Journal of Advanced Research in Fluid Mechanics and Thermal Sciences 123, Issue 1 (2024) 69-85

Journal of Advanced Research in Fluid Mechanics and Thermal Sciences

Journal homepage: https://semarakilmu.com.my/journals/index.php/fluid_mechanics_thermal_sciences/index
ISSN: 2289-7879

A Comparative Study of Building Energy Consumption Prediction Methods

Mashila Amilludin[1,2], Azhan Ab. Rahman[1,2,*], Arfah Ahmad[1,2], Mohd Firdaus Mohd Ab. Halim[1,2], Norazlan Hashim[3], Mochammad Facta[4]

1 Fakulti Teknologi dan Kejuruteraan Elektrik, Universiti Teknikal Malaysia Melaka (UTeM), Melaka, Malaysia
2 Centre of Robotics and Industrial Automation, Universiti Teknikal Malaysia Melaka (UTeM), Melaka, Malaysia
3 School of Electrical Engineering, College of Engineering, Universiti Teknologi Mara (UiTM), Shah Alam, Malaysia
4 Fakultas Teknik, Universitas Diponegoro, Semarang, Indonesia

ARTICLE INFO

Article history:
Received 10 June 2024
Received in revised form 29 September 2024
Accepted 13 October 2024
Available online 30 October 2024

Keywords:
Temperature; energy; prediction method; building

ABSTRACT

Continued usage of fossil fuel-based electricity generation is one of the main factors for the greenhouse effect. As the greenhouse effect increase, the ambient temperature of the earth will also increase. As the temperature increase, the usage of air conditioning in buildings is also expected to increase. Since that more than 50 % of building energy consumption goes to air conditioning, it is utterly important to look into the methods that can be used to predict the impact of variations in temperature towards building energy consumption. This paper hence tries to perform a conceptual review on the prevalent methods that have been used for prediction of energy consumption, grouped in the category known as the white box, grey box and black box. The comparison of the different methods used in forecasting of energy consumption is discussed in terms of application, input, specific approach, advantage and disadvantage. It is hoped that the work carried out in this paper will be the reference to jump start the research work of others in this field of study.

1. Introduction

Buildings consume almost 40% of global electrical power on a daily basis as a result of fast expanding populations, growing economies, and advances in technology, according to Khan *et al.,* [1]. Buildings contribute significantly to global consumption of energy and greenhouse gas emissions, particularly carbon dioxide, which is responsible for environmental pollution, climate change, and global warming, according to Kunalan *et al.,* [2] and Norouziasl and Jafari [3]. Offices and other business buildings that use a lot of energy globally are particularly vulnerable to this problem.

According to Vijayan [4], energy utilisation, power consumption, and energy conservation are urgent concerns due to the rising demand for electricity around the world. In order to avoid overproduction and to improve energy management and efficiency, accurate demand forecasting is crucial. The importance of developing reliable models for consumption of energy prediction is

* *Corresponding author.*
E-mail address: azhanrahman@utem.edu.my

https://doi.org/10.37934/arfmts.123.1.6985

underscored by Mohapatra *et al.,* [5] in order to facilitate informed decision-making and effective energy management. In a similar vein, Daut *et al.,* [6] underscore the need for accurate industrial energy use predictions in order to optimise production plans and procedures while avoiding disruptions to ongoing business operations. In order to clarify the idea of energy consumption forecasting, this paper gives a thorough explanation of three different methodologies which are the white box, grey box, and black box methods. After that, it provides a thorough analysis of these key forecasting techniques in relation to building energy usage according to Chalal *et al.,* [7]. Each of these modeling methods contributes uniquely to the effectiveness of energy management systems.

Through a methodical approach to these goals, the study aims to offer significant perspectives on the range of the three approaches used which are white box, black box and grey box in energy consumption forecasting, emphasising their individual advantages and disadvantages. This study also concludes with a comprehensive comparison of the three approaches in the context of projecting energy consumption. Integration of these modelling methodologies allows energy management systems to be customised for different applications, increasing energy efficiency, cutting expenses, and increasing the viability and accessibility of renewable energy sources. This integration helps the larger objective of moving towards a more economical and sustainable energy future while also improving the performance of Energy Management System which is has investigated by Chalal *et al.,* [7]. Research gap for this paper is conducting in-depth comparative studies to assess the accuracy of white-box, black-box, and grey-box models across different building types, climates and operational scenarios.

2. Building Energy Consumption Prediction Method

Several modelling techniques are available for analysing building energy use, and each has advantages of its own depending on the needed level of accuracy and the data that is available. In building power prediction and modelling purposes, there are three basic approaches which are white box, black box and grey box according to Wang *et al.,* [8]. White-box models, developed by researchers with detailed knowledge of the building's internal mechanics, use physical equations to represent energy flows and consumption accurately. These models are highly reliable for optimization and control applications but require comprehensive data on material properties, geometrical dimensions, and operational conditions. In a study by Zhao and Magoulès [9], white box approaches calculate the consumption of energy by solving the heat and mass balance inside a building also referred to as a physical method because of its internal logic using thermodynamic principles. In contrast, black-box models rely on historical data and machine learning techniques to predict energy consumption without detailed knowledge of the building's internal structure. Developed by data researchers, these models are ideal for scenarios where quick, reasonably accurate predictions are needed, and detailed system information is not accessible. Black-box models, such as neural networks by Wang and Srinivasan [10], use past consumption data and external factors like weather conditions for their predictions. Grey-box models offer a middle ground, combining simplified physical equations with empirical data to balance accuracy and practicality. These models are useful when partial knowledge of the system is available, and some historical data can be used for calibration. By integrating both physical and data-driven approaches, grey-box models provide reliable energy consumption predictions, making them suitable for applications where both model accuracy and simplicity are important. For example, to increase the efficacy of HVAC control frameworks according to Foucquier *et al.,* [11], the grey box approach was first introduced in the 1990s, which blends white box and black box approaches which are stated in Bourdeau *et al.,* [12] and Lü *et al.,* [13]. Each method plays a crucial role in understanding and optimizing building energy consumption, tailored to specific needs and data availability. Figure 1 shows an illustration of understanding for energy consumption method for example white-box method, black-box method and grey-box method.

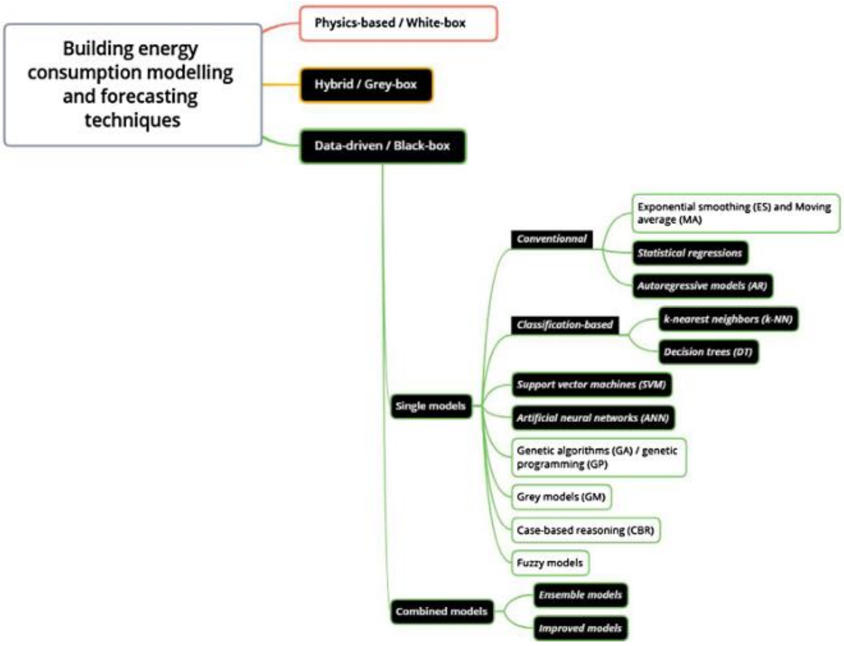

Fig. 1. Illustration of understanding for energy consumption method

2.1 White-box

First approach is white box. According to Chen *et al.,* [14], the terms white box, also known as physical model and building physical energy model. White-box models are typically developed by researchers who have extensive knowledge of the internal mechanics of the system being modelled. These models are often used in controlled environments such as laboratories, research institutions, and industries where detailed system data is available. White-box models rely on detailed physical equations to represent the system's behaviour, requiring comprehensive information about material properties, geometrical dimensions, and operational conditions. They are particularly useful when accurate predictions are essential, and sufficient data about the system is accessible. The primary advantage of white-box models lies in their ability to provide precise control and prediction of system behaviour, making them highly reliable for optimization and control applications. The process of developing a white-box model involves defining the system, gathering detailed data, formulating mathematical equations based on physical laws, calibrating the model with experimental data, and validating it through simulations to ensure accuracy. An example is a thermodynamic model of an HVAC system, which uses equations for heat transfer, fluid dynamics, and material properties to predict energy consumption under various operating conditions. In a study by Harish and Kumar [15], white-box models are those that were created to modelling subsystems for Heating, Ventilation, and Air Conditioning (HVAC) system. To easily generate and analyse these mathematical formulas, a variety of open-source or commercial software tools, including DOE2, EnergyPlus, TRNSYS, Dymola and Matlab, can be utilised for building energy modelling which are stated in Imam *et al.,* [16] and Ha *et al.,* [17]. It is difficult to gather all the information needed since many input factors are needed to construct an appropriate physical technique. More importantly, Blázquez-García *et al.,* [18] mention that there may be a discrepancy between the assumptions made by the model and reality. According to the hazy and untrustworthy data, various forms of energy have been the subject of studies on white box approaches which is stated by Oh *et al.,* [19], for example, energy use for a building lift by Li *et al.,* [20], electrical usage for cooling by Yu [21] and Kang *et al.,* [22], demand for energy by Ghedamsi *et al.,* [23] and Lee [24], prediction of load by Lindberg *et al.,* [25], electrical and thermal load by Szul and Kokoszka [26], and heating usage by Zhang *et al.,* [27].

Figure 2 is an example of the framework for EnergyPlus in term of white box approach, which is from the collection of input data to the simulation solver, which is using EnergyPlus and comes out with a result output.

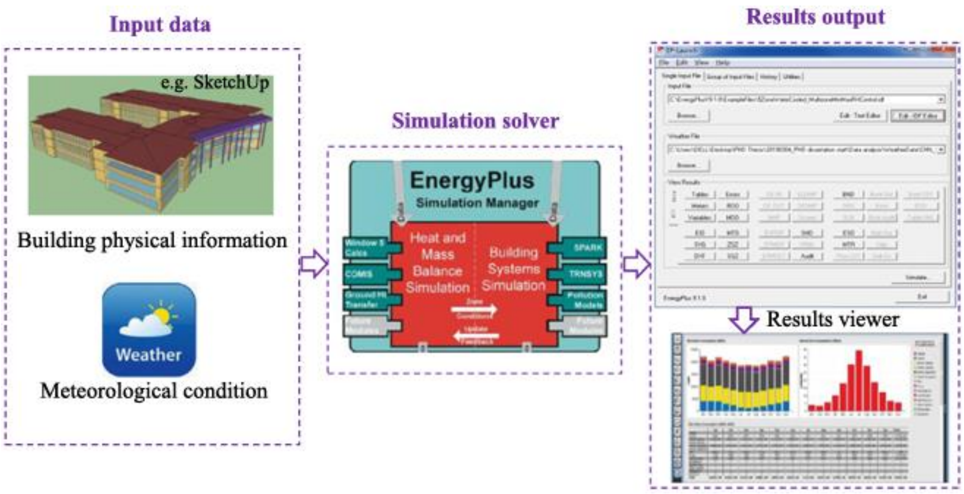

Fig. 2. Example of framework for EnergyPlus

There are examples of ideas from previous study for white box approach. From the Table 1, according to Wei *et al.,* [28], Shahcheraghian *et al.,* [29], Nie *et al.,* [30], González *et al.,* [31], and Adilkhanova *et al.,* [32], temperature is widely used for input or variable in energy consumption prediction methods, for example air temperature. Wei *et al.,* [28] introduced the new hybrid of white box which are a blend of principal component analysis (PCA), multiple linear regression (MLR) and algorithm, weighted parallel model architecture (WPMA) which named is PCA-WPMA-MLR for forecasting energy consumption. All the white box methods use similar variables of input which are time, occupants, lighting etc. There is a few white box software that can be used for example EnergyPlus, TRNSYS, CitySim and IDA-ICE based on analysis by Shahcheraghian *et al.,* [29]. Instead of using EnergyPlus, TRNSYS etc, white box can be developed by using Python according to Nie *et al.,* [30] and SOLENE-Microclimat according to Adilkhanova *et al.,* [32]. It was suggested by González *et al.,* [31] that two index kinds be used for the white box model assessment, for example the root mean square error's coefficient of variation, also called as CV(RMSE), comes initially. The Root Mean Square Error (RMSE) is weighted by the real data mean to obtain the CV(RMSE).

Table 1
List of previous study for white box approach

Author	Application	Input	Paper approach
Wei *et al.,* [28]	Provide interpretable and accurate results for daily natural gas consumption forecasting from high dimensional and large samples.	• Time • Occupants • Lighting • Occupants heating • Total internal load	• PCA • MLR • WPMA • PMA • Hybrid white box • PCA-WPMA-MLR
Shahcheraghian *et al.,* [29]	Simulation tools were categorized for building energy management into two primary classes: white and black-box models.	• Building envelope characteristics • HVAC system configurations • Internal heat contributions • Equipment specifications • Occupancy patterns • Thermal zones • Geographical location • Meteorological data	Whitebox software • EnergyPlus • TRNSYS • CitySim • IDA-ICE
Nie *et al.,* [30]	Co-simulation and white box Modelling based that allows architects to obtain recommended thermal values from corresponding climatic parameters during early design stage to swiftly evaluate the energy saving potential.	• Thermal zone • Building characteristic • Air temperature	• EnergyPlus • Python
González *et al.,* [31]	Simulation of the base model with all the proposed weather file combinations. This was an attempt to discover which climate file, by simulating it with a baseline model, best fit reality	• Weather • Temperature • Floor area	White box- quality assessment of models based on the • Coefficient of Variation of the Root Mean Square Error (CV(RMSE)) • Square Pearson Correlation Coefficient
Adilkhanova *et al.,* [32]	Existing literature was examined using the instruments for UHI estimation, thoroughly evaluates the widely used physical method and data driven based methods applied for UHI predictions and explores the viability of combining both approaches during the urban area's development and operation levels for efficient UHI (urban heat island effect) forecasting and prevention.	• Air temperature • Humidity • Wind velocity. • Surface temperature	Simulation tool • SOLENE-Microclimat

2.2 Black-box

Black box techniques are widely used to predict and forecast energy usage that investigated by Kusiak *et al.,* [33], Dong *et al.,* [34], Yang and Shen [35], and Ahmad *et al.,* [36]. Black-box models are developed for who may not have detailed knowledge of the internal workings of the system but have access to historical data. These models are commonly applied in commercial settings where only input and output data are available, such as in scenarios where detailed system data is proprietary or not accessible. Black-box models utilize statistical and machine learning techniques to predict system behaviour based on historical data. They are used when a quick and reasonably accurate model is needed, and detailed system information is unavailable. The simplicity and adaptability of black-box models make them ideal for real-time applications. The development process includes collecting historical input and output data, preprocessing the data to handle missing values and noise, selecting suitable machine learning algorithms (such as neural networks or regression models), training the model, validating it against testing data, and deploying it for real-time predictions. An example is a neural network model that predicts building energy consumption based on past consumption data and weather conditions. The term "black box" or "data driven" is also used and have been the subject of studies by Li *et al.,* [37], Tardioli *et al.,* [38], and Chen *et al.,* [39]. Promising black box techniques are presented in this part, such as support vector machines (SVM), linear regression (LR) that has been analysed by Bourdeau *et al.,* [40] and artificial neural networks (ANN) by Li and Wen [41]. Unfortunately, because to the great dimensionality of the material, black box approaches suffer from inefficient technique prediction has been analysed by Amara *et al.,* [42] and was focused on energy usage were Zhao and Magoulès [43] and Zhang *et al.,* [44], thermal comfort in the interior was studied by Kiprijanovska *et al.,* [45], electricity utilisation investigated by Ciulla and D'Amico [46], heat and electricity use was mention by Chae *et al.,* [47], and occupant and usage of energy by Li *et al.,* [48] have all been the subject of research on data-driven techniques.

There are examples of the ideas from previous study for black box approach. From the Table 2, according to the Kusiakand Li [49] and Kusiak *et al.,* [50] in their previous and recent paper, this method determines how data driven model which is the Multilayer perceptron (MLP) has been seen as a better fit for HVAC system parts. For the analysing data from Tomažič *et al.,* [51], four distinct forecasting models were built using the techniques of linear regression, k-nearest neighbours (k-NN), fuzzy, and evolving. The fuzzy approach was determined to be most precise when considering the mistakes in predicting the amount of electricity use since the coefficient of determination has the greatest value and the root mean square error has the smallest value. According to Kontokosta and Tull [52], the highest precise forecast for estimating the consumption of energy for the collection of Local Law 84 energy disclosure data (LL84) buildings is provided by SVM. According to Kontokosta and Tull [52], Zhao and Magoulès [53], and Tso *et al.,* [54] that to estimate the energy consumption of buildings throughout an urban area, there are three machine learning methods are utilised: Random Forest (RF), Ordinary least squares (OLS) Regression and Support Vector Machines (SVM) and every of these theories may explain various levels of nonlinearity and has certain advantages and disadvantages.

According to the Li *et al.,* [55], the ANN has been used to forecast several kinds of consumption of energy for buildings and on wider scales within the last 20 years. For Regression Model is the most fundamental type of forecasting time series methods is regression modelling, which is typically predicated on the notion of using the variation process to change the time series into one that is steady, and it is used for larger scale. For Fuzzy Model is one of the techniques that combines traditional time series models with the concept of fuzzy sets for predicting problems. Soon after, regression issues were included to this technique. From the Oliveira-Lima *et al.,* [56], the input of parking lot occupancy as a variable for energy consumption prediction instead of time, lighting etc.

Table 2
List of previous study for black box approach

Author	Application	Input	Paper approach
Kusiak *et al.,* [50]	Minimization of energy consumption of an HVAC system while maintaining the thermal comfort of a building with uncertain occupancy level.	• Time • Occupants • Lighting • Occupants heating • Total internal load	• Multilayer perceptron (MLP)
Tomažič *et al.,* [51]	The production of a certain type of steel is possible in shorter time compared to the existing process.	• Total scrap weight • Total carbon • Mean temperature • Total oxygen • Chemical energy	• K-NN method • Linear regression • evolving method • fuzzy method
Kontokosta and Tull [52]	Evaluation of several prediction algorithms, which includes ordinary least squares regression, support vector regression, and random forest, and two feature selection methods for building-specific annual energy use prediction and energy use intensity (EUI) from existing property and land use administrative records.	• Log building area. • Floor area ratio • Year build	• Regression model • OLS Regression • Random Forest (RF). • ANN • Support Vector Machines (SVM)
Li *et al.,* [55]	Detail analysis based on evolutionary algorithms hybridized models that combine evolutionary algorithms with common black box models to enhance estimation precision and robustness.	• Historical energy data • Meteorology • Calendar • Occupancy	• ANN (Artificial neural networks) • SVM (Support vector machines) • Fuzzy model • Regression model
Oliveira-Lima *et al.,* [56]	Estimation of occupancy profiles assisted by alternative data sources, improving the accuracy and feasibility of the building occupancy profile estimation and energy consumption	• Electrical load • Parking lot occupancy • occupancy profile	• ANN (Artificial neural networks)

2.3 Grey Box

One of the biggest challenges facing building energy users is the design and testing process of "white box" software. There are a lot of fundamental input factors needed, which makes it time-consuming to construct the modelling on a physical software framework and expensive to run the model. In a manner, the data driven models incorporate both linear and nonlinear significance among the input and output parameters. But to develop these algorithms and produce precise forecasts in a variety of scenarios, a substantial amount of historical information must be gathered over an extended period. A grey box solution has been suggested to address this conundrum. The benefits associated with the physical and hybrid models are combined, as it simulates building energy consumption using a reduced building model and easily accessed information. Grey-box models are created by researchers who possess partial knowledge of the system and have access to some historical data according to Dong *et al.,* [57]. These models are used in semi-controlled environments where some system details are known, and data is available for calibration. Grey-box models combine the detailed physical equations of white-box models with the empirical data-driven approach of black-box models, offering a balance between complexity and practicality. They are applied when there is sufficient understanding of the system but not enough to develop a complete white-box model. Grey-box models provide a compromise between the accuracy of white-box models and the simplicity of black-box models, making them useful for applications where both accuracy and simplicity are important. Developing a grey-box model involves defining the system, developing simplified physical equations, collecting physical and historical data, calibrating the model with historical data, integrating physical equations with data-driven techniques, and validating the model to ensure accuracy.

There are examples of ideas from previous study for grey box approach. Table 3 shows, according to the Dong *et al.,* [57], analysis that compares between 5 of data driven model and hybrid model which are artificial neural networks (ANN), Gaussian mixture models (GMM), Support Vector Regression (SVR), Least squares support vector machines (LS-SVM) and Gaussian process regression (GPR) and the conclusion of the analysis revealed that the grey box modelling technique performs somewhat better for predicting one hour in advance. All the outcomes for the 24-hour predictions are comparable by using the same parameter which are internal heat gain, solar radiation, and outside air temperature. According to Green and Garimella [58], result analysis the similar findings were obtained by the genetic algorithm utilizing the hybrid model and black box models, with reductions in relations of both energy and funds. Based on the author when solve household energy modelling challenges, Extreme gradient boosting (XGB) has been shown to yield better outcomes than well-known machine learning methods like SVM and ANN.

Harb *et al.,* [59] state that the hybrid or grey models are frequently shown as xRyC networks with lumped variables that are analogous to an electrical circuit, where y represents the thermal capacitance and x represents as represents the number of thermal resistances. Based on their analysis 3R2C and 4R2C are selected. The construction model's mathematical equations are developed in accordance with the selected circuit. Based on the Shamsi *et al.,* [60], the changing conditions of the building are represented by a conventional Resistance Capacitance (RC) network, which assumes a steady-state heat transmission throughout the building environment. Grey box or hybrid method which is thermal equivalent circuit (TEC) technique, which treats the temperatures and heat flows of a temperature system as the currents and voltages of an electrical circuit, is used to develop the freezer designs based on the Sossan *et al.,* [61] analysis. Figure 3 shows the example of the 3R2C design according to Harb *et al.,* [59].

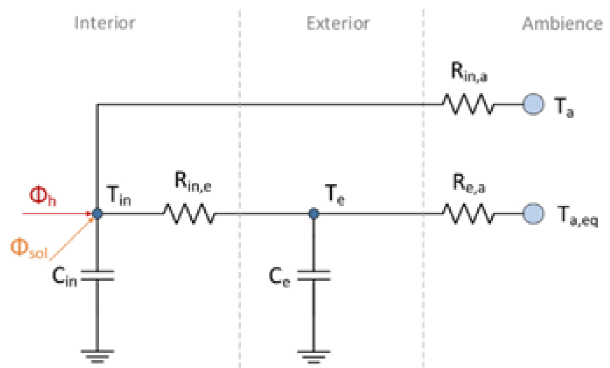

Fig. 3. Framework of the 3R2C design

Table 3
List of previous study for grey box approach

Author	Application	Input	Paper approach
Dong *et al.*, [57]	A novel hybrid forecasting technique was created to anticipate home electrical consumption one hour in advance and one day in advance by combining physical model and black box model.	• internal heat gain, • solar radiation • outdoor air temperature.	• Hybrid • ANN • SVR • GPR • GMM
Green and Garimella [58]	Extreme Gradient Boosting (XGB) was used to develop data driven models of power draw and subsequent temperatures of the water heater and HVAC system.	• Temperature • HVAC System	• XGB
Harb *et al.*, [59]	An approach to simulate the thermal behaviour of buildings based on collected measurement data using grey-box models.	• Indoor air temperature (average for the whole building) • Building heat consumption • Outdoor air temperature • Global solar radiation on horizontal surface.	• 3R2C • 4R2C
Shamsi *et al.*, [60]	A comprehensive approach for testing and assessing the scalability, adaptability, and connectivity of the framework for reduced-order hybrid models.	• Weather • Building Site Information • Building Physical Parameters • HVAC System	• Reduced-order grey box models
Sossan *et al.*, [61]	Grey box modelling was used to recognize suitable power consumption to temperature models of domestic refrigeration by experimental measurements from instrumented freezer.	• Global heat capacity • Coefficient	• Thermal equivalent circuit (TEC)

2.4 Formula for Accuracy Matric

Several important criteria are frequently used to evaluate the accuracy of building energy consumption models, regardless of whether they are black-box, grey-box, or white-box models. These measures aid in assessing how closely the model's predictions correspond with the real energy usage information. Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2) are the most used accuracy metrics. Based on Figure 4, for accuracy, each of the black box and grey box model will be compared using in Eq. (1) Mean Absolute Error (MAE) to display the typical variance between the expected and actual values. Then, Mean Absolute Percentage Error (MAPE) in Eq. (2) is used to convey accuracy rate. While, Root Mean Square Error (RMSE) in Eq. (3) on the other hand is used to examine the accuracy of different expectations standards. The effect of absolute errors, Coefficient of Variation of RMSE (CVRMSE) in Eq. (4) which is to defines expectation error and produces a measure without units and Determination Coefficient (R^2) in Eq. (5) which is to evaluates how well a model replicates the actual data according to Bourdeau *et al.,* [12]. For Eq. (1) to Eq. (5) are the example of each accuracy metric.

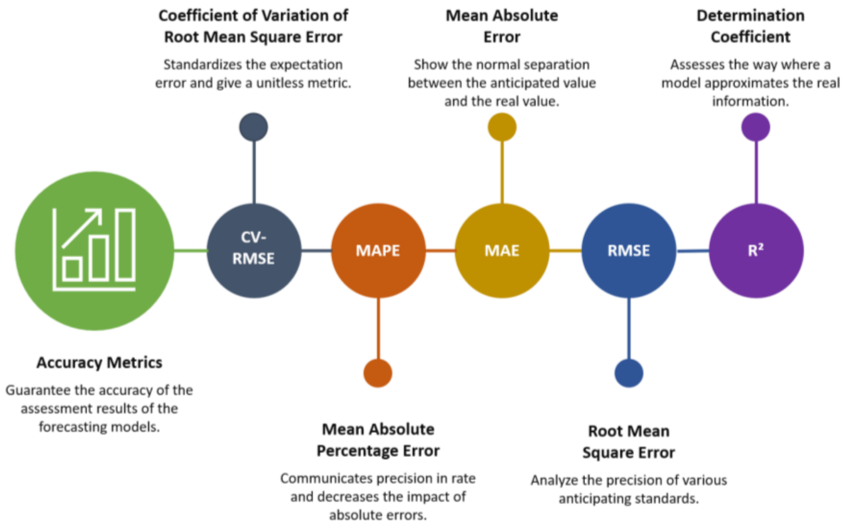

Fig. 4. Example of the accuracy metric for forecasting model

Mean Absolute Error (MAE):

Without considering the direction of the mistakes, MAE calculates the average magnitude of the errors between the actual and anticipated values. The meaning of the absolute variations between the expected and actual values is used to compute it according to Sukarti *et al.,* [62].

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_{forecasting\ i} - y_{observed,i}\right| \tag{1}$$

where the $y_{forecasting,i}$ is estimated energy usage at time point i , $y_{observed,i}$ is the actual energy usage data at time point i ,$\bar{y}_{observed,i}$ is the average of the actual data energy usage over the respected timeframe, and n is sum of the data in dataset respected for performance assessment according to Bourdeau *et al.,* [12]. For the Eq. (2) to Eq. (5) are the expanding formula from the MAE equation.

Mean Absolute Percentage Error (MAPE):

The average of the absolute percentage errors between the expected and actual values is used to determine the accuracy of predictions expressed as a percentage. When comparing the outcomes of different models or datasets, this statistic is useful because it provides an intelligible percentage error. Because it gives insight into the average relative error size for each approach, it is frequently used to compare various methods. Unfortunately, one of its drawbacks is that the % inaccuracy grows for low values of the real variable, making it less than perfect as an indicator according to Nęcka *et al.,* [63].

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_{forecasting,i} - y_{observed,i}}{\bar{y}_{observed,i}}\right| * 100 \qquad (2)$$

Root Mean Squared Error (RMSE):

This error metric has the same units as the target variable and is calculated as the square root of MSE. RSME is an indicator that contrasts the absolute amount of the variation between the measured and projected values according to Moon *et al.,* [64].

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{forecasting\ i} - y_{observed,i}\right)^{2}} \qquad (3)$$

Coefficient of Variation of RMSE (CVRMSE):

These are specific metrics often used in building energy modelling to account for systematic biases and the variability of errors relative to the mean of the observed data. It calculates the difference between expected and actual energy use. The root mean square error is divided by the average energy usage to compute the CVRMSE, a calculation that was developed in 2014 by the American Society of Heating, Refrigerating, and Air-Conditioning Engineers. As a benchmark for assessing the precision of model predictions, this parameter is essential to both ASHRAE Guideline 14 and the International Performance Measurement and Verification Protocol (IPMVP) according to Sukarti *et al.,* [62].

$$CVRMSE = \sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}\left(y_{forecasting\ i} - y_{observed,i}\right)^{2}}{\bar{y}_{observed,i}}} * 100 \qquad (4)$$

Coefficient of Determination (R^2):

R^2 measures the proportion of the variance in the dependent variable that is predictable from the independent variables. crucial metric for comprehending a model's goodness-of-fit. A high R2 value is preferred in M&V energy modelling since it shows that the model accounts for a sizable percentage of the variable in energy use, making it a reliable tool for analysis and prediction under ASHRAE and IPMVP standards according to Sukarti *et al.,* [62].

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_{forecasting\ i} - y_{observed,i}\right)^2}{\sum_{i=1}^{n} \left(y_{forecasting\ i} - \bar{y}_{observed,i}\right)^2} \tag{5}$$

By employing these criteria, researchers can assess and contrast the efficacy of various building energy consumption models, guaranteeing that the selected model yields dependable and precise forecasts. Precise models are essential for maximising energy use, cutting expenses, and raising buildings' total energy efficiency.

3. Results

Table 4 describes the advantages and disadvantages of the three approaches, which are the white box, black box, and grey box. The advantages of the white box are that the variables are simple to alter, and no prior information is needed. And white box disadvantages are that inserting building variables and data takes a great deal of work, there are significant expenses for computation, and prior experience in thermal dynamics or simulation is necessary, based on An *et al.,* [65]. White Box can be used to predict the building envelope characteristics, HVAC system configurations, internal heat contributions, and equipment specifications by using the software, for example, EnergyPlus, TRNSYS, and Python, according to Shahcheraghian *et al.,* [29] and Nie *et al.,* [30]. Then, for the black box approach, no specialised knowledge is needed, the time required for calculation and modelling is minimal, and generalising the created model is simple, but a significant amount of past information is needed. Overfit is simple, and most of the time it cannot be explained, according to An *et al.,* [65]. It can be used for forecasting total scrap weight, total carbon, mean temperature, total oxygen, and chemical energy by using, for example, regression models such as OLS regression, random forest (RF), artificial neural networks, and support vector machines (SVM), according to Tomažič *et al.,* [51] and Kontokosta and Tull [52]. And for grey box approach is this approach less information is needed. It is necessary to have just limitations on physical variables, but it is difficult to build because it combines two different fields of science, which are the white box and the black box, based on An *et al.,* [65]. It can be used to predict indoor air temperature (average for the whole building), building heat consumption, outdoor air temperature, and global solar radiation on a horizontal surface by using, for example, XGB, 3R2C, and 4R2C, according to Green and Garimella [58] and Harb *et al.,* [59].

Table 4
Advantages, disadvantages, application and type of prediction method of different prediction approaches

Model	Advantage	Disadvantage	Application	Type
White box model	• The variables are simple to alter. • No prior information is needed.	• Inserting building variables and data takes a great deal of work. • Significant expenses for computation. • Prior experience in thermal dynamics or simulation is necessary.	• Building envelope characteristics • HVAC system configurations • Internal heat contributions • Equipment specifications • Occupancy patterns • Thermal zones • Geographical Location • Meteorological data	• EnergyPlus • TRNSYS • CitySim • IDA-ICE • Phyton
Black box	• No specialized knowledge is needed. • The time required for calculation and modelling is minimal. • Generalising the created model is simple.	• A significant amount of past information is needed. • Overfitting is simple, and most of the time it cannot be explained	• Total scrap weight • Total carbon • Mean temperature • Total oxygen • Chemical energy	• Regression model • OLS Regression • Random Forest (RF) • Artificial Neural Network • Support Vector Machines (SVM)
Grey box	• Less information is needed. • It is necessary to have just limitations on physical variables.	• It combines two different fields of science. • It is difficult to build.	• Indoor air temperature (average for the whole building) • Building heat consumption • Outdoor air temperature • Global solar radiation on horizontal surface.	• XGB • 3R2C • 4R2C

4. Conclusions

Optimising energy use, cost reduction and improving energy efficiency all depend on how well-versed the system and available data one has. Depending on this, three different modelling approaches as summarized in Figure 5; white-box, black-box and grey-box offer varying benefits as presented in this paper. White-box models employ detailed physical equations to achieve high accuracy, but they require extensive system data. Black-box models, on the other hand, rely on historical data and machine learning techniques to provide detailed accuracy at a lower implementation cost. Grey-box models combine elements of both, using simplified physical equations supplemented by empirical data, providing a balance between accuracy and simplicity. To evaluate the performance of these models, several accuracy metrics are used, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Coefficient of Determination (R^2), Normalized Mean Bias Error (NMBE), Coefficient of Variation of RMSE (CVRMSE) and Mean Absolute Percentage Error (MAPE). These indicators help in measuring how well the models' predictions correlate with real energy usage data, directing improvements and assuring reliable predictions. Each model type and accuracy metric play an important role in the field of building energy modelling, helping initiatives to make buildings more energy-efficient and sustainable. By selecting the proper modelling approach and evaluating its effectiveness using these criteria, researchers and practitioners may successfully address energy consumption concerns in buildings.

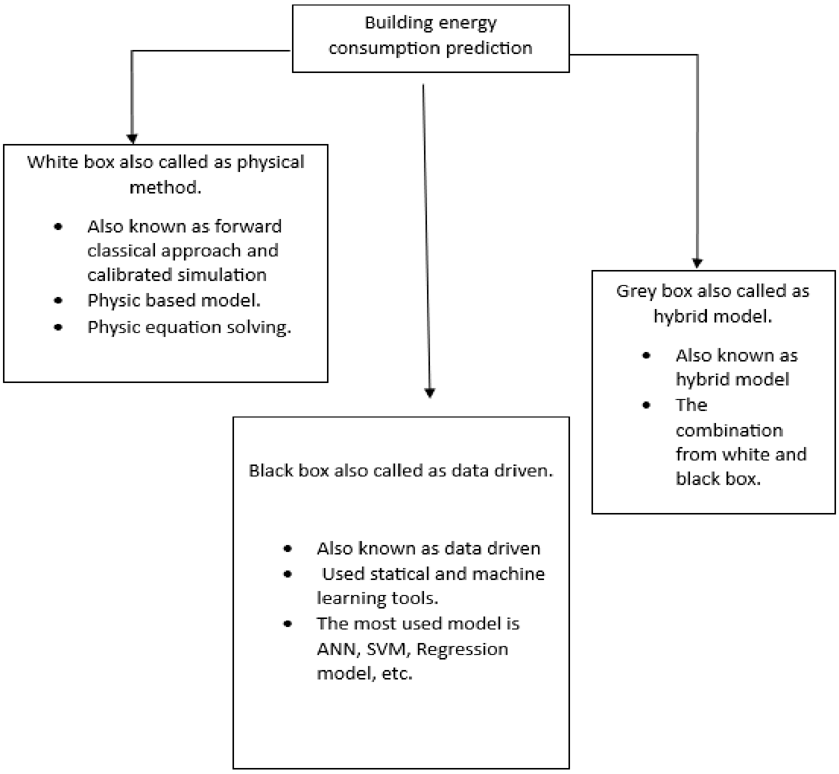

Fig. 5. Summary of the three methods of energy forecasting modelling

Acknowledgement

This research was supported by the Faculty of Electrical Technology and Engineering, Universiti Teknikal Malaysia Melaka [Research Publication Fund].

Reference

[1] Khan, Samee Ullah, Noman Khan, Fath U. Min Ullah, Min Je Kim, Mi Young Lee, and Sung Wook Baik. "Towards intelligent building energy management: AI-based framework for power consumption and generation forecasting." *Energy and Buildings* 279 (2023): 112705. https://doi.org/10.1016/j.enbuild.2022.112705

[2] Kunalan, Kerishmaa Theavy, Cheng Yee Ng, and Nauman Riyaz Maldar. "A performance investigation of a multi-staging hydrokinetic turbine for river flow." *Progress in Energy and Environment* 17 (2021): 17-31. https://doi.org/10.37934/progee.17.1.1731

[3] Norouziasl, Seddigheh, and Amirhosein Jafari. "Identifying the most influential parameters in predicting lighting energy consumption in office buildings using data-driven method." *Journal of Building Engineering* 72 (2023): 106590. https://doi.org/10.1016/j.jobe.2023.106590

[4] Vijayan, Priya. "Energy consumption prediction in low energy buildings using machine learning and artificial intelligence for energy efficiency." In *2022 8th International Youth Conference on Energy (IYCE)*, pp. 1-6. IEEE, 2022. https://doi.org/10.1109/IYCE54153.2022.9857548

[5] Mohapatra, Sunil Kumar, Sushruta Mishra, Hrudaya Kumar Tripathy, and Ahmed Alkhayyat. "A sustainable data-driven energy consumption assessment model for building infrastructures in resource constraint environment." *Sustainable Energy Technologies and Assessments* 53 (2022): 102697. https://doi.org/10.1016/j.seta.2022.102697

[6] Daut, Mohammad Azhar Mat, Mohammad Yusri Hassan, Hayati Abdullah, Hasimah Abdul Rahman, Md Pauzi Abdullah, and Faridah Hussin. "Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review." *Renewable and Sustainable Energy Reviews* 70 (2017): 1108-1118. https://doi.org/10.1016/j.rser.2016.12.015

[7] Chalal, Moulay Larbi, Medjdoub Benachir, Michael White, and Raid Shrahily. "Energy planning and forecasting approaches for supporting physical improvement strategies in the building sector: A review." *Renewable and Sustainable Energy Reviews* 64 (2016): 761-776. https://doi.org/10.1016/j.rser.2016.06.040

[8] Wang, Zhe, Tianzhen Hong, and Mary Ann Piette. "Building thermal load prediction through shallow machine learning and deep learning." *Applied Energy* 263 (2020): 114683. https://doi.org/10.1016/j.apenergy.2020.114683

[9] Zhao, Hai-xiang, and Frédéric Magoulès. "A review on the prediction of building energy consumption." *Renewable and Sustainable Energy Reviews* 16, no. 6 (2012): 3586-3592. https://doi.org/10.1016/j.rser.2012.02.049

[10] Wang, Zeyu, and Ravi S. Srinivasan. "A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models." *Renewable and Sustainable Energy Reviews* 75 (2017): 796-808. https://doi.org/10.1016/j.rser.2016.10.079

[11] Foucquier, Aurélie, Sylvain Robert, Frédéric Suard, Louis Stéphan, and Arnaud Jay. "State of the art in building modelling and energy performances prediction: A review." *Renewable and Sustainable Energy Reviews* 23 (2013): 272-288. https://doi.org/10.1016/j.rser.2013.03.004

[12] Bourdeau, Mathieu, Xiao qiang Zhai, Elyes Nefzaoui, Xiaofeng Guo, and Patrice Chatellier. "Modeling and forecasting building energy consumption: A review of data-driven techniques." *Sustainable Cities and Society* 48 (2019): 101533. https://doi.org/10.1016/j.scs.2019.101533

[13] Lü, Xiaoshu, Tao Lu, Charles J. Kibert, and Martti Viljanen. "Modeling and forecasting energy consumption for heterogeneous buildings using a physical-statistical approach." *Applied Energy* 144 (2015): 261-275. https://doi.org/10.1016/j.apenergy.2014.12.019

[14] Chen, Yongbao, Mingyue Guo, Zhisen Chen, Zhe Chen, and Ying Ji. "Physical energy and data-driven models in building energy prediction: A review." *Energy Reports* 8 (2022): 2656-2671. https://doi.org/10.1016/j.egyr.2022.01.162

[15] Harish, V. S. K. V., and Arun Kumar. "A review on modeling and simulation of building energy systems." *Renewable and Sustainable Energy Reviews* 56 (2016): 1272-1292. https://doi.org/10.1016/j.rser.2015.12.040

[16] Imam, Salah, David A. Coley, and Ian Walker. "The building performance gap: Are modellers literate?." *Building Services Engineering Research and Technology* 38, no. 3 (2017): 351-375. https://doi.org/10.1177/0143624416684641

[17] Ha, Sungkyun, Sungho Tae, and Rakhyun Kim. "Energy demand forecast models for commercial buildings in South Korea." *Energies* 12, no. 12 (2019): 2313. https://doi.org/10.3390/en12122313

[18] Blázquez-García, Ane, Angel Conde, Aitor Milo, Roberto Sánchez, and Irantzu Barrio. "Short-term office building elevator energy consumption forecast using SARIMA." *Journal of Building Performance Simulation* 13, no. 1 (2020): 69-78. https://doi.org/10.1080/19401493.2019.1698657

[19] Oh, Seung Jin, Kim Choon Ng, Kyaw Thu, Wongee Chun, and Kian Jon Ernest Chua. "Forecasting long-term electricity demand for cooling of Singapore's buildings incorporating an innovative air-conditioning technology." *Energy and Buildings* 127 (2016): 183-193. https://doi.org/10.1016/j.enbuild.2016.05.073

[20] Li, Xiwang, Jin Wen, Ran Liu, and Xiaohui Zhou. "Commercial building cooling energy forecasting using proactive system identification: A whole building experiment study." *Science and Technology for the Built Environment* 22, no. 6 (2016): 674-691. https://doi.org/10.1080/23744731.2016.1188654

[21] Yu, Dongwei. "A two-step approach to forecasting city-wide building energy demand." *Energy and Buildings* 160 (2018): 1-9. https://doi.org/10.1016/j.enbuild.2017.11.063

[22] Kang, Minkyung, Mario Bergés, and Burcu Akinci. "Forecasting Airport Building Electricity Demand on the Basis of Flight Schedule Information for Demand Response Applications." *Transportation Research Record* 2603, no. 1 (2017): 29-38. https://doi.org/10.3141/2603-04

[23] Ghedamsi, Rébha, Noureddine Settou, Abderrahmane Gouareh, Adem Khamouli, Nadia Saifi, Bakhta Recioui, and Boubekker Dokkar. "Modeling and forecasting energy consumption for residential buildings in Algeria using bottom-up approach." *Energy and Buildings* 121 (2016): 309-317. https://doi.org/10.1016/j.enbuild.2015.12.030

[24] Lee, Dongho. "Low-cost and simple short-term load forecasting for energy management systems in small and middle-sized office buildings." *Energy Exploration & Exploitation* 39, no. 2 (2021): 637-656. https://doi.org/10.1177/0144598719900964

[25] Lindberg, Karen Byskov, Steffen J. Bakker, and Igor Sartori. "Modelling electric and heat load profiles of non-residential buildings for use in long-term aggregate load forecasts." *Utilities Policy* 58 (2019): 63-88. https://doi.org/10.1016/j.jup.2019.03.004

[26] Szul, Tomasz, and Stanisław Kokoszka. "Application of Rough Set Theory (RST) to forecast energy consumption in buildings undergoing thermal modernization." *Energies* 13, no. 6 (2020): 1309. https://doi.org/10.3390/en13061309

[27] Zhang, Xiang, Christoffer Rasmussen, Dirk Saelens, and Staf Roels. "Time-dependent solar aperture estimation of a building: Comparing grey-box and white-box approaches." *Renewable and Sustainable Energy Reviews* 161 (2022): 112337. https://doi.org/10.1016/j.rser.2022.112337
[28] Wei, Nan, Lihua Yin, Chao Li, Changjun Li, Christine Chan, and Fanhua Zeng. "Forecasting the daily natural gas consumption with an accurate white-box model." *Energy* 232 (2021): 121036. https://doi.org/10.1016/j.energy.2021.121036
[29] Shahcheraghian, Amir, Hatef Madani, and Adrian Ilinca. "From white to black-box models: A review of simulation tools for building energy management and their application in consulting practices." *Energies* 17, no. 2 (2024): 376. https://doi.org/10.3390/en17020376
[30] Nie, Zichuan, Sisi Chen, Shen Zhang, Haitang Wu, Tobias Weiss, and Lihua Zhao. "Adaptive Façades Strategy: An architect-friendly computational approach based on co-simulation and white-box models for the early design stage." *Energy and Buildings* 296 (2023): 113320. https://doi.org/10.1016/j.enbuild.2023.113320
[31] González, Vicente Gutiérrez, Germán Ramos Ruiz, and Carlos Fernández Bandera. "Impact of actual weather datasets for calibrating white-box building energy models base on monitored data." *Energies* 14, no. 4 (2021): 1187. https://doi.org/10.3390/en14041187
[32] Adilkhanova, Indira, Jack Ngarambe, and Geun Young Yun. "Recent advances in black box and white-box models for urban heat island prediction: Implications of fusing the two methods." *Renewable and Sustainable Energy Reviews* 165 (2022): 112520. https://doi.org/10.1016/j.rser.2022.112520
[33] Kusiak, Andrew, Guanglin Xu, and Zijun Zhang. "Minimization of energy consumption in HVAC systems with data-driven models and an interior-point method." *Energy Conversion and Management* 85 (2014): 146-153. https://doi.org/10.1016/j.enconman.2014.05.053
[34] Dong, Bing, Cheng Cao, and Siew Eang Lee. "Applying support vector machines to predict building energy consumption in tropical region." *Energy and Buildings* 37, no. 5 (2005): 545-553. https://doi.org/10.1016/j.enbuild.2004.09.009
[35] Yang, Shan-lin, and Chao Shen. "A review of electric load classification in smart grid environment." *Renewable and Sustainable Energy Reviews* 24 (2013): 103-110. https://doi.org/10.1016/j.rser.2013.03.023
[36] Ahmad, Ahmad S., Mohammad Y. Hassan, Md Pauzi Abdullah, Hasimah A. Rahman, F. Hussin, Hayati Abdullah, and Rahman Saidur. "A review on applications of ANN and SVM for building electrical energy consumption forecasting." *Renewable and Sustainable Energy Reviews* 33 (2014): 102-109. https://doi.org/10.1016/j.rser.2014.01.069
[37] Li, Yanfei, Zheng O'Neill, Liang Zhang, Jianli Chen, Piljae Im, and Jason DeGraw. "Grey-box modeling and application for building energy simulations-A critical review." *Renewable and Sustainable Energy Reviews* 146 (2021): 111174. https://doi.org/10.1016/j.rser.2021.111174
[38] Tardioli, Giovanni, Ruth Kerrigan, Mike Oates, O'Donnell James, and Donal Finn. "Data driven approaches for prediction of building energy consumption at urban level." *Energy Procedia* 78 (2015): 3378-3383. https://doi.org/10.1016/j.egypro.2015.11.754
[39] Chen, Yongbao, Mingyue Guo, Zhisen Chen, Zhe Chen, and Ying Ji. "Physical energy and data-driven models in building energy prediction: A review." *Energy Reports* 8 (2022): 2656-2671. https://doi.org/10.1016/j.egyr.2022.01.162
[40] Bourdeau, Mathieu, Xiao qiang Zhai, Elyes Nefzaoui, Xiaofeng Guo, and Patrice Chatellier. "Modeling and forecasting building energy consumption: A review of data-driven techniques." *Sustainable Cities and Society* 48 (2019): 101533. https://doi.org/10.1016/j.scs.2019.101533
[41] Li, Xiwang, and Jin Wen. "Review of building energy modeling for control and operation." *Renewable and Sustainable Energy Reviews* 37 (2014): 517-537. https://doi.org/10.1016/j.rser.2014.05.056
[42] Amara, Fatima, Kodjo Agbossou, Alben Cardenas, Yves Dubé, and Sousso Kelouwani. "Comparison and simulation of building thermal models for effective energy management." *Smart Grid and Renewable Energy* 6, no. 04 (2015): 95. https://doi.org/10.4236/sgre.2015.64009
[43] Zhao, Hai-Xiang, and Frédéric Magoulès. "Feature selection for predicting building energy consumption based on statistical learning method." *Journal of Algorithms & Computational Technology* 6, no. 1 (2012): 59-77. https://doi.org/10.1260/1748-3018.6.1.59
[44] Zhang, Fan, Chirag Deb, Siew Eang Lee, Junjing Yang, and Kwok Wei Shah. "Time series forecasting for building energy consumption using weighted Support Vector Regression with differential evolution optimization technique." *Energy and Buildings* 126 (2016): 94-103. https://doi.org/10.1016/j.enbuild.2016.05.028
[45] Kiprijanovska, Ivana, Simon Stankoski, Igor Ilievski, Slobodan Jovanovski, Matjaž Gams, and Hristijan Gjoreski. "Houseec: Day-ahead household electrical energy consumption forecasting using deep learning." *Energies* 13, no. 10 (2020): 2672. https://doi.org/10.3390/en13102672
[46] Ciulla, G., and A. D'Amico. "Building energy performance forecasting: A multiple linear regression approach." *Applied Energy* 253 (2019): 113500. https://doi.org/10.1016/j.apenergy.2019.113500

[47] Chae, Young Tae, Raya Horesh, Youngdeok Hwang, and Young M. Lee. "Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings." *Energy and Buildings* 111 (2016): 184-194. https://doi.org/10.1016/j.enbuild.2015.11.045
[48] Li, Zihao, Daniel Friedrich, and Gareth P. Harrison. "Demand forecasting for a mixed-use building using agent-schedule information with a data-driven model." *Energies* 13, no. 4 (2020): 780. https://doi.org/10.3390/en13040780
[49] Kusiak, Andrew, and Mingyang Li. "Cooling output optimization of an air handling unit." *Applied Energy* 87, no. 3 (2010): 901-909. https://doi.org/10.1016/j.apenergy.2009.06.010
[50] Kusiak, Andrew, Mingyang Li, and Haiyang Zheng. "Virtual models of indoor-air-quality sensors." *Applied Energy* 87, no. 6 (2010): 2087-2094. https://doi.org/10.1016/j.apenergy.2009.12.008
[51] Tomažič, Simon, Goran Andonovski, Igor Škrjanc, and Vito Logar. "Data-driven modelling and optimization of energy consumption in EAF." *Metals* 12, no. 5 (2022): 816. https://doi.org/10.3390/met12050816
[52] Kontokosta, Constantine E., and Christopher Tull. "A data-driven predictive model of city-scale energy use in buildings." *Applied Energy* 197 (2017): 303-317. https://doi.org/10.1016/j.apenergy.2017.04.005
[53] Zhao, Hai-xiang, and Frédéric Magoulès. "A review on the prediction of building energy consumption." *Renewable and Sustainable Energy Reviews* 16, no. 6 (2012): 3586-3592. https://doi.org/10.1016/j.rser.2012.02.049
[54] Tso, Geoffrey K. F., and Kelvin K. W. Yau. "Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks." *Energy* 32, no. 9 (2007): 1761-1768. https://doi.org/10.1016/j.energy.2006.11.010
[55] Li, Kangji, Wenping Xue, Gang Tan, and Anthony S. Denzer. "A state of the art review on the prediction of building energy consumption using data-driven technique and evolutionary algorithms." *Building Services Engineering Research and Technology* 41, no. 1 (2020): 108-127. https://doi.org/10.1177/0143624419843647
[56] Oliveira-Lima, Jose A., Ramiro Morais, J. F. Martins, Adrian Florea, and Celson Lima. "Load forecast on intelligent buildings based on temporary occupancy monitoring." *Energy and Buildings* 116 (2016): 512-521. https://doi.org/10.1016/j.enbuild.2016.01.028
[57] Dong, Bing, Zhaoxuan Li, S. M. Mahbobur Rahman, and Rolando Vega. "A hybrid model approach for forecasting future residential electricity consumption." *Energy and Buildings* 117 (2016): 341-351. https://doi.org/10.1016/j.enbuild.2015.09.033
[58] Green, Christy, and Srinivas Garimella. "Residential microgrid optimization using grey-box and black-box modeling methods." *Energy and Buildings* 235 (2021): 110705. https://doi.org/10.1016/j.enbuild.2020.110705
[59] Harb, Hassan, Neven Boyanov, Luis Hernandez, Rita Streblow, and Dirk Müller. "Development and validation of grey-box models for forecasting the thermal response of occupied buildings." *Energy and Buildings* 117 (2016): 199-207. https://doi.org/10.1016/j.enbuild.2016.02.021
[60] Shamsi, Mohammad Haris, Usman Ali, Eleni Mangina, and James O'Donnell. "Feature assessment frameworks to evaluate reduced-order grey-box building energy models." *Applied Energy* 298 (2021): 117174. https://doi.org/10.1016/j.apenergy.2021.117174
[61] Sossan, Fabrizio, Venkatachalam Lakshmanan, Giuseppe Tommaso Costanzo, Mattia Marinelli, Philip J. Douglass, and Henrik Bindner. "Grey-box modelling of a household refrigeration unit using time series data in application to demand side management." *Sustainable Energy, Grids and Networks* 5 (2016): 1-12. https://doi.org/10.1016/j.segan.2015.10.003
[62] Sukarti, Suziee, Mohamad Fani Sulaima, Aida Fazliana Abdul Kadir, Muhamad Hafizul Shamsor, and Siaw Wei Yao. "Advancing industrial building energy measurement and verification (M&V) with deep learning: Evaluating data size and feature selection impact." *Energy and Buildings* 319 (2024): 114457. https://doi.org/10.1016/j.enbuild.2024.114457
[63] Nęcka, Krzysztof, Tomasz Szul, Joanna Piotrowska-Woroniak, and Krzysztof Pancerz. "Assessment of Municipal Waste Forecasting Methods in Poland Considering Socioeconomic Aspects." *Energies* 17, no. 14 (2024): 3524. https://doi.org/10.3390/en17143524
[64] Moon, Juhwan, Jaejoon Lee, Sangwon Lee, and Hongsik Yun. "Urban river dissolved oxygen prediction model using machine learning." *Water* 14, no. 12 (2022): 1899. https://doi.org/10.3390/w14121899
[65] An, Dawn, Nam H. Kim, and Joo-Ho Choi. "Practical options for selecting data-driven or physics-based prognostics algorithms with reviews." *Reliability Engineering & System Safety* 133 (2015): 223-236. https://doi.org/10.1016/j.ress.2014.09.014