al. [21] | Old building | 210 | 5 | Central Italy, climatic area D | Comparative analysis of the energy performances of an old building using a semi-stationary software and a dynamic one | Correct estimation of the energy demands, taking into account the dynamic properties of the structures | Semi-stationary (based on standard UNI EN ISO 13790) and dynamic approach (transfer functions method) | ISO 13790 standard variables | MC11300: steady-statepart analysis/TRNSYS: dynamic analysis |
| Čongradac et al. [22] | Emergency Center (Hospital) | 8350 | 5 | Serbia –Novi Sad | Presenting the tool for assessing the heating and cooling energy consumption | Ease of use, simplified set of input data, as well as the omission of a complex dynamic modeling | Static modeling approach | Thermo-physical variables similar as those in ISO 13790 standard | EnergyPlus, Riuska and Standard EN 13790/macros and Visual Basic functions of EXCEL |
| Mantovani and Ferrarini [23] | Commercial building (shoping center) | 26,369 | 5 | Campo de Fiori shopping center-northern, Italy | The design of an MPC architecture for the optimal temperature control of a real commercial building | A non-linear MPC approach for thermal energy control | Dynamic modeling and predictive control approach | Air and water temperatures, mass flows, conductive and convective heat, efficiency of heat pumps, fan coil models | Matlab/Energy plus/Audytor OZC (based on EN ISO 13790 standard) |
| Ferrarini and Mantovani [24] | Large commercial building | 26,369 | 5 | Gavirate, Italy | Modeling, control and energy management of a large-commercial building | Vertical air temperature stratification, aimed at efficient energy control | Dynamic modeling/classical and advanced control approach (PID & MPC) | Air and water temperatures, mass flows, conductive and convective heat, efficiency of heat pumps, fan coil models. | MATLAB/UNI EN ISO 13790 standard and dynamic modeling with Matlab/Simulink |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Category: Conventional Statistics/Regression based | | | | | | | | | |
| Krese, Prek, and Butala [25] | Office | 7200 | 13 | Ljubljana, Slovenia | To improve the cooling degree method taking into account both the sensible and latent loads and use it to analyze electric energy consumption data from an existing building and compared against the conventional cooling degree day approach | An improved cooling degree method which takes into account both the sensible and latent loads, is used to analyze electric energy consumption data from an existing building and compared against the conventional cooling degree day approach | Statistical analysis to improve cooling degree days method and piecewise linear regression | Monthly electric energy consumption | N/A |
| Fumo, Mago, and Luck [26] | Office | 715 | 1 | Atlanta and Meridian, USA | Employ a series of predetermined coefficients to the monthly energy consumption data from electrical. | The use of predetermined coefficients relieve the user from the burden of performing or learning how to perform a detailed dynamic simulation of the building. The information given by these coefficients could cover information missing from utility bills to perform an energy analysis | Uses an 'EnergyPlus normalized energy consumption coefficients' (E + NECC) as normalized energy profiles | Hourly electrical and fuel energy consumption | EnergyPlus to generate data |
| Smith et al. [27] | Office | 4980 | 3 | Baltimore, USA | Uses EnergyPlus normalized energy consumption coefficients to estimate the energy profiles of buildings with similar characteristics to a given benchmark model | the coefficient methodology decreases the error limits in almost all of the test points | Uses an 'EnergyPlus normalized energy consumption coefficients' (E + NECC) as normalized energy profiles | Hourly energy consumption | EnergyPlus to generate data |
| Catalina, Virgone, and Blanco [28] | Residential | Different buildings shapes with areas from 150 to 300 m2 | 1 | Paris, Marseille, Chambery, Strasbourg, Rouen, Brussels, France | The development of regression models to predict the monthly heating demand | Easy use equations to be applied by architects and professionals at early desing stage, with small range of error | Multiple regression analysis | Monthly heating demand | TRNSYS to generate data |

Table 1 (*Continued*)

| Publication/Reference | Building description | | | | Aim of the work | Key contribution | Algorithm/Method | Data variable | Software used |
|---|---|---|---|---|---|---|---|---|---|
| | Type of building | Size [m$^2$] | Number of floors | Location | | | | | |
| Catalina, Iordache, and Caracaleanu [29] | Residential | 3907 | 11 | Bucharest, Romania | Develop a simple and large applicable model to estimate heating energy demand based on three inputs: heat loss coefficient, the south equivalent surface and the difference between the indoor set point temperature and the sol-air temperature | A correction made in the model to better take into account human behavior improves heating consumption predictions under real building's operation conditions | Iteratively reweighted least squares (IRLS) | Heating energy demand | TRNSYS to generate data |
| Asadi, Amiri, and Mottahedi [30] | Commercial | 2,322.6 | 2 | Houston, USA | To build a simple but precise model to predict energy consumption, based on regression analysis with massive data results as inputs to cover a comprehensive set of variables | Monte Carlo was used to define a set of 70,000 simulation scenarios. High precision of predictions was obtained, within an error or 5% | Linear regression model and standardized regression coefficients | Annual heating and cooling demand | DOE-2 to generate data |
| Amiri, Mottahedi, and Asadi [31] | Commercial | 2322.6 | 2 | San Jose, USA | To create a multiple regression model, flexible and simple, to evaluate the building energy consumption and performance | Monte Carlo was uased to define a set of 30,000 simulation scenarios | Stepwise regression. multiple linear regression | Annual energy consumption | DOE-2 to generate data |
| Pedersen, Stang, and Ulseth [32] | Residential, office, educational | Case studies range from 70 to 7000 m$^2$ | N/A | Trondheim, Norway | Perform a load prediction method for heat and electricity demand in buildings and a method for load aggregation based on the building categories' load profiles | Model flexible and simple with high accuracy to evaluate the building energy consumption and performance | Piece-wise linear regression and probability distribution analyses | Annual heating and electricity demand | DOE-2 to generate data |
| Fumo and Rafe Biswas [33] | Residential | N/A | N/A | Texas, USA | Analyze previous information on regression analyses on prediction of energy consumption in buildings | Results from a case study, as the time interval of the observed data increases, the quality of the models improves | Simple linear, simple quadratic and multiple linear | Hourly and daily energy consumption | N/A |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Yun et al. [34] | Residential, office | 464.5 | 1 | Atlanta, USA | Develop an efficient ARX indexed model more accurate, easily implementable and computationally efficient AI-based models for cooling and heating loads in buildings | the performance of a properly indexed ARX model is better than that of non-indexed models and comparable to that of the BPNN | ARX model (Fourth order auto regressive model with exogenous inputs) | Cooling and heating loads | EnergyPlus to generate data |
| E. Wang, Shen, and Grosskopf [35] | Residential | Wide range given 480 case study | Wide range given 480 case study | Iowa, USA | Selective residual-clustering benchmarking method is proposed for building envelope energy efficiency evaluation with multi or high dimensional data set. | PCA allows to represent multi correlated variables with less principal uncorrelated components in terms of data variability. Results obtained are comparable with reliable infrared thermography | Multivariate linear regression analysis with principal component analysis to address the multicollinearity risk, PCR, PCA, MRA, Fuzzy C-Means | Energy efficiency of existing building envelopes | N/A |
| Qiang et al. [36] | Office | 12,770/49,800 | N/A | Tianjin, China | An improved mutlivariable linear regression model is presented, based on a better selection of meteorological variables and better description of internal factors | A dynamic two-step correction is proposed. PCA is an applicable measure to avoid the negative effect of multicollinearity on prediction | Multivariable linear regression | Daily cooling load | N/A |
| Majcen, Itard, and Visscher [37] | Residential | 105 | N/A | The Netherlands | Analyze key factors that cause discrepancies between theoretical and actual gas consumptions by using regression analysis. | Average indoor temperature and ventilation rate were found to have a large influence on the theoretical gas consumption whereas number of occupants and internal heat load have limited impact | Descriptive statistics and regression analysis | Gas consumption | N/A |

Table 1 (*Continued*)

| Publication/Reference | Building description | | | | Aim of the work | Key contribution | Algorithm/Method | Data variable | Software used |
|---|---|---|---|---|---|---|---|---|---|
| | Type of building | Size [$m^2$] | Number of floors | Location | | | | | |
| Majcen, Itard, and Visscher [38] | Residential | N/A | N/A | The Netherlands | Examine discrepancies between the normalized theoretical and actual heating consumption, in order to improve energy label certification calculation method. | Occupant behavior has larger effect that the considered by calculation method. Factors with significant effect are different for overpredicting and underpredicting cases | Multiple regression | Gas consumption | N/A |
| Hoşgör and Fischbeck [39] | Residential | N/A | N/A | Gainesville, Florida, Texas | To explore the effect of statistical modeling structural and demographic characteristics on residential energy efficiency parameters using Princeton Scorekeeping Method and publicly available data on house energy efficiency | Publicly available information can help predict energy efficiency parameters and savings potential. Predictive regression models can be applied anywhere and models with R2 values higher than 30% can be interpreted to have a relatively high explanatory power | PRISM (Princeton Scorekeeping Method) | Electricity, heating and cooling demand | N/A |
| US Environmental Protection Agency [40–42] | Medical | Building portfolio | Building portfolio | USA; Canada | To use regression models to identify the aspects of building activity that are significant drivers of energy use, normalize those factors and propose a method to score energy efficiency in Hospitals. | The methodology allows to compare energy use prediction for a building to its acutal energy use and give a score of performance, relative to the national population | Weighted ordinary least squares regression | Energy consumption expressed in source energy use intensity | N/A |
| Christiansen et al. [43] | Medical | 90 individual buildings, 400,000 $m^2$ | N/A | Hamburg, Germany | To give a better understanding of the time-dependent energy consumption of a medical building laboratory plug loads under consideration of uncertainties | Only a few plug load groups contribute the greater part of the total electrical energy demand | Cumulative load predictions | Electricity consumption | N/A |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Zhou et al. [45] | Office | 15 buildings, from 24,000 to 99,000 $m^2$ | N/A | Beijing and Hong Kong, China | To analyze the main characteristics of lighting energy use over various timescales capturing energy use patterns | The results are applicable to large office buildings without daylighting controls or any other automatic lighting controls. Lighting energy use was found to be mainly driven by the occupant schedule and the influence of the outdoor illuminance was very weak | Least squares regression | Hourly lighting's electricity consumption | N/A |
| Palacios-Garcia et al. [46] | Residential | N/A | N/A | Andalusia region, Spain | Perform a high-resolution stochastic model for simulating lighting consumption profile with high temporal resolution and analyze the economic and environmental impact of applying LED technology's penetration into domestic lighting systems | Results demonstrated a strong relationship between sunlight availability and active occupancy of dwellings with electricity consumption for lighting | Stochastic model | Hourly lighting's electricity consumption | N/A |
| C. Yan, Wang, and Xiao [47] | Commercial | 321,000/54,490 | 108 floors the first case study, not provided in the second | Hong Kong and Beijing, China | Develop a simply-use energy performance assessment method of cooling load in existing buildings, based on the electricity consumption data and cooling energy balances between demand side and supply side of HVAC systems. | The proposed simplified method provides satisfactory results on the annual analysis, given a higher error rate for monthly analysis. | Optimization algorithm using trial-and-error method | Disaggregated energy consumptions and the energy performance indicators of HVAC systems | N/A |

Table 1 (*Continued*)

| Publication/Reference | Building description | | | | Aim of the work | Key contribution | Algorithm/Method | Data variable | Software used |
|---|---|---|---|---|---|---|---|---|---|
| | Type of building | Size [m$^2$] | Number of floors | Location | | | | | |
| Category: Machine Learning Models | | | | | | | | | |
| Bagnasco et al. [48] | Medical clinic, three building set | 9500 | 3 blocks | Turin, Italy | Predict EEC from previous values | Original data and ANN algorithm | Time series: ANN (MLP) with RPROP | Daily – season separation –, previous day consumption | Matlab |
| Jovanović, Sretenović, and Živković [49] | University campus, 35 buildings | 300,000 | | Trondhei, Norway | Predict EEC from previous values | Original data and ensemble ANN | Time series: Ensemble of 3 ANNs (FFNN + RBFN + ANFIS) | Daily – cold period (January-March), only working days – previous day info | Matlab |
| Papantoniou and Kolokotsa [50] | Several cities | N/A | | Ancona (Italy), Chania (Greece), Granada (Spain), Mollet (Spain) | Predict air temperature for a 4h–24 h horizon | Predicting outdoor air temperature with ANN | Time series: ANN | Every 12 h or 24 h, one year – | Matlab |
| Chae et al. [51] | Three office buildings in urban area | 15,224 | | Korea | Predict EEC every 15-min | Predicting electricity consumption for next day with 15-min data resolution | Time series: ANN with Bayesian regularization (comparison with SVM, LR, RBF, lazy. . .) | Every 15-min, few weeks of 2012 – short-term monitoring – HVAC set temp, OT, RH, sky, WS, HVAC variables | N/A |
| Popoola, Munda, and Mpanda [52] | Several cities | N/A | | South Africa | Estimate and understand lighting load | Lighting load profile prediction with neuro-fuzzy systems | Regression: ANFIS | Survey data for 417 buildings – occupancy, income | Matlab |
| Platon, Dehkordi, and Martel [53] | Institutional building | 16,790 | | Calgary, Canada | Predict EEC for a 1h–6 h horizon | Prediction of electricity consumptuion with 1 h to 6 h horizon | Time series: ANN, PCA, CBR | Hourly, one year March 2013-May 2014 – OT, RH, IT, 8 vars. of HVAC | N/A |
| Koo and Hong [54] | School | 2000 | | test in Seoul, South Korea | Historical trends CO2 emission (energy performance) of a building | Studying historical trends in the energy performance of existing buildings | Regression: CBR and GA as optimizer (comparison with MRA and ANN) | 1999–2010, yearly – different building and using factors (people, classes, etc.) | Evolver for GA |
| Edwards, New, and Parker [55] | 3 test residential houses | N/A | | Knox County, Tennessee, USA | Predict EEC for the next hour | Predicting hourly electricity consumption | Time series: LR, ANN, SVR | Year 2010, every 15 min – artificial occupancy and usage of test houses – many variables on house controlled condition | Matlab, LIBSVM, LS-SVMlab |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Yu et al. [56] | 80 residential building | N/A | | 6 districts in Japan | Energy demand modeling | Energy demand modeling | Classification: Decision trees | Year 2003, every 15 min – 10 years weather, indoor, occupants | N/A |
| Li, Su, and Chu [57] | 2 buildings: a big building and a 10-floors library | 25,542 | 10 floors | Hangzhou, China | Predict EEC | Predicting energy consumption with genetic-neuro fuzzy systems | Time series: ANN, GA-ANFIS | Sep 1989-Feb 1990 (hourly) and Oct-Nov 2009 (hourly) – OT, SR, RH, WS | N/A |
| Tsanas and Xifara [58] | 12 simulated buildings. | 771.75 | | Athens, Greece | Predict heating load and cooling load | Analysis of energy performance in buildings by machine learning | Regression: Ensemble learning (Random Forests), IRLS | Simulation, 768 samples, (surface, wall, roof areas, height, orientation, glazing) | N/A |
| Castelli et al. [59] | Same than [58] | 771.75 | | Same than [58] | Same than [58] | Analysis of energy performance in buildings by evolutionary computation | Regression: Genetic Programming (with local search and linear scaling) | Same than [58] | N/A |
| Category: Grey Box/Hybrid models | | | | | | | | | |
| Raftery, Keane, and O'Donnell [61] | Office | 30,000 | 4 floors | Leixlip, Ireland | Dynamic Simulation Model Calibration (Methodology) | A systematic evidence-based methodology for the calibariotn of dynamic simulation models, using E+ and SVN tools | Dynamic Simulation with systematic version control | All E+ with hourly measurements of plugs & lights electrical energy consumption | EnergyPlus, TortoiseSVN |
| Raftery, Keane, and Costa [62] | Office | 30,000 | 4 floors | Leixlip, Ireland | Dynamic Simulation Model Calibration (Case study) | Application case study of the methodology proposed in [61] | Dynamic Simulation with systematic version control | All E+ with hourly measurements of plugs & lights electrical energy consumption | EnergyPlus, TortoiseSVN |
| Coakley et al. [63] | Office | 700 | 3 floors | Galway, Ireland | Dynamic Simulation Model Calibration (Methodology) | A methodology for the calibariotn of dynamic simulation models combining evidence-based model developent with statistical optimization techniques | Uncertainty analysis on Monte Carlo based multiple dynamic simulation outputs | All E+ with hourly measurements of space temp, CO2, electrical & heat energy consumption | OpenStudio, EnergyPlus |
| Coakley, Raftery, and Molloy [64] | Office | 700 | 3 floors | Galway, Ireland | Dynamic Simulation Model Calibration (Case study) | Application case study of the methodology proposed in [67] | Uncertainty analysis on Monte Carlo based multiple dynamic simulation outputs | All E+ with hourly measurements of space temp, CO2, electrical & heat energy consumption | OpenStudio, EnergyPlus |
| Mustafaraj et al. [65] | Office with underfloor heating | 4500 | 3 floors | Cork, Ireland | Dynamic Simulation Model Calibration | Methodology to calibrate the dynamic models for predicting the thermal behavior of underfloor heating, heat-pump and natural ventilation | Dynamic Simulation | All E+ with monthly electricity & gas bills and hourly measurements of temperature & humidity | DesignBuilder, EnergyPlus |

Table 1 (*Continued*)

| Publication/Reference | Building description | | | | Aim of the work | Key contribution | Algorithm/Method | Data variable | Software used |
|---|---|---|---|---|---|---|---|---|---|
| | Type of building | Size [m$^2$] | Number of floors | Location | | | | | |
| Roberti, Oberegger, and Gasparella [66] | Historical building | 3000 | 3 floors plus attic and basement | Bolzano, Italy | Dynamic Simulation Model Calibration | Sensitivity analysis on parameters of historiic buildings' models | Dynamic Simulation | All E+ with hourly space temperature | EnergyPlus |
| Royapoor and Roskilly [67] | Office | 8365 | 5 floor | Newcastle, UK | Dynamic Simulation Model Calibration | Building model calibration case study | Dynamic Simulation | All E+ with hourly mesurements of space temperature, electrical & gas energy consumption | EnergyPlus |
| Lü et al. [68] | 4 sport hauls | 3000 to 6000 | Single Volume | Finland | Simplified models for the prediction of energy consumption | Physical model with stochastic parameters | Physical, ARIMA, SVD, Convex hull | Qenvelope, Qsolar, Qventilation, Qoccupancy Qlighting | Matlab LibSVM |
| Heo, Choudhary, and Augenbroe [69] | Office | N/A | N/A | Cambridge, UK | Calibration of normative energy models for retrofit analysis | Uncertainty quantification and calibration of quasi steady state ISO normative model | Bayesian Calibration | Envelope thermal properties, internal loads, ventilation, HVAC generation efficiency & distribution loss factors | EnergyPlus (for validation and performance comparison) |
| Brohus et al. [70] | 5 zone test model | N/A | 1 | N/A | Uncertainty quantification in a physical model | Uncertainty quantification by means of stochastic differential equations | Stochastic Differential Equations | Zone thermal capacity, specific heat loss, temperature, internal heat generation | |
| L. Wang, Mathew, and Pang [71] | DOE reference office building | 4982 | 3 | Multiple (virtual) | Uncertainty quantification in a dynamic simulation model | Investigation of uncertainties and understanding of the impacts of key operation parameters in energy consumption | Monte Carlo based multiple dynamic simulations | All E+ | EnergyPlus |
| Zhao et al. [74] | DOE reference office building | 4982 | 3 | Multiple (virtual) | Create an occupant behavior and schedule modeling method | Development of an indirect practical data mining approach using office appliance power consumption data in order to learn the occupant behavior | Decision Tree, LWNB, SVM, LR, LWR and Dynamic Simulation | Occupancy Schedules | EnergyPlus |
| D'Oca and Hong [75] | Office | 17,402 | 5 | Frankfurt, Germany | Create an occupant behavior and schedule modeling method | Development of a three-step data mining framework to discover occupancy patterns in office spaces | Decision Tree, k-means | Occupancy Schedules | RapidMiner |

algorithms, especially related to model predictive control, although most of them are better classified as grey models, as in Refs. [23,24].

As mentioned before, white box models require a thermal engineering expert to model and interpret their results. The study of dynamic driving forces, made by the reviewed works, clearly calls for a previous extended knowledge on dynamic heat transmission in buildings. The output of these models is not directly interpretable by building managers and the adaptation of the technique results to manageable information is done under the user criteria, overlooking sometimes interesting information for building managers. These models perform an exhaustive modeling of outdoor conditions, requiring detailed data (usually hourly data) on solar radiation, outdoor temperatures and wind velocity. The access to these databases is not always free for researchers and users. In addition, it is common to find data gaps for specific locations far from important cities and climate stations, making it difficult to model buildings in certain locations.

### 3.2. Black box/Statistical/Inverse models

#### 3.2.1. Conventional statistics/Regression-based models

An improved method for the application of Cooling Degree Days (CDD), base temperature determination and CDD calculation technique including latent loads is presented in [25]. An approach to simplify and avoid detailed hourly simulations that uses predetermined coefficients to be applied to monthly energy consumption actual data from energy bills is presented in [26], showing errors within 10% [27].

Simplified regression models producing required data by dynamic simulation can overcome lack of adequate measurement data as in [28], utilizing different regression inputs for 16 French cities, with the deviation between predicted and simulated results being 5.1% with average error of 2%. The same methodology with different regression inputs is used in [29,30]. An extensive Monte Carlo simulation campaign is used in [30], with the regression equation showing a maximum error less than 5% to simulation outputs. The Monte Carlo simulation with the DOE-2 simulator generating 30 thousand design parameter combinations and using 17 key building design variables is presented in [31], with the resulting statistical analysis of data including stepwise regression, linear regression equations and the most effective parameter sensitivity analysis.

Estimation of heat and electricity load profiles based on regression analysis (heat load) and statistical analysis (electricity load) of district heat and electricity consumption measurements is discussed in [32]. Various regression analyses are performed in [33], suggesting the use of both the coefficient of determination and the root mean square error metrics for model quality comparison and assessment. A computationally efficient autoregressive model for thermal load prediction using different sets of coefficients is presented in [34], validating prediction accuracy with the EnergyPlus® simulator.

The Principal Component Regression can solve multi-collinearity effects transforming collinear variables to orthogonal components [35]; the method is validated through infrared thermography showing superiority against statistical rating method. The prediction accuracy of cooling load in office buildings can be improved by simultaneous application of Principal Component Analysis of meteorological factors, cumulative effect of high temperatures and dynamic two step correction; the validation was done in Tianjin office buildings showing a prediction accuracy of a mean absolute relative error less than 8% [36].

Energy labeling data and primary energy consumption of Netherlands dwellings, with nearly 200k entries being used in a top-down approach [37,38] reveal different parameter influences of theoretical and actual gas and electricity consumptions. The PRInceton Scorekeeping Method (PRISM) is used to examine the energy-efficiency profile of individual single-family houses from Gainesville, Florida [39], by processing weather and usage data as inputs to an iterative regression approach computing energy efficiency parameters. Various regressions have been tried over building databases of Portfolio Manager/EnergyStar scoring applications; the most notable and relevant ones were those addressing US and Canada hospital population [40–42].

A model approach focusing on medical equipment and over 33,500 h of measurement in the University Medical Centre of Hamburg shows that cumulative load predictions for an entire building are possible with an error of less than 6% [43]. The overall energy footprint of a CT scan is calculated in [44].

The stochastic nature of lighting energy use due to occupant behavior is analyzed in [45], based on relative measurements from 15 large Beijing and Hong Kong office buildings and a stochastic lighting energy use model is proposed to improve simulation accuracy. Similarly, [46] a stochastic model is proposed to be used in simulations of residential building cases.

A specific usage of disaggregation techniques for energy bills has been studied in [47], proposing an optimization algorithm to establish best possible cooling energy balances and disaggregate energy consumption of different users. The algorithm has been validated through cooling season measurement data from two Hong Kong and Beijing buildings.

Although allowing a detailed energy consumption comprehension, the statistics and regression-based techniques rely on a large amount of historical information, apart from the data needed to calibrate and validate the model. This sort of information may not be always available to users, both due to technical and managerial reasons. One of the main limitations is that energy consumption must be assigned beforehand to end-uses, lacking the chance of detecting marginal consumers [4]. Furthermore, these techniques require a former estimation of occupant behavior, taking into account the demonstrated variability in determining occupant behavior in building energy modeling.

Although these techniques perform accurate predictions and reduce error from 6% [43] to 2% [28], they are not the best option for detecting the reasons of consumption and designing energy saving measures, as the models are more focused in the prediction, rather than the identification of energy saving opportunities.

The Regression analysis is a validated technique for explaining major consumers in buildings. However, residuals are usually not accurately explained, as no specific pattern is found [29].

There is still a gap in knowledge about regression techniques for explaining residuals, those smalls energy consumers that, although may not be significant in amount, reflect non-considered phenomena in the buildings that hide behaviors or appliances beyond the building manager's control. The accurate consideration and explanation of residuals is still an open research challenge in the evolution of these techniques.

#### 3.2.2. Machine learning

A large number of papers exploit the potential of Artificial Neural Networks (ANN) in energy consumption predictions. A multi-layer perceptron ANN, based on a backpropagation training algorithm for load prediction is presented in [48]. Various types of ANNs for the prediction of the heating energy consumption of a university campus are studied in [49], trained and tested on actual measurement data; usage of an ensemble of more than one types leads to better results. Different Matlab implemented neural network topologies for the prediction of outdoor air temperatures using data from four European cities are shown in [50]. A short-term (15 min) forecasting model for a commercial building energy usage based on an ANN with Bayesian regularization is presented in [51].

An Adaptive Neural Fuzzy Inference System (ANFIS) is proposed for residential lighting load prediction in [52], showing better correlation and root mean square error to regression models on metered data. The ANN and Case-Based Reasoning (CBR) techniques were used for an hourly electricity consumption prediction in a Canadian facility in [53], with ANN models outperforming CBR models. Simplified CBR (S-CBR) was applied on the energy performance of a Seoul school building and validated in [54]. A comparison of the application of 7 machine learning techniques to data from the residential and commercial building sector is presented in [55], with ANN-based methods performing better in the commercial building and the Least Square Support Vector Machines outperforming ANNs in the residential buildings.

Besides the ANNs, other approaches receive attention in literature. A decision tree-based predictive model is presented in [56], facilitating the easy extraction of information, accurately predicting building energy demand levels (92%) and providing a combination of factors and thresholds, leading to high building energy performances. A hybrid Genetic Algorithm-Adaptive Network based Fuzzy Inference System (GA-ANFIS) is presented in [57], providing better prediction accuracy to ANNs. A random forest-based statistical machine learning framework is used in [58] to estimate heating and cooling load, validated through simulation of 768 residential buildings. The same dataset is used in [59] for residential building load estimation, via a genetic programming-based framework combined with a local search method and linear scaling.

Similar to regression techniques, a large amount of historical data is necessary for training and predicting energy consumption. Another limitation is the time-consuming calculations and specific software tools user-expertise. One disadvantage observed in the reviewed works is the need for pre-processing a large amount of data in order to decide the number of networks before building the model [48] or identify significant variables and outliers [49]. Like regression models, machine learning techniques do not give any explanation to outliers as residual data is removed in the preprocessing analysis [49].

Such techniques could predict energy consumption more accurately. However, they face the challenge of a deeper explanation of existing phenomena, as they do not calculate dynamic heat transfer phenomena. In fact, from the user point of view, it could be claimed that machine learning techniques represent an opposite approach to white box and physical techniques.

Machine learning developers also face the problem of easy generalization to different buildings without requiring significant change of the model or by endangering the precision of predictions.

### 3.3. Grey box/Hybrid models

A recent review of approaches to model calibration is presented in [60], assessing various analytical and mathematical/statistical tools. Yet, no consensus exists on standard calibration procedures and methods to be generally used on a variety of buildings.

A systematic evidence-based methodology for calibration of simulation models is presented in [61–64]. Parameter values reference the source of information used to make changes to the initial model, using version control software to store the records of the calibration process. A demonstration case calibrating an Ireland Intel campus four-floor office building is presented in [62], with the results showing excellent correlation with measured HVAC consumption data. The methodology is combined with statistical Monte Carlo-based optimization techniques in [63,64], applied in a naturally ventilated library building at the National University of Ireland, Galway.

A detailed example of calibration flow for an EnergyPlus® simulation of a building with underfloor heating system and natural ventilation is shown in [65], taking into account heat pump, energy consumption and zone temperature measurements. The possibility of poor calibrated models based on only one measured parameter is shown in [66], showcased for a medieval building EnergyPlus model. A similar example appears in [67], where a set of two environmental sensors and a weather station are used for annual space air temperature predictions.

A hybrid physical–statistical approach is described in [68], where stochastic parameters are introduced into the physical model and the statistical time series model is formulated to reflect model uncertainties, while a methodology based on Bayesian calibration of the normative EN ISO 13790 energy models is presented in [69], focusing on model parameter uncertainty quantification to generate probabilistic predictions of retrofit performances. The uncertainty is also quantified in [70] by means of stochastic differential equations applied to a general heat balance for an arbitrary number of loads and zones in a building, to determine the dynamic thermal response under random conditions. Uncertainty in energy consumption due to actual weather and building operational practices is investigated in [71], using simulation-based analysis of a medium size office building and Monte Carlo sampling of possible parameter combinations.

The need for more accurate occupant behavior models is among the results of [72], showing differences of 50% in average between design time predicted energy use of a low-energy building in Sweden, obtained through dynamic simulation, and actual measurements after tenants moved in. State-of-the-art occupant-related data collection and monitoring, modeling approaches, model evaluation, and model implementation into simulation tools is presented in [73].

An indirect data mining approach to learn occupant passive behavior and create the occupancy schedules of the EnergyPlus dynamic simulator is also presented in [74]. A similar data mining framework in presented in [75], where a learning process is used to extrapolate office occupancy patterns and working user profiles from big data streams in order to feed typical building energy modeling tools.

Accurate occupant behavior models deal with difficulties in the acquisition of information from building occupants. These models rely on their responses for a first modeling stage but need an exhaustive fitting once the first results are obtained. The step of occupant interviews and response analysis is also time-consuming. In addition, occupants are not always accessible for interview (for example in medical buildings).

Visual-based approaches such as the Energy Performance Augmented Reality are considered as powerful tools to know the real state of behavior of the building. The authors of [76] proposed a model approach that combines digital and thermal imagery with fluid dynamics models. The approach proposed consisted of three parts: 1) thermal data and digital building data collection with a thermal camera; 2) building energy performance simulation through a computational fluid dynamics analysis; 3) both models are superimposed in a common 3D environment, obtaining reasonable accuracy. In [77] this model was also used to visualize deviations between buildings' state and simulated energy performances and visualize the potential performance problems in the Energy Performance Augmented Reality environment. The model identified thermal bridges in the tested rooms.

In [78] the authors used a Graphic Processing Unit structured by Motion and Multi-View Stereo algorithms to reconstruct in 3D the geometrical conditions of the building that was studied. Then, this model was superimposed to a 3D thermal point model. The model was used to represent six interior and exterior spaces, concluding that thermal imagery is a feasible and relatively quick method for analysing the actual energy performance of existing buildings. In [79] this method was used to conduct a cost-benefit analysis of different retrofit alternatives of two existing buildings. The results

**Table 2**
Comparison of selected approaches.

| Approach | Simplicity | Completeness | Generality | Usefulness | Innovation |
|---|---|---|---|---|---|
| Bill-based methods | High | Medium | High | Low | Low |
| Monitoring-based methods | Low | Medium | Medium | High | High |
| Dynamic simulations | Low | High | Low | High | Medium |

demonstrated the reliability and accuracy of the method in estimating the return on investment from retrofitting thermal performance problems.

Visual models containing thermal values facilitate the recognition of temperature distribution and the detection of building performance failures. These methods facilitate the detection of building performance deviations and identify disparities between building information and real conditions. Usually, these techniques are combined with more detailed approaches in order to extract information from the visual analysis.

Vision based methods have potential in reducing time and effort in collecting data and high level of accuracy in detecting thermal bridges and defaults in the building. These methods present an adequate balance between effort and quality of the analysis that they perform, and they also present a great advantage by facilitating the visualization of the data and the immediacy of their analysis. To the contrary of other approaches, visual-based methods do not require detailed previous information of the building in order to provide immediate results without the need of exhaustive data analysis. These methods accurately examine the exterior energy performance of the building in real time. However, they are not easily applicable to interior performance and generally they need to be supported by another approach.

Compared with the approaches cited in previous sections, these models have the limitation that are not applicable in all the project's phases, but only in the operation phase of existing buildings. Some aspects still need to be improved: for example, achieving more accuracy and reliability in the identification of the threshold for performance detection under different external and internal conditions. These approaches require an exhaustive on-site inspection of the building, and some drawbacks could come across during this process, such as difficulties accessing some rooms or conflict with the performed activities (for example, in educative buildings or medical centres).

## 4. Discussion on the suitability of approaches

Some authors [6,80] provide qualitative comparison frameworks for the identified methods on the axis of the application and use-case on the level of building details or on the amount of measurement data needed, on the computation time and on the level of insight to the underlying physical processes revealed. Quantitative comparisons exist in the literature too, but they are inherently less generic, as they must compare a restricted set of explicit method instances (i.e. explicit model implementations) [7–9].

In this work, we follow a horizontal, selective but highly generic view. We sort out three of the presented approaches: bill-based methods, monitoring-based methods and dynamic simulations offering a comparison against a set of specific parameters as shown in Table 2 and Fig. 2.

The selected approaches are characterized based on the following features:

- Simplicity: inversely relates to development effort, the total work done to apply the approach, the required information volume, specialized skills of staff, need of an interdisciplinary team, etc. Lower values of these concepts lead to a higher simplicity (lower complexity) which is preferable, as it has a higher guarantee of being successfully and on time applied.
- Completeness: is the quality of explaining the total reality involved in an energy consumption system. This property depends largely on the degree of specificity reached and can vary significantly among different methods of the same approach.
- Generality: stands for the quality of the obtained results, being general enough as to be useful for a standardized comparison among different buildings. Higher generality is preferable as the effort to extrapolate conclusions is lower and easiness to introduce the approach in new buildings is greater.
- Usefulness: relates to the utility of the derived knowledge for making decisions on energy efficiency strategies. Models that discover complex and interesting variable relationships and get further insight are preferable, since they represent an advance in the field. The level of detail of the results of the models compared in this article is variable. This parameter evaluates the exploitation of the results and predictions obtained by the model for its use in a later analysis, especially its applicability for the decision-making in the prioritization of economic investments in order to reduce the energy demand of the studied building. This feature also values the utility of results of each model for the stakeholders in investments for energy efficiency in buildings.
- Innovation: represents the space to provide original results by using cutting edge techniques. Although the field of energy modeling has been highly explored and refined in recent years, as it has been pointed out in the critical analysis of the limitations of each of the approaches compared, there are still open research challenges that need to be addressed in the future. This feature evaluates the degree of flexibility that each approach presents in order to improve itself and the introduction of new tools to continue the innovation in its field of application.

Table 2 summarizes the level of achievement of the five features above explained by each model compared. Three levels of achievement are identified: low, medium and high.

This tabulation system allows to clearly differentiate the strengths and weaknesses of each of the approaches compared. The evaluation has been made based on: the literature review made in the previous sections, the critical analysis performed during this literature review, the study of the depth of detail of the works reviewed. A low level means that the feature is not an identifiable or achieved characteristic by the model. A medium level indicates that the model presents this feature, although with deficiencies or shows obvious improvement. Finally, a high level means that this feature is clearly identifiable at a satisfactory level during the use of the model and the results obtained.

From Table 2, it can be seen that the bill-based methods reach a high level of simplicity and generality, while an intermediate level in completeness is reported. However, it shows a low score on both usefulness and innovation. Bill-based methods are more easily applicable and, therefore, more general; nonetheless, they do not go beyond the state-of-the-art, so the innovation degree is poor.

The level of simplicity of monitoring-based methods is low, as it can be hard to implement (depending on the measurement they may require sensors or specific information) and may be difficult to extrapolate to other environments. However, margin to innovate is very good and the obtained knowledge very useful for decision-making, reaching a high score on both usefulness and

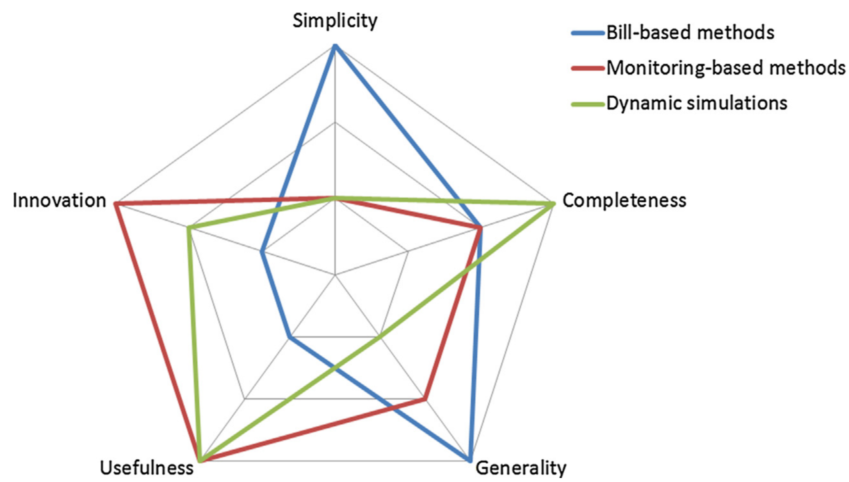


**Fig. 2.** Comparison of selected approaches.

innovation. Finally, dynamic simulations need an important development effort and are hardly general; however, they provide the most detailed description of energy use distribution and are useful for energy retrofitting in buildings, showing a medium score in general. Finally, dynamic simulations reach high levels of completeness, as they allow a detailed description of the energy consumption of the building. Dynamic simulations also reach high levels in usefulness, as the results obtained are effective and applicable in decision-making regarding economic investment in energy retrofit of buildings. They have a medium score in innovation (halfway between detailed monitoring-based methods and generalists bill-based methods). Regarding simplicity, they are characterized by a low score given that specific knowledge by the user is necessary prior to perform a simulation. They also obtain a low score in general, since these models require specific details of the building (especially regarding construction materials and occupation profiles) and their exportation to other buildings is not direct and requires a detailed change of parameters.

The advantages of the three approaches are fairly matched, so choosing the best method is a matter of importance of the aforementioned properties. To guarantee minimal and general results, bill-based methods seem the best option; in order to innovate, monitoring-based methods are recommended; for obtaining the deepest knowledge, the dynamic system method is preferable.

With a level of intermediate effort in data collection and by attributing much of the quality of the taken information to the user instead of to the existence of monitoring systems (as in machine learning), hybrid models allow obtaining predictions with low error rates. In addition, the approach is useful for identifying opportunities for energy saving.

Fig. 2 compares these approaches according to the five proposed criteria.

A known barrier among the open research challenges in delivering optimal hybrid models is the data collection process. Machine learning and calibrated methods need detailed metered information from the building, usually collected by advanced meters, whose cost is still not feasible for most of the buildings or housing owners. In order to achieve a higher market penetration of such meters, the challenge of their cost reduction has to be met. Furthermore, model predictions are necessary to be compared with real energy bills. Researchers usually find barriers in accessing such information, usually stored by energy companies. Access to larger portions of information on energy consumption of districts or cities would provide a starting point to implement accurate predictive models at high scale. This would in turn help the identification of big consumers and the implementation of specific energy saving measures at district level. This is also associated to a challenge in the legal dimension, in order to make such data available to the research community, without including sensitive information.

## 5. Concluding remarks

A revision of existing approaches for modeling energy consumption and efficiency in buildings has been conducted.

The main features that characterize the methodologies are identified. A performance analysis of the methodologies is conducted, and a rating system is proposed. According to this rating, to guarantee minimal and general results, bill-based methods are the best option. Measurement-based methods present higher degree of innovation, whereas to get the deepest knowledge, dynamic system modeling is the best option.

This assessment methodology facilitates the comparison of different approaches when energy modeling in buildings is concerned. The selection of the most appropriate method is relevant to the individual expectations and needs.

A hybridization of the analyzed approaches could offer a more complete solution, by taking profit of their main advantages and mitigating their individual drawbacks. In this context, bill-based methods could be utilized to set dynamic models that can be subsequently optimized by measurement-based methods.

## Acknowledgments

The work presented in this paper has been funded in the framework of "Support Tool for Energy Efficiency Programmes in medical centres – STEER" project, Grant Agreement 655694, Research and Innovation Staff Exchange (RISE), H2020 – MSCA – RISE – 2014, European Commission.

## References

[1] U.S. Energy Information International, Administration Energy Outlook 2016, 2016.
[2] EU, Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings (recast), Off. J. Eur. Union. (2010) 13–35. 10.3000/17252555.L_2010.153.eng.
[3] US Department of Energy, Energy Efficiency Trends in Residential and Commercial Buildings, 2008.
[4] L.G. Swan, V.I. Ugursal, Modeling of end-use energy consumption in the residential sector: a review of modeling techniques, Renew. Sustain. Energy Rev. 13 (2009) 1819–1835, http://dx.doi.org/10.1016/j.rser.2008.09.033.
[5] U.S. Department of Energy, M&V Guidelines: Measurement and Verification for Performance-Based Contracts Version 4.0, 2015.

[6] A. Foucquier, S. Robert, F. Suard, L. Stéphan, A. Jay, State of the art in building modelling and energy performances prediction: a review, Renew. Sustain. Energy Rev. 23 (2013) 272–288, http://dx.doi.org/10.1016/j.rser.2013.03.004.
[7] G.K.F. Tso, K.K.W. Yau, Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks, Energy 32 (2007) 1761–1768, http://dx.doi.org/10.1016/j.energy.2006.11.010.
[8] A.H. Neto, F.A.S. Fiorelli, Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption, Energy Build. 40 (2008) 2169–2176, http://dx.doi.org/10.1016/j.enbuild.2008.06.013.
[9] A. Afram, F. Janabi-Sharifi, Black-box modeling of residential HVAC system and comparison of gray-box and black-box modeling methods, Energy Build. 94 (2015) 121–149, http://dx.doi.org/10.1016/j.enbuild.2015.02.045.
[10] S. Wang, C. Yan, F. Xiao, Quantitative energy performance assessment methods for existing buildings, Energy Build. 55 (2012) 873–888, http://dx.doi.org/10.1016/j.enbuild.2012.08.037.
[11] J.K. Kissock, J.S., Haberl, D.E. Claridge, Inverse modeling toolkit: User's guide. ASHRAE Final Report for RP-1050 (2001).
[12] U.S. Department of Energy, Characteristics and Performance of Existing Load Disaggregation Technologies, 2015.
[13] American Society of Heating Refrigerating and Air-Conditioning Engineers, ASHRAE Handbook -Fundamentals. Chapter 19: Energy estimating and modeling methods, in: Chapter 19 Energy Estim. Model. Methods (2013).
[14] S. Yoshida, K. Ito, R. Yokoyama, Sensitivity analysis in structure optimization of energy supply systems for a hospital, Energy Convers. Manage. 48 (2007) 2836–2843, http://dx.doi.org/10.1016/j.enconman.2007.06.045.
[15] X. Yang, L. Zhao, M. Bruse, Q. Meng, An integrated simulation method for building energy performance assessment in urban environments, Energy Build. 54 (2012) 243–251, http://dx.doi.org/10.1016/j.enbuild.2012.07.042.
[16] I. Ballarini, V. Corrado, Analysis of the building energy balance to investigate the effect of thermal insulation in summer conditions, Energy Build. 52 (2012) 168–180, http://dx.doi.org/10.1016/j.enbuild.2012.06.004.
[17] B.L. Gowreesunker, S.A. Tassou, M. Kolokotroni, Improved simulation of phase change processes in applications where conduction is the dominant heat transfer mode, Energy Build. 47 (2012) 353–359, http://dx.doi.org/10.1016/j.enbuild.2011.12.008.
[18] G. Dall'O', L. Sarto, N. Sanna, A. Martucci, Comparison between predicted and actual energy performance for summer cooling in high-performance residential buildings in the Lombardy region (Italy), Energy Build. 54 (2012) 234–242, http://dx.doi.org/10.1016/j.enbuild.2012.08.005.
[19] G. Dall'O', L. Sarto, A. Galante, G. Pasetti, Comparison between predicted and actual energy performance for winter heating in high-performance residential buildings in the Lombardy region (Italy), Energy Build. 47 (2012) 247–253, http://dx.doi.org/10.1016/j.enbuild.2011.11.046.
[20] P. Michalak, The simple hourly method of EN ISO 13790 standard in Matlab/Simulink: a comparative study for the climatic conditions of Poland, Energy 75 (2014) 568–578, http://dx.doi.org/10.1016/j.energy.2014.08.019.
[21] R. De Lieto Vollaro, C. Guattari, L. Evangelisti, G. Battista, E. Carnielo, P. Gori, Building energy performance analysis: a case study, Energy Build. 87 (2015) 87–94, http://dx.doi.org/10.1016/j.enbuild.2014.10.080.
[22] V. Čongradac, B. Prebiračević, N. Jorgovanović, D. Stanišić, Assessing the energy consumption for heating and cooling in hospitals, Energy Build. 48 (2012) 146–154, http://dx.doi.org/10.1016/j.enbuild.2012.01.022.
[23] G. Mantovani, L. Ferrarini, Temperature control of a commercial building with model predictive control techniques, IEEE Trans. Ind. Electron. 62 (2015) 2651–2660, http://dx.doi.org/10.1109/TIE.2014.2387095.
[24] L. Ferrarini, G. Mantovani, Modeling and control of thermal energy of a large commercial building, 2013 IEEE Int. Work. Intell. Energy Syst. (2013).
[25] G. Krese, M. Prek, V. Butala, Analysis of building electric energy consumption data using an improved cooling degree day method, J. Mech. Eng. 58 (2012) 107–114.
[26] N. Fumo, P. Mago, R. Luck, Methodology to estimate building energy consumption using EnergyPlus Benchmark Models, Energy Build. 42 (2010) 2331–2337, http://dx.doi.org/10.1016/j.enbuild.2010.07.027.
[27] A. Smith, N. Fumo, R. Luck, P.J. Mago, Robustness of a methodology for estimating hourly energy consumption of buildings using monthly utility bills, Energy Build. 43 (2011) 779–786, http://dx.doi.org/10.1016/j.enbuild.2010.11.012.
[28] T. Catalina, J. Virgone, E. Blanco, Development and validation of regression models to predict monthly heating demand for residential buildings, Energy Build. 40 (2008) 1825–1832, http://dx.doi.org/10.1016/j.enbuild.2008.04.001.
[29] T. Catalina, V. Iordache, B. Caracaleanu, Multiple regression model for fast prediction of the heating energy demand, Energy Build. 57 (2013) 302–312, http://dx.doi.org/10.1016/j.enbuild.2012.11.010.
[30] S. Asadi, S.S. Amiri, M. Mottahedi, On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design, Energy Build. 85 (2014) 246–255, http://dx.doi.org/10.1016/j.enbuild.2014.07.096.
[31] S.S. Amiri, M. Mottahedi, S. Asadi, Using multiple regression analysis to develop energy consumption indicators for commercial buildings in the U.S, Energy Build. 109 (2015) 209–216, http://dx.doi.org/10.1016/j.enbuild.2015.09.073.
[32] L. Pedersen, J. Stang, R. Ulseth, Load prediction method for heat and electricity demand in buildings for the purpose of planning for mixed energy distribution systems, Energy Build. 40 (2008) 1124–1134, http://dx.doi.org/10.1016/j.enbuild.2007.10.014.
[33] N. Fumo, M.A. Rafe Biswas, Regression analysis for prediction of residential energy consumption, Renew. Sustain. Energy Rev. 47 (2015) 332–343, http://dx.doi.org/10.1016/j.rser.2015.03.035.
[34] K. Yun, R. Luck, P.J. Mago, H. Cho, Building hourly thermal load prediction using an indexed ARX model, Energy Build. 54 (2012) 225–233, http://dx.doi.org/10.1016/j.enbuild.2012.08.007.
[35] E. Wang, Z. Shen, K. Grosskopf, Benchmarking energy performance of building envelopes through a selective residual-clustering approach using high dimensional dataset, Energy Build. 75 (2014) 10–22, http://dx.doi.org/10.1016/j.enbuild.2013.12.055.
[36] G. Qiang, T. Zhe, D. Yan, Z. Neng, An improved office building cooling load prediction model based on multivariable linear regression, Energy Build. 107 (2015) 445–455, http://dx.doi.org/10.1016/j.enbuild.2015.08.041.
[37] D. Majcen, L. Itard, H. Visscher, Actual and theoretical gas consumption in Dutch dwellings: what causes the differences? Energy Policy 61 (2013) 460–471, http://dx.doi.org/10.1016/j.enpol.2013.06.018.
[38] D. Majcen, L. Itard, H. Visscher, Statistical model of the heating prediction gap in Dutch dwellings: relative importance of building, household and behavioural characteristics, Energy Build. 105 (2015) 43–59, http://dx.doi.org/10.1016/j.enbuild.2015.07.009.
[39] E. Hoşgör, P.S. Fischbeck, Virtual home energy auditing at scale: predicting residential energy efficiency using publicly available data, Energy Build. 92 (2015) 67–80, http://dx.doi.org/10.1016/j.enbuild.2015.01.037.
[40] US Environmental Protection Agency, ENERGY STAR Score for Hospitals in Canada, 2014.
[41] US Environmental Protection Agency, ENERGY STAR Score for Hospitals in the United States, 2014.
[42] US Environmental Protection Agency, DataTrends: Energy Use in Hospitals, 2015.
[43] N. Christiansen, M. Kaltschmitt, F. Dzukowski, F. Isensee, Electricity consumption of medical plug loads in hospital laboratories: identification, evaluation, prediction and verification, Energy Build. 107 (2015) 392–406, http://dx.doi.org/10.1016/j.enbuild.2015.08.022.
[44] A. Esmaeili, J.M. Twomey, M.R. Overcash, S.A. Soltani, C. McGuire, K. Ali, Scope for energy improvement for hospital imaging services in the USA, J. Health Serv. Res. Policy 20 (2015) 67–73, http://dx.doi.org/10.1177/1355819614554845.
[45] X. Zhou, D. Yan, T. Hong, X. Ren, Data analysis and stochastic modeling of lighting energy use in large office buildings in China, Energy Build. 86 (2015) 275–287, http://dx.doi.org/10.1016/j.enbuild.2014.09.071.
[46] E.J. Palacios-Garcia, A. Chen, I. Santiago, F.J. Bellido-Outeiriño, J.M. Flores-Arias, A. Moreno-Munoz, Stochastic model for lighting's electricity consumption in the residential sector Impact of energy saving actions, Energy Build. 89 (2015) 245–259, http://dx.doi.org/10.1016/j.enbuild.2014.12.028.
[47] C. Yan, S. Wang, F. Xiao, A simplified energy performance assessment method for existing buildings based on energy bill disaggregation, Energy Build. 55 (2012) 563–574, http://dx.doi.org/10.1016/j.enbuild.2012.09.043.
[48] A. Bagnasco, F. Fresi, M. Saviozzi, F. Silvestro, A. Vinci, Electrical consumption forecasting in hospital facilities: an application case, Energy Build. 103 (2015) 261–270, http://dx.doi.org/10.1016/j.enbuild.2015.05.056.
[49] R.Ž. Jovanović, A.A. Sretenović, B.D. Živković, Ensemble of various neural networks for prediction of heating energy consumption, Energy Build. 94 (2015) 189–199, http://dx.doi.org/10.1016/j.enbuild.2015.02.052.
[50] S. Papantoniou, D.-D. Kolokotsa, Prediction of outdoor air temperature using neural networks: application in 4 European cities, Energy Build. 114 (2016) 72–79, http://dx.doi.org/10.1016/j.enbuild.2015.06.054.
[51] Y.T. Chae, R. Horesh, Y. Hwang, Y.M. Lee, Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings, Energy Build. 111 (2016) 184–194, http://dx.doi.org/10.1016/j.enbuild.2015.11.045.
[52] O. Popoola, J. Munda, A. Mpanda, Residential lighting load profile modelling, Energy Build. 90 (2015) 29–40, http://dx.doi.org/10.1016/j.enbuild.2015.01.005.
[53] R. Platon, V.R. Dehkordi, J. Martel, Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis, Energy Build. 92 (2015) 10–18, http://dx.doi.org/10.1016/j.enbuild.2015.01.047.
[54] C. Koo, T. Hong, A dynamic energy performance curve for evaluating the historical trends in the energy performance of existing buildings using a simplified case-based reasoning approach, Energy Build. 92 (2015) 338–350, http://dx.doi.org/10.1016/j.enbuild.2015.02.004.
[55] R.E. Edwards, J. New, L.E. Parker, Predicting future hourly residential electrical consumption: a machine learning case study, Energy Build. 49 (2012) 591–603, http://dx.doi.org/10.1016/j.enbuild.2012.03.010.
[56] Z. Yu, F. Haghighat, B.C.M. Fung, H. Yoshino, A decision tree method for building energy demand modeling, Energy Build. 42 (2010) 1637–1646, http://dx.doi.org/10.1016/j.enbuild.2010.04.006.
[57] K. Li, H. Su, J. Chu, Forecasting building energy consumption using neural networks and hybrid neuro-fuzzy system: a comparative study, Energy Build. 43 (2011) 2893–2899, http://dx.doi.org/10.1016/j.enbuild.2011.07.010.
[58] A. Tsanas, A. Xifara, Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools, Energy Build. 49 (2012) 560–567, http://dx.doi.org/10.1016/j.enbuild.2012.03.003.
[59] M. Castelli, L. Trujillo, L. Vanneschi, A. Popovič, Prediction of energy performance of residential buildings: a genetic programming approach, Energy Build. 102 (2015) 67–74, http://dx.doi.org/10.1016/j.enbuild.2015.05.013.

[60] D. Coakley, P. Raftery, M. Keane, A review of methods to match building energy simulation models to measured data, Renew. Sustain. Energy Rev. 37 (2014) 123–141, http://dx.doi.org/10.1016/j.rser.2014.05.007.
[61] P. Raftery, M. Keane, J. O'Donnell, Calibrating whole building energy models: an evidence-based methodology, Energy Build. 43 (2011) 2356–2364, http://dx.doi.org/10.1016/j.enbuild.2011.05.020.
[62] P. Raftery, M. Keane, A. Costa, Calibrating whole building energy models: detailed case study using hourly measured data, Energy Build. 43 (2011) 3666–3679, http://dx.doi.org/10.1016/j.enbuild.2011.09.039.
[63] D. Coakley, P. Raftery, P. Molloy, G. White, Calibration of a detailed bes model to measured data using an evidence-based analytical optimisation approach, Proc. Build. Simul. 2011 12th Conf. Int. Build. Perform. Simul. Assoc. (2011).
[64] D. Coakley, P. Raftery, P. Molloy, Calibration of whole building energy simulation models: detailed case study of a naturally ventilated building using hourly measured data, in: 1st Build Simul. Optim. Conf., Loughbrgh, 2012.
[65] G. Mustafaraj, D. Marini, A. Costa, M. Keane, Model calibration for building energy efficiency simulation, Appl. Energy 130 (2014) 72–85, http://dx.doi.org/10.1016/j.apenergy.2014.05.019.
[66] F. Roberti, U.F. Oberegger, A. Gasparella, Calibrating historic building energy models to hourly indoor air and surface temperatures: methodology and case study, Energy Build. 108 (2015) 236–243, http://dx.doi.org/10.1016/j.enbuild.2015.09.010.
[67] M. Royapoor, T. Roskilly, Building model calibration using energy and environmental data, Energy Build. 94 (2015) 109–120, http://dx.doi.org/10.1016/j.enbuild.2015.02.050.
[68] X. Lü, T. Lu, C.J. Kibert, M. Viljanen, Modeling and forecasting energy consumption for heterogeneous buildings using a physical–statistical approach, Appl. Energy 144 (2015) 261–275, http://dx.doi.org/10.1016/j.apenergy.2014.12.019.
[69] Y. Heo, R. Choudhary, G.A. Augenbroe, Calibration of building energy models for retrofit analysis under uncertainty, Energy Build. 47 (2012) 550–560, http://dx.doi.org/10.1016/j.enbuild.2011.12.029.
[70] H. Brohus, C. Frier, P. Heiselberg, F. Haghighat, Quantification of uncertainty in predicting building energy consumption: a stochastic approach, Energy Build. 55 (2012) 127–140, http://dx.doi.org/10.1016/j.enbuild.2012.07.013.
[71] L. Wang, P. Mathew, X. Pang, Uncertainties in energy consumption introduced by building operations and weather for a medium-size office building, Energy Build. 53 (2012) 152–158, http://dx.doi.org/10.1016/j.enbuild.2012.06.017.
[72] F. Karlsson, P. Rohdin, M. Persso, Measured and predicted energy demand of a low energy building: important aspects when using Building Energy Simulation, Build. Serv. Eng. Res. Technol. 28 (2007) 223–235.
[73] D. Yan, W. O'Brien, T. Hong, X. Feng, H. Burak Gunay, F. Tahmasebi, A. Mahdavi, Occupant behavior modeling for building performance simulation: current state and future challenges, Energy Build. 107 (2015) 264–278, http://dx.doi.org/10.1016/j.enbuild.2015.08.032.
[74] J. Zhao, B. Lasternas, K.P. Lam, R. Yun, V. Loftness, Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining, Energy Build. 82 (2014) 341–355, http://dx.doi.org/10.1016/j.enbuild.2014.07.033.
[75] S. D'Oca, T. Hong, Occupancy schedules learning process through a data mining framework, Energy Build. 88 (2015) 395–408, http://dx.doi.org/10.1016/j.enbuild.2014.11.065.
[76] Y. Ham, M. Golparvar-Fard, EPAR, Energy Performance Augmented Reality models for identification of building energy performance deviations between actual measurements and simulation results, Energy Build. 63 (2013) 15–28, http://dx.doi.org/10.1016/j.enbuild.2013.02.054.
[77] M. Golparvar-Fard, Y. Ham, Automated diagnostics and visualization of potential energy performance problems in existing buildings using energy performance augmented reality models, J. Comput. Civ. Eng. 28 (2014) 17–29, http://dx.doi.org/10.1061/(ASCE)CP. 1943-5487.0000311.
[78] Y. Ham, M. Golparvar-fard, An automated vision-based method for rapid 3D energy performance modeling of existing buildings using thermal and digital imagery, Adv. Eng. Inform. 27 (2013) 395–409, http://dx.doi.org/10.1016/j.aei.2013.03.005.
[79] Y. Ham, M. Golparvar-Fard, Three-dimensional thermography-based method for cost-benefit analysis of energy efficiency building envelope retrofits, J. Comput. Civ. Eng. ASCE 29 (2015) B4014009, http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000406.
[80] N. Fumo, A review on the basics of building energy estimation, Renew. Sustain. Energy Rev. 31 (2014) 53–60, http://dx.doi.org/10.1016/j.rser.2013.11.040.