Sustainable Cities and Society 48 (2019) 101533

Contents lists available at ScienceDirect

Sustainable Cities and Society

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

Engineering Advance

Modeling and forecasting building energy consumption: A review of data-driven techniques

Mathieu Bourdeau[a], Xiao qiang Zhai[a,*], Elyes Nefzaoui[b,c], Xiaofeng Guo[b], Patrice Chatellier[d]

[a] Institute of Refrigeration and Cryogenics, Shanghai Jiao Tong University, Shanghai 200240, PR China
[b] ESIEE Paris, Université Paris-Est, 2 Bd Blaise Pascal, 93162 Noisy-le-Grand Cedex, France
[c] Université Paris-Est, ESYCOM (EA 2552), CNAM, ESIEE Paris, UPEMLV, F-77454, Marne-la-Vallée, France
[d] Université Paris-Est, IFSTTAR, 14-20 Bd Newton, Champs-sur-Marne, Marne-la-Vallée, France

ARTICLE INFO

Keywords:
Building energy consumption
Building load forecasting
Data-driven techniques
Machine learning

ABSTRACT

Building energy consumption modeling and forecasting is essential to address buildings energy efficiency problems and take up current challenges of human comfort, urbanization growth and the consequent energy consumption increase. In a context of integrated smart infrastructures, data-driven techniques rely on data analysis and machine learning to provide flexible methods for building energy prediction. The present paper offers a review of studies developing data-driven models for building scale applications. The prevalent methods are introduced with a focus on the input data characteristics and data pre-processing methods, the building typologies considered, the targeted energy end-uses and forecasting horizons, and accuracy assessment. A special attention is also given to different machine learning approaches. Based on the results of this review, the latest technical improvements and research efforts are synthesized. The key role of occupants' behavior integration in data-driven modeling is discussed. Limitations and research gaps are highlighted. Future research opportunities are also identified.

1. Introduction

Buildings account for a significant part of the global energy consumption with 30% in average and a third of the associated CO_2 emissions (International Energy Agency, 2016). Despite developments to improve building energy efficiency, the International Energy Agency has highlighted in 2017 that current investments were not on track for building sector to achieve the 2 °C-scenario targeted by Paris Climate Agreement (International Energy Agency, 2017). Meanwhile, some of the major global warming contributors and signatories of the Agreement such as China – 2538 Mega ton oil equivalent consumed and 8796 Mega ton of CO_2 produced in 2016 ("Global Energy Statistical Yearbook, 2017|World Energy Statistics|Enerdata," 2017) – are facing challenges with a growing urbanization and an annual increase of their building stock (Tsinghua University Building Energy Research Center, 2016).

Abbreviations: AC, air conditioning; AI, artificial intelligence; ANFIS, adaptive network-based fuzzy inference system; ANN, artificial neural network; AR, autoregressive; ARMA, autoregressive moving average; ARX, autoregressive exogenous; ARIMA, autoregressive integrated moving average; ARIMAX, autoregressive integrated moving average exogenous; BECMF, building energy consumption modeling and forecasting; BDT, boosting decision tree; BPNN, back-propagation neural network; CART, classification and regression tree; CHAID, chi-square automatic interaction detector; CRBM, conditional restricted Boltzmann machine; CV-RMSE, coefficient of variation of root mean squared error; DBN, deep belief network; DE, differential evolution; DNN, deep neural network; DPT, dew point temperature; DT, decision tree; ELM, extreme learning machine; EM, expectation maximization; FCRBM, factored conditional restricted Boltzmann machine; FFNN, feed-forward neural network; GA, genetic algorithm; GP, genetic programming; GBDT, gradient boosting decision tree; HVAC, heating, ventilation and air-conditioning; IAT, indoor air temperature; *k*-NN, *k* nearest neighbors; LEED, leadership in energy and environmental design; LS-SVM, least square support vector machine; MAE, mean absolute error; MAPE, mean absolute percentage error; MARS, multivariate adaptive regression splines; MLP, multilayer perceptron; MLR, multiple linear regression; NARX, non-linear autoregressive exogenous; OAT, outdoor air temperature; OLS, ordinary least square (regression); PCA, principal component analysis; PSO, particle swarm optimization; RBF, radial basis function; RBFNN, radial basis function neural network; RC, resistance capacitance; RF, random forest; RH, relative humidity; RMSE, root mean square error; RNN, recurrent neural network; SARIMA, seasonal autoregressive integrated moving average; SARIMAX, seasonal autoregressive integrated moving average exogenous; SARSA, state-action reward state-action; SOM, self-organizing map; SR, solar radiation; SVM, support vector machine; SVR, support vector regression; WD, wind direction; WS, wind speed

* Corresponding author.
E-mail address: xqzhai@sjtu.edu.cn (X.q. Zhai).

https://doi.org/10.1016/j.scs.2019.101533
Received 17 November 2018; Received in revised form 20 February 2019; Accepted 2 April 2019
Available online 14 April 2019

To address the stakes of rapidly growing urbanization, the increasing need of human comfort and consequent energy consumption increase, solutions emerge in the development of smart sustainable infrastructures (Silva, Khan, & Han, 2018). Smart and low to zero energy buildings play a significant role (Kylili & Fokaides, 2015) in many aspects including global energy efficiency, energy conservation measures and the integration of renewable energy systems. Hence, building energy consumption modeling and forecasting is a key tool to achieve smart and sustainable designs. Indeed, it can assist in higher energy efficient designs by comparing several strategies for both pre- (Tahmassebi & Gandomi, 2018) and post-occupancy studies (Ruparathna, Hewage, & Sadiq, 2017). It can also guide energy management at local and global scales (Xu, Taylor, Pisello, & Culligan, 2012).

Among the three main approaches in building energy consumption modeling and forecasting (BECMF) – physics-based, data-driven and hybrid models (Dong, Li, Rahman, & Vega, 2016), data-driven techniques emerge as the most suitable option to ensure the integration of buildings in smart environments. Smart infrastructures rely on sensor networks which generate large amounts of energy-related data (Rathore et al., 2018). For instance, massive smart-meters deployment programs have been launched with in Europe, the United States of America and China during the past decade with ambitious goals to achieve by 2020 (Liu, Marnay, Feng, Zhou, & Karali, 2017; Obey, 2009; Smart Metering deployment in the European Union|JRC Smart Electricity Systems and Interoperability. (n.d.).; U.S. Energy Information Administration (EIA), 2018). Then, as the name suggests, data-driven methods propose modeling and forecasting frameworks based on data analysis schemes rather than on classical physics-based modeling tools (Foucquier, Robert, Suard, Stéphan, & Jay, 2013). Furthermore, these frameworks include algorithms that take benefit from the recent significant developments in the field of machine learning in recent years (Wang & Srinivasan, 2017), providing flexibility and reliability to modeling and forecasting tools. Consequently, data-driven building energy consumption modeling techniques have recently drawn an increasing attention, providing new case studies, algorithms and results while technical challenges remain (Bourdeau, Guo, & Nefzaoui, 2018).

Thus, we report in the present paper a review on data-driven building energy modeling techniques. It aims to introduce the most prevalent techniques and to further provide an up-to-date overview of recent studies and advancements in BECMF studies, as well as research gaps and promising research directions. The paper is organized as follows: in Section 2 the data-driven forecasting process is described, performance assessment metrics are defined, and the prevalent techniques are presented namely autoregressive models (AR), statistical regressions, *k*-nearest neighbors (*k*-NN), decision trees (DT), support vector machine (SVM) and artificial neural networks (ANN). Section 3 discusses and compares the application of different machine learning approaches for data-driven techniques. Section 4 summarizes the characteristics of the data used and pre-processing methods in data-driven forecasting processes. Section 5 discusses the challenges in terms of building typologies, energy end-uses, forecasting horizons and the implications of the lack of occupant-related data. It also summarizes the current trends and latest technical achievements in machine learning applications to building energy forecasting studies, while highlighting their limitations and possible solutions.

1.1. Research methodology

The research methodology followed six steps. The first step relied on the analysis of existing review papers to highlight (1) the trends in research and applications of BECMF techniques over the past decade which has witnessed massive smart-metering deployment programs and an increasing amount of energy demand data production (Liu et al., 2017; Obey, 2009; U.S. Energy Information Administration (EIA), 2018), (2) the main existing categories and classes of techniques to (3) build a classification and to cross-check the corresponding nomenclature. Indeed, the many recent applications and technical improvements of BECMF methods have introduced numerous names of techniques that may confuse non-experts.

Based on the developed classification, a key-word search was conducted for the broad field of BECMF and more specifically for each types of techniques. Google Scholar was used as it provides relevant information. The number of citations helped target reference articles. These articles highlighted more recent citing papers that build on previous research work to introduce novel applications and methods. Authors' names were also used to search for related and relevant similar studies.

A selection of articles retrieved from the key-word search was performed based on three criteria: (1) the publication date in the past decade to consider relatively recent research work; (2) the focus on forecasting energy consumption and load demand in buildings (including overall energy, thermal energy with combined and separated cooling and heating loads, and other loads such as lighting or plug load); (3) the application-scale focusing on building-scale studies (neighborhoods, cities, regions and countries were excluded).

The fourth step aimed to highlight specific information for each selected studies, and summarized in a table for study comparisons. It included the type(s) of technique(s) implemented, the characteristics of the building(s) case study(ies) (number of case studies, building type(s), location(s)), the characteristics of the input (data type, granularity, amount) and output data (type(s) of end-use(s), forecasting horizon, accuracy) and the modeling tools or software used.

With the collected information, research articles were selected to serve at least one of the three following purposes: (1) to provide a solid but accessible theoretical and application reference for one specific type of technique; (2) to present an original approach in terms of forecasting technique and/or application case-study, input data, end-use(s) or multidisciplinary work; (3) to propose a comparative study with insights on the different methods implemented, their performances, advantages and weaknesses. Also, this step helped identify relevant approaches not covered by existing review articles.

Finally, a last comparison was conducted with existing state-of-the-art papers to select the most relevant examples of BECMF methods. An effort was made to select original research works that were little or not already reported. Moreover, techniques with few new application studies compared to these that have already been covered were excluded. Nevertheless, these techniques were included in the classification of BECMF methods presented in the following section. They are briefly introduced when encountered in the selected articles.

1.2. Classification of methods for building energy consumption modeling and forecasting

Numerous and various techniques have been developed, adapted and used for BECMF. The significant research efforts on the topic over the past twenty years have led to several previous reviews describing the existing methods and using different nomenclatures. To aid the readers' understanding of the different techniques, and re-contextualize the scope of the present review, this part of the study compares the different classifiers encountered in fifteen reviewed state-of-the-art articles (Ahmad, Chen, Guo, & Wang, 2018; Amasyali & El-Gohary, 2018; ASHRAE, 2009, chap. 19; Chalal, Benachir, White, & Shrahily, 2016; Deb, Zhang, Yang, Lee, & Shah, 2017; Foucquier et al., 2013; Fumo, 2014; Mat Daut et al., 2017; Pedersen, 2007; Swan & Ugursal, 2009; Tardioli, Kerrigan, Oates, O'Donnell, & Finn, 2015; Wang & Srinivasan, 2017; Wei et al., 2018; Yildiz, Bilbao, & Sproul, 2017; Zhao & Magoulès, 2012). A unifying nomenclature is then presented in Fig. 1. The suggested classification is based on the differences in modeling processes, without building type or energy end-use distinctions and considering building-scale applications.

The review work has highlighted three main categories of BECMF

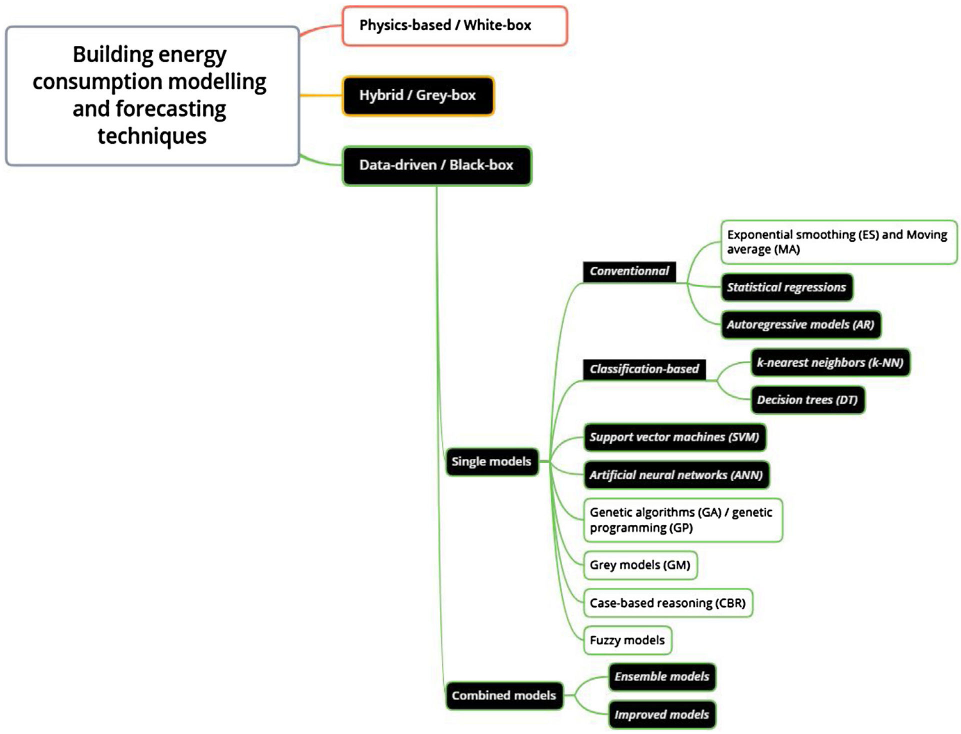

Fig. 1. Summary classification of building energy consumption modeling and forecasting methods (techniques with white font on black background are the techniques covered in this paper).

models. The first, physics-based models, is also commonly referred as "white-box" (Tardioli et al., 2015). It uses a transparent process based on physics equations solving to describe the energy behavior of buildings. Physics-based modeling has been introduced with different names, either called "forward classical approach" and "calibrated simulation approach" (ASHRAE, 2009, chap. 19), "energy simulation programs" (Pedersen, 2007), "engineering methods"/approach (Fumo, 2014; Swan & Ugursal, 2009; H. Zhao & Magoulès, 2012) "physical modelings" (Foucquier et al., 2013), or "thermal models" (Yildiz et al., 2017). Sub-classifications of these models have been proposed as well depending on the origin of input data (Swan & Ugursal, 2009), the level of details implemented in the modeling (Foucquier et al., 2013) and the modeling calibration methods (ASHRAE, 2009, chap. 19; Fumo, 2014).

The second category is data-driven models. They mainly rely on time-series statistical analyses and machine learning algorithms to assess and forecast the building energy consumption. They are also often named "black-box" models (ASHRAE, 2009, chap. 19; Tardioli et al., 2015) to emphasize that the relationship between inputs and outputs can hardly be transposed to physics-based analysis with these techniques. "Data-driven" techniques (Ahmad et al., 2018; Amasyali & El-Gohary, 2018; Tardioli et al., 2015; Wei et al., 2018), have also been named "time series […] techniques" (Deb et al., 2017) and "statistical" (Chalal et al., 2016; Swan & Ugursal, 2009; Zhao & Magoulès, 2012). Furthermore, data-driven "statistical analyses", "regressions-based models" and "auto-regressive models" regarded as more conventional methods (Mat Daut et al., 2017; Pedersen, 2007; Yildiz et al., 2017; Zhao & Magoulès, 2012) have been differentiated from artificial intelligence models referred as "intelligent computer systems"/techniques (Pedersen, 2007), "intelligent techniques" (Fumo, 2014), "AI approach" (Mat Daut et al., 2017) or "machine learning models" (Yildiz et al., 2017). More recently, Wang and Srinivasan (2017) reviewed data-driven models in building energy consumption prediction by opposing "single models" and "ensemble models". The former uses a single algorithm for a straightforward forecasting process, while the latter build a framework managing the strengths and weaknesses of techniques. Within all these classes of data-driven models, several specific popular techniques have been described (Deb et al., 2017; Zhao & Magoulès, 2012). They are organized as follow in the proposed classification for data-driven methods: single models are divided between (1) classical techniques with moving average & exponential smoothing (MA & ES), autoregressive models (AR) and statistical regressions, (2) classification-based techniques applied to forecasting purpose with *k*-nearest neighbors (*k*-NN) and decision trees (DT), (3) support vector machines (SVM), (4) artificial neural networks (ANN), (5) genetic algorithms (GA), (6) grey modeling, (7) case-based reasoning and (8) fuzzy models. On the opposite, a category named combined models includes both ensemble models and improved models. The latter refers to the combination of single data-driven techniques and optimization methods (Mat Daut et al., 2017).

Finally, the third and last of the main category of models is hybrid models. It describes the combination of physics-based and data-driven methods. They are also called "gray-box" or "grey-box approach" (Foucquier et al., 2013; Tardioli et al., 2015), as the combination of white-box and black-box methods. Other techniques have also been named "hybrid models" (Chalal et al., 2016; Mat Daut et al., 2017) but were referring to the improvement of single data-driven techniques with optimization methods, or the combination of several machine learning algorithms. In the proposed classification these are called improved models as described in the previous paragraph.

1.3. Overview of the papers reviewed

The study covers eight classes of data-driven models with autoregressive models (AR), statistical regressions, *k* nearest neighbors (*k*-NN), decision trees (DT), support vector machines (SVM), artificial neural networks (ANN), ensemble and improved techniques. Hybrid models are also discussed in the discussions of this review. Based the research methodology previously described, techniques including moving average, exponential smoothing, genetic algorithms (GA), grey models, case-based reasoning and fuzzy-based models are not described in detail in the present work (black font on white background in Fig. 1). They have been investigated during the review process, however,

Table 1
Counting of the number of reviewed research papers implementing the different BECMF techniques from 2007 to 2019.

BECMF techniques		2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total	
AR					1		1		2	3	1				8	
Statistical regressions	Supervised	1					2		3	2	2	3	3		16	17
	Unsupervised								1						1	
k-NN									2		2	1			5	
DT		1			1				1	1		1	2		7	
SVM	Supervised			1					4	4	3	4	2		18	20
	Unsupervised								1						1	
	Transfer learning												1		1	
ANN	Supervised		1	1		1	1		4	4	3	2	2		20	22
	Unsupervised											1			1	
	Transfer learning												1		1	
DNN	Supervised										2	1		1	4	6
	Unsupervised											1			1	
	Reinforcement										1				1	
Ensemble	Supervised						1		2		1	3	4		11	16
	Unsupervised								1			2			3	
Improved										2	2	1	1		6	
Hybrid						1					1		1		3	

results of the literature review highlighted case-based reasoning (Kolodner, 2014), fuzzy-based models (Song & Chissom, 1993) and grey models (Deng, 1989) have had few new applications compared to these already covered in other review papers. Also, for grey modeling, fuzzy-based models, exponential smoothing and moving average, the applications were mainly focusing country-scale studies while the present work limits the applications to building-scale. GA (Mitchell, 1998) can be found in the literature but have been sparsely applied alone as the prediction technique for building energy consumption forecasting. They are mostly implemented as an optimization tool as described later in this review. Finally, physics-based models are out of the scope of this paper, focusing on data-driven methods. Hence, these techniques are only briefly presented when implemented in the reviewed articles.

In the present article, a total number of fifty original research papers have been reviewed. Moreover, a counting of the number of studies implementing the different reviewed BECMF techniques has been performed. It is presented in Table 1 with the number of papers per year and the total number of papers over the 2007–2019 period and for each specific approach. A distinction is made between more conventional ANN and deep (learning) neural networks (DNN), as well as between supervised, unsupervised, reinforcement and transfer learning approaches.

2. Data-driven techniques

2.1. Data-driven forecasting process: training, validation and testing

Data-driven techniques use statistical and machine learning tools to develop an energy model of a building. Most techniques focus on time series data analysis but also frequently include basic knowledge on the buildings' characteristics. The data-driven modeling process involves three steps that rely on three different sets of data. These datasets usually result from the division of a main original one and include the same input variables but with different combinations of values and for different periods of time (Bishop, 2006).

The first step is the training of the algorithm. The model is run on the training dataset to produce results. These results are compared to the original training data and based on the results of the comparison, the different parameters of the algorithm can be adjusted to fit on the training dataset (Wahid & Kim, 2016). The second step is the validation. A validation dataset is used to provide an unbiased evaluation of the implemented algorithm, already fit on training data, and to tune its key modeling parameters to enhance the fitting of the model. The validation dataset must be different from the training dataset to prevent overfitting (Zhang, Deb, Lee, Yang, & Shah, 2016). Otherwise, it would result in a model performing very well with a specific set of data but poorly with other datasets.

Finally, the third step is the testing step when the algorithm developed is run on the remaining part of the data to provide a final unbiased evaluation of the modeling and forecasting performances. It is commonly admitted that the model parameters and structure should not be modified based on the results of this final step (Fan, Xiao, & Wang, 2014). Several methods have been used to pre-process the datasets and select relevant input data with adapted training–validation–testing ratios. They will be presented along with the review of the different studies in the following sections. Nevertheless, it should be highlighted that in practice training and validation steps are not always explicitly separated.

2.2. Accuracy metrics

Forecasting performances of data-driven algorithms are tested using accuracy metrics. The most common are the mean absolute percentage error (MAPE), the root mean square error (RMSE), the coefficient of variation of RMSE (CV-RMSE) and the mean average error (MAE) assessed in 53%, 47%, 38% and 36% of the reviewed studies respectively. The coefficient of determination (R^2), the mean square error (MSE), the mean relative error (MRE), the mean bias error (MBE) and the normalized mean bias error (NMBE) can also be found in 27%, 16%, 9%, 2% and 4% of reviewed studies respectively. Finally, some authors defined specific accuracy metrics such as the relative error (Liu, Chen, & Mori, 2015), average error (Neto & Fiorelli, 2008) and accuracy rate (Wahid & Kim, 2016; Yu, Haghighat, Fung, & Yoshino, 2010).

Different metrics provide different information on the forecasting performances and the model behavior for different datasets (Hyndman & Koehler, 2006). Contrariwise unit-based metrics (i.e. MAE or RMSE for instance), performance evaluation based on error percentages provide normalized information. Thus, they should be preferred for comparisons between different models, studies and building typologies. Since MAPE was the most used in reviewed articles, this metric has been selected as the reference accuracy metric for the present review work (and in Appendix A). Otherwise, when MAPE was not available other metrics included CV-RMSE, RMSE, MAE and R^2 were provided to illustrate the forecasting performances of the different implemented algorithms. The specific definition of each of these metrics can be found in Eqs. (1)–(9).

$$\text{Mean Absolute Error (MAE)} = \frac{1}{n}\sum_{i=1}^{n} |y_{forecasting,i} - y_{observed,i}| \tag{1}$$

$$\text{Mean Absolute Percentage Error (MAPE) (\%)} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_{forecasting,i}-y_{observed,i}}{y_{observed,i}}\right|*100 \tag{2}$$

$$\text{Mean Square Error (MSE) (W or kW)} = \frac{1}{n}\sum_{i=1}^{n}(y_{forecasting,i}-y_{observed,i})^2 \tag{3}$$

$$\text{Root Mean Square Error (RMSE) (W or kW)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{forecasting,i}-y_{observed,i})^2} \tag{4}$$

$$\text{Cofficient of Variation of RMSE (CVRMSE) (\%)} = \sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}(y_{forecasting,i}-y_{observed,i})^2}{\bar{y}_{observed}}}*100 \tag{5}$$

Also called Normalized RMSE (NRMSE) or Root Mean Square Percentage Error (RMSPE)

$$\text{Cofficient of Determination } (R^2) \text{ (unitless)} = 1-\frac{\sum_{i=1}^{n}(y_{forecasting,i}-y_{observed,i})^2}{\sum_{i=1}^{n}(y_{observed,i}-\bar{y}_{observed})^2} \tag{6}$$

$$\text{Mean Bias Error (MBE) (W or kW)} = \frac{1}{n}\sum_{i=1}^{n}(y_{forecasting,i}-y_{observed,i}) \tag{7}$$

$$\text{Normalized Mead Bias Error (NMBE) (\%)} = \frac{\frac{1}{n}\sum_{i=1}^{n}(y_{forecasting,i}-y_{observed,i})}{\bar{y}_{observed}}*100 \tag{8}$$

$$\text{Mean Relative Error (MRE)} = \left|\frac{\bar{y}_{forecasting,i}-\bar{y}_{observed,i}}{\bar{y}_{observed,i}}\right| \tag{9}$$

With $y_{forecasting,i}$ is the forecasted energy consumption at time point i, $y_{observed,i}$ is the real energy consumption data at time point i, $\bar{y}_{observed,i}$ is the average of the real energy data consumption over the considered timeframe, and n is the total number of data in the dataset considered for performance evaluation.

2.3. Single models

Single models are data-driven techniques implementing only one predictive algorithm for a forecasting problem. In this paper, single models include conventional methods with autoregressive models and statistical regressions, classification-based methods with k nearest neighbors and decision trees, support vector machine and artificial neural networks.

2.3.1. Conventional methods

Conventional methods refer to autoregressive models and statistical regressions, two popular techniques which have been widely implemented for BECMF. They provide a good balance between implementation simplicity and forecasting accuracy. However, they have shown significant limitations with respect to the forecasting horizon and the ability to model nonlinear data patterns.

2.3.1.1. Autoregressive models (AR). Autoregressive modeling is one of the most classical modeling and forecasting techniques and is based on statistical analysis of time-series. It only requires the training set to be stationary: this means that statistical properties of the time-series should be time-invariant, or in other words that the energy consumption at a specific time should be similar to this of the recent past. Common models include AR and auto-regressive integrated moving average (ARIMA) models, also called Box–Jenkins models (Box, Jenkins, & Reinsel, 2008). *ARIMA(p,d,q)* is composed of lagged terms from the input time series with the autoregressive part *AR(p)* of order p, and of lags of the forecasting error with the moving average part *MA(q)* of order q. When the input time series is not stationary it can be differentiated: the order d indicates the degree of differentiation. When the time series already is stationary, and therefore differentiation is not necessary, ARIMA can also be noted *ARMA(p,q)*. The output of ARIMA modeling is a linear equation, combining both the autoregressive and moving average parts as follows:

$$\hat{Y}_t = C + \sum_{i=1}^{p}\varphi X_{t-i} - \sum_{j=1}^{q}\theta_j\varepsilon_{t-j} \tag{10}$$

With t the time-step, $\hat{Y}$ the predicted value and X the time series values. φ is the coefficient of the autoregressive model, θ the is coefficient of the moving average model and C is a constant. ε is the forecasting error.

AR models are relatively simple to implement. Basic autoregressive models only consider the recent past historical load demand data points to predict its future states. Therefore, they can only provide short-term forecasting, which limits their application scope and accuracy. Several technical improvements have been developed to overcome these issues. SARIMA models (seasonal ARIMA) (Jeong, Koo, & Hong, 2014) append additional seasonal terms to a standard ARIMA to account for events and trends happening at a regular pace. They are noted *SARIMA(p,d,q)(P,D,Q)*$_s$ with p, d and q related to the non-seasonal part of the data as presented above and P, D, Q the lagged terms of the seasonal part of the data for a lag of S (the period of the events/trends). Also, ARIMAX models (ARIMA with inclusion of eXogenous variables) (Newsham & Birt, 2010) consider the impact of parameters other than the past load demand on the energy consumption such as weather conditions or occupancy. They are added to the standard ARIMA models as a linear combination the past b terms of their corresponding time series. ARIMAX models can then be noted *ARIMAX(p,d,q,b)*. Finally, both SARIMA and ARIMAX can be combined.

Indeed, Newsham and Birt (2010) developed a SARIMAX model with occupancy data from network logins and daily power seasonality, using IBM SPSS Statistics ("IBM SPSS Statistics, " n.d.). It aimed to forecast occupancy-related electricity load (lighting, office and lab equipment, plug loads, without chiller power) of an office and research building in Ontario, Canada. The model was compared to a SARIMA model and results showed a MAPE of 1.24% and 1.22% for SARIMA and SARIMAX respectively. Yun, Luck, Mago, & Cho (2012) implemented four 4th-order ARX models to predict separately ahour-ahead building cooling and heating loads. Three models were indexed with time (different hours of the day) or time periods and temperature levels. Test cases were benchmark buildings simulated in EnergyPlus ("EnergyPlus, " n.d.), a physics-based modeling software. Building typologies included small-office, medium-office, midrise apartment and high-rise apartment buildings. ARX models were compared to a simple AR model, a multi-linear regression (MLR) model and a back-propagation neural network (BPNN). The ARX model indexed with three time periods of the day and five OAT levels performed better for every four building types and for both cooling and heating periods. Dagnely, Ruette, Tourwé, and Tsiporkova (2015) developed a seventh-order AR model to forecast the next 72-h electricity load demand of an office building in Brussels, Belgium. They proposed a comparison with an ordinary least square regression (OLS) using Python Statsmodels ("StatsModels: Statistics in Python — statsmodels 0.9.0 documentation, " n.d.) and a support vector regression (SVR) using Python Scikit-Learn (Scikit-Learn, 2019"Scikit-Learn: machine learning in Python, " n.d.). Various inputs combinations were considered including day type, occupancy, OAT, SR and previous-week same-day logged energy consumption called "recency". The AR model gave the best MAE of 2.01 kW. OLS performances ranged between a MAE of 2.05 kW for all variables and a MAE of 3.74 kW with temperature only. SVM performances ranged between a MAE of 1.94 for all variables and "recency" only, and a MAE of 3.46 kW with temperature only.

2.3.1.2. Statistical regressions. Statistical regressions aim to model a relationship between an output and contributing inputs, also called explanatory variables, in the form of an equation. For BECMF, several types of statistical regressions can be found in the literature. These include multiple linear regressions (MLR), also called conditional demand analysis (CDA) (Parti & Parti, 1980); ordinary least square regressions (OLS) (Dagnely et al., 2015); piecewise linear regressions also called segmented regressions (Zheng, Zhuang, Lian, & Yu, 2017); general linear regressions (Chou & Bui, 2014); elastic net regressions (Fan, Xiao, & Zhao, 2017); Bayesian regressions (Gelman et al., 2013); and Gaussian process regressions (Rasmussen & Williams, 2006). In case that the modeling pattern shows high non-linearity, literatures reported the efficiency of multivariate adaptive regression splines (MARS) (Friedman, 1991).

Statistical regressions have been widely implemented for both pre-occupancy (design phase) and post-occupancy forecasting studies such as energy retrofit impact assessment. Their popularity is mostly related to their simple implementation and relatively explicit formulation to link output energy consumption to input explanatory variables. Moreover, forecasting performance of statistical regressions are reasonably good for most applications. Nevertheless, if the simplicity of statistical regressions generally is a strong advantage it also induces one of their major drawbacks. Indeed, most regression techniques are unable to deal with non-linear phenomena which are common in building energy efficiency studies. Moreover, a large amount of data is also required to capture all possible scenarios.

For instance, Amber et al. (2017) implemented a MLR to predict the daily electricity consumption of an administrative building and an academic building in London, England. Input data included daily mean OAT, RH, SR and WS, weekday index and building type, and were narrowed down to daily mean OAT, weekday index and building type after a collinearity study. Over five years of data collected. Four years were used to train the regression model and one year to test it. Results showed a MAPE of 8.58% for the administrative building and a MAPE of 9.76% for the academic building. Pulido-Arcas, Pérez-Fargallo, and Rubio-Bellido (2016) developed MLR forecasting models for office buildings in Chile using a government database. It included building characteristics with the number of stories, floor area, form ratio, wall-to-window ratio (WWR), coefficient of performance (COP), energy efficiency ratio, and heating and cooling emission factors. Models were adapted to nine locations with specific climate datasets. They were used to assess total energy consumption (electricity and natural gas), comparing 77,000 possible office buildings for each climate locations. Nine regression models were prepared and energy consumption forecasting results (MAE) ranged between 0.11 kW and 0.41 kW.

For non-linear dynamics, statistical regressions with MARS can compete with more complex data-driven methods presented in detail in the following sections of this review. For cooling and heating load forecasting Sekhar Roy, Roy, and Balas (2018) compared MARS with a linear regression, a Gaussian process, a simple ANN, a radial basis function neural network (RBFNN), an extreme learning machine (ELM) model and an ensemble model of MARS and ELM. Models were trained and tested with an open database of 768 building samples. For heating load prediction, the ensemble model gave the highest accuracy and MARS ranked second, followed by ANN, Gaussian Process, ELM, Linear regression and RBFNN with a MAE of 0.037 kW, 0.077 kW, 0.085 kW, 0.175 kW, 0.189 kW, 0.196 kW and 0.354 kW, respectively. For cooling load, models also ranked the same with the ensemble model first, then MARS and ELM, with MAE of 0.127 kW, 0.146 kW and 0.238 kW, respectively.

2.3.2. Classification-based methods

Classification-based methods have been successfully implemented for modeling and forecasting purpose. Two of their most popular representatives are *k*-nearest neighbors (*k*-NN) and decision trees (DT). Both are intuitive techniques with high forecasting accuracy. However, both are also limited by the need of comprehensive input dataset. They allow only qualitative analyses of their results.

2.3.2.1. K-nearest neighbors (k-NN). *K*-nearest neighbors is a popular technique for pattern clustering and classification that was first introduced by Fix and Hodges, Jr (1951). When applied on time-series, it relies on the idea that similar patterns can be identified and classified according to their properties: for instance, energy demand or consumption can be related to occupancy, weather data and other relevant parameters. Thus, given a set of historic observations (energy consumption and other variables), clusters are first created. They are constructed with respect to a user-defined feature: peak load, average, magnitude of daily load variation, daily consumption (integral), etc. These features are calculated for each time series and then used for classification. A less user dependent process relies on the calculation of distances between each pair of time-series for which a metric is then to be defined. A simple Euclidian distance for example can be used and interpreted as a difference of energy consumption (Toffanin, 2016). Then, new observation data are compared with the clusters by defining a degree of closeness based on two criteria. The first criteria, the parameter *k*, sets the number of neighbors (the closest data points) to which the target observations should be compared. The second criteria is the metric for comparison and classification. It is usually the same as for the previous clustering step. Once the comparison is made and the new observation data are associated to a specific cluster, energy forecasting can be performed.

K-NN modeling is intuitive, relatively simple to implement and shows good forecasting accuracy. In most reviewed studies, it has been applied for short-term forecasting horizon with hourly time-step. Similarly, to other classification methods, it has the advantage to enable the utilization of categorical variables for energy driver considerations and to create the neighbors groups. However, its forecasting ability relies on the amount of input data available: the accuracy depends on the presence of similar "conditions" resulting a similar output in the database.

Valgaev and Kupzog (2016) developed a *k*-NN model for 24-h-ahead overall electricity load forecasting of mixed-use buildings with different sizes and aggregated end-consumers load. An Irish energy database comprising over 6000 low-voltage buildings was used to model daily profiles from smart-metering and day-type was also distinguished. Three building sizes (25, 50 and 100 end-consumers) were then generated, with 70% of residential spaces and 30% of commercial spaces, respectively, and with samples of 100 buildings for each size. Accuracy results were assessed with MRE and showed that a higher number of end-consumers induced less accurate forecasting results, with 0.975, 0.968 and 0.940 for 25, 50 and 100 end-consumers, respectively. Wahid and Kim (2016) implemented *k*-NN for next-day total electricity consumption forecasting of residential buildings using both MATLAB ("MATLAB – MathWorks – MATLAB Simulink, " n.d.) and Weka ("Weka 3 – Data Mining with Open Source Machine Learning Software in Java, " n.d.) software. Appliance-level hourly energy consumption data were collected for 520 apartments in Seoul, South Korea. Apartments were then divided between low and high-power demand ones, considering the daily profiles generated. Different training-testing ratios were considered, and results showed that the most robust ratio was 60% training–40% testing, with 95.96% of accurately forecasted results. Ma, Song, and Zhang (2017) proposed a method with combined weight selection of similar days, applied on a government office building in Jiangsu Province, China. Based on day type and daily weather type (sunny, cloudy, rainy, and overcast), they extracted hourly OAT, lighting and plug, and air-conditioning loads to create daily electricity load profiles. One reference working day and one reference vacation day were then used with each weather scenario to forecast hourly air conditioning (AC) load. An eQuest (physics-based) model ("eQUEST, " n.d.) was also implemented for comparison. The relative errors for physics-based modeling and k-NN models, for working day/vacation day ranged between [5.59%; 13.6%]/[5.23%; 17.6%] and [1.31%; 5.05%]/[0.83%; 3.68%], respectively. Lachut, Banerjee, & Rollins (2014) developed 5-NN ($k = 5$) forecasting models to predict building-level power demand in residential buildings. They used their own dataset recorded on seven different buildings. Data provided electricity loads with 30-s time-step and were aggregated to hourly, 6-h, daily and weekly time-step, depending on the forecasting horizon. k-NN was compared with Bayesian regression, SVM and ARMA(1,1) using past 24-h loads with time-related information (hour of the day, day of the week and quarter of the day). Results showed that k-NN performed better for both appliance and home levels with a one-week forecasting horizon. However, it was the least performing algorithms for 1 h-, 6 h- and one day-ahead forecasting horizons.

2.3.2.2. Decision trees (DT). Decision trees are a popular machine learning method also applied in regression problems for forecasting applications. It follows the simple idea of a tree growing from roots to leaves. Hence, a DT starts with a root node leading to other successive non-leaf nodes. At each node, a test is performed by considering a specific condition on an input variable, either binary or categorical, and the branches keep splitting until leaf-nodes are reached to figure a possible value of the predicted output (Fig. 2). There is then a path to follow from the root node to the leaf-nodes through decision-making.

Several types of DT have been developed. The most common for building energy consumption forecasting are classification and regression trees (CART) (Breiman, Friedman, Olshen, & Stone, 1984), chi-squared automatic interaction detector (CHAID) (Kass, 1980), ID3 (J R Quinlan, 1986), C4.5 (John Ross Quinlan, 1993), and C5.0. CART refer to classification trees when the predicted output is the class the data belongs to, and to regression trees when the predicted output is a number (for forecasting applications). CHAID detects interdependency between the different variables of a dataset and therefore allows to study the influence of explanatory variables on the result. Finally, C4.5 is an entropy measurement-based DT and improved version of ID3. C5.0 is an optimized version of C4.5 in terms of computation speed, memory allocation and tree sizing.

DT are flexible techniques that have been applied for both early design stage and post-occupancy studies. The accuracy of prediction results is comparable to other single data-driven techniques such as artificial neural networks and support vector machines. However, DT have the significant advantage to be easy to apprehend and with reasonably complicated implementation and operation.

Tso and Yau (2007) compared stepwise regression, multi-layer perceptron (MLP) and DT models of residential households in Hong-Kong. 1516 buildings were surveyed during two distinct periods, in winter and summer. They collected power rating appliances for each end-use and their corresponding half-hour time-step usage patterns. A database was then created and divided into three categories with residential housing types, household characteristics and appliance ownership details. Each of the models was trained with the same database, for both periods. Results for electricity consumption forecasting showed that the DT performed slightly better for summer time with RMSE of 39.36 kWh, compared to the ANN (RMSE of 39.53 kWh) and the regression (RMSE of 39.42 kWh). However, higher accuracy was achieved for winter time with the ANN (44.14 kWh) and with the regression (44.18 kWh) than with DT (44.40 kWh). Chou and Bui (2014) compared several models including CART, CHAID, SVR, ANN, general linear regression and ensemble models with different combinations of these techniques. They were implemented in IBM SPSS Modeler (IBM SPSS Modeler, 2019"IBM SPSS Modeler, " n.d.) to predict separate heating and cooling load of twelve different building types. Input variables were extracted from an open-database of 768 building samples simulated in Ecotect tool (Tsanas & Xifara, 2012). It included relative compactness, surface area, wall area, roof area, overall height, orientation, glazing area and glazing area distribution. For cooling load assessment results (MAPE), methods ranked as follows: SVM with 2.99%, the four different ensemble models between 3.46% and 3.54%, then CART & CHAID both with 4.02%, ANN with 4.40% and regression with 4.96%. For heating load (MAPE), the method ranking was: SVR with 1.13%, the four different ensemble models between 1.56% and 1.61%, CART with 2.10%, ANN with 2.36% CHAID with 2.41% and regression with 4.59%. Finally, Yu et al. (2010) implemented a C4.5 DT to model building energy use intensity with Weka Software ("Weka 3 – Data Mining with Open Source Machine Learning Software in Java, " n.d.). It was based on an 80-residential-building database from six different Japanese districts. The database included energy uses of the different energy sources in each building and at different time-steps, with outdoor air temperature, building characteristics and other information such as occupant number and energy saving measures. The C4.5 DT assessed the energy use intensity for each building and classified them as "HIGH" or "LOW". Results showed a 92% success rate for the classification.

2.3.3. Support vector machines (SVM)

Support vector machines (Cortes & Vapnik, 1995) are a popular and efficient technique for non-linear problems solving. It gives accurate results even with a relatively limited amount of available data. For forecasting applications, the process is similar to the resolution of a regression problem and is called support vector regression (SVR) (Smola & Schölkopf, 2004). Therefore, as for all regression problems,

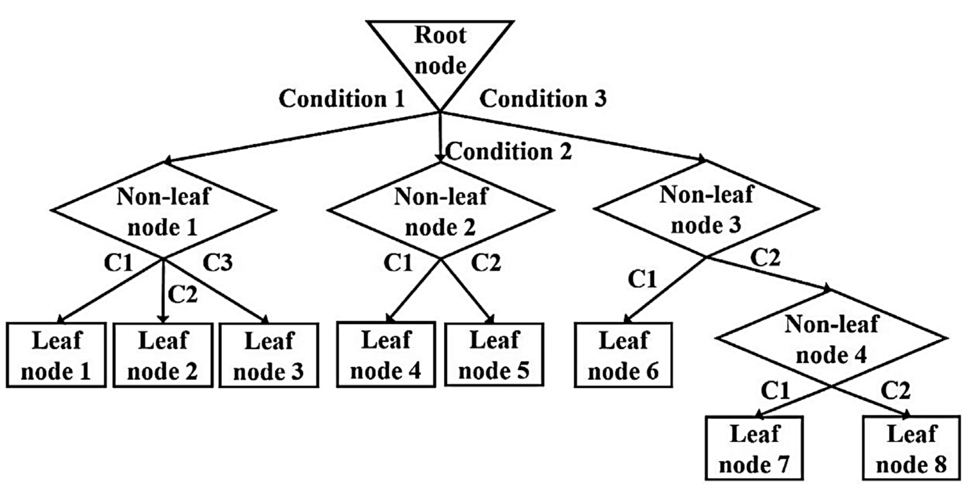

Fig. 2. Schematic of a DT (decision tree) with input-variable-based conditional separating into non-leaf nodes until final leaf-nodes are reached.

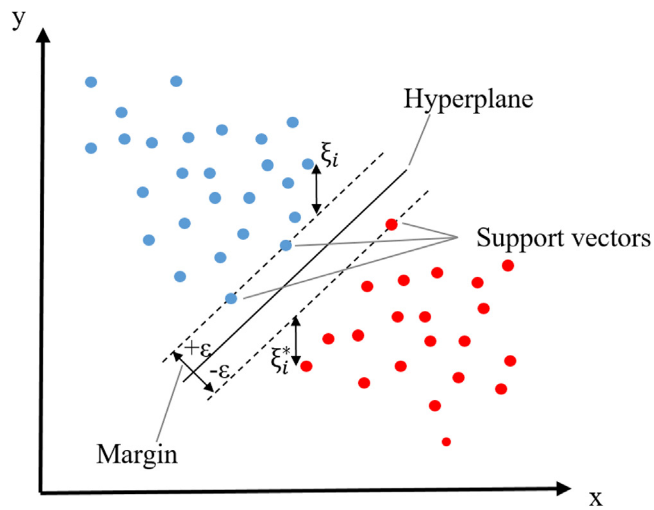

Fig. 3. Representation of the division of a dataset into two subsets using SVR.

the goal is to find a best-fitting function which in SVR modeling is developed based on the search of a decision hyperplane splitting a given dataset into two sub-datasets. Moreover, the particularity of SVR is that it tolerates an error in the regression. This error is called largest margin and is defined as the maximum distance between the separation boundary of the hyperplane and the closest data samples, named support vectors (Fig. 3). Finally, the key feature of SVR is the definition of a kernel function. The idea behind the kernel function is to change the representation space of the datasets to a higher dimension where there is probably a linear separation of the two datasets. Indeed, sometimes a dataset cannot be directly and linearly divided into two sub-datasets. Then, the kernel selection has a significant impact on the performance of SVM model. It can be linear, polynomial or a radial basis function (RBF), also called Gaussian kernel. Several sub-types of SVM can also be found in the literature such as a simplification of standard SVM called least-square SVM (LS-SVM) (Li, Lu, Ding, Xu, & Li, 2009), ε-SVR and ν-SVR (Zhang et al., 2016). However, a detailed description of these techniques is beyond the scope of this study.

SVR have been implemented with most time-steps, input variables and considering real or simulated data. Nevertheless, despite competitive forecasting performances and application flexibility, SVM present a major drawback: the calibration of their parameter is a difficult but decisive process for prediction accuracy. For instance, the kernel function for example is challenging to accurately determine and significantly affects the accuracy of the forecasting. Therefore, the optimization of SVM parameters has become a key challenge in building energy studies (Chen & Yang, 2018; Fu, Li, Zhang, ö Xu, 2015). Finally, as a black-box model SVM is completely non-transparent in terms of physics-based interpretation.

Paudel et al. (2017) implemented SVM with LIBSVM library (Chang & Lin, 2013) to a TRNSYS-simulated ("TRNSYS: Transient System Simulation Tool, " n.d.) residential low energy building, in four different French cities and with four different climatic conditions. They aimed to forecast the combined cooling and heating energy demand with input variables including OAT, SR, solar gain through window and on walls, past-time steps of these variables, occupancy profile and day indicator. Two different kernel functions were selected. A linear kernel was used to train a SVM and determine the weight of weather data on the energy consumption. A RBF kernel was used for the prediction of the energy load. Three simulation scenarios were implemented, with different combinations of input variables after a "relevant data" selection using the linear kernel SVM. Results gave median RMSE of 13.1 kW, 4.2 kW and 3.2 kW for the three scenarios, respectively. Also, the relevant data selection was compared to an all-input model for the third scenario and showed that input selection could improve the forecasting accuracy with median RMSE of 3.4 kW against 9.1 kW.

Support vector machines also perform rather well compared to other popular data-driven techniques such as ANN or statistical regressions. For instance, Massana, Pous, Burgas, Melendez, & Colomer (2015) proposed a comparison of the three methods implemented on Weka software (Weka, 2019"Weka 3 – Data Mining with Open Source Machine Learning Software in Java, " n.d.) for short-term electric load forecasting of an university office building in Girona, Spain. Seven different scenarios were tested for different combinations of input variables with filtered and non-filtered instances. The SVR model had higher accuracy, followed by a multilayer perceptron (MLP) and a multilinear regression (MLR) models. This study as for (Paudel et al., 2017), also highlighted that the modeling accuracy increased with the selection of variables and the filtering of instances. With all variables and all instances, the MAPE was of 24.3%, 23.72% and 14.32% while with filtered instances for occupation and OAT, the MAPE was of 5.2%, 1% and 0.06%, for MLR, MLP and SVR, respectively. Li et al. (2009) studied short-term cooling load forecasting of a DeST-simulated ("DeST simulation software, " n.d.) office building in Guangzhou, China. They built a LS-SVM model with Gaussian function kernel using mySVM software kit (mySVM, 2019"mySVM – TU Dortmund, " n.d.) that was compared it to a back propagation neural network (BPNN) implemented with MATLAB ("MATLAB – MathWorks – MATLAB Simulink, " n.d.). Both models used hourly OAT, RH and SR as input variables. Results showed that LS-SVM was more accurate than BPNN with a CV-RMSE of 5.56% and 11.8%, respectively. Fu et al. (2015) proposed a ε-SVR with Gaussian kernel function for an historical record storage and office building in Shanghai, China. They aimed to perform day-ahead prediction of the four major building loads separately and aggregated using day type indicator, OAT, DPT and previous 48 h electricity loads. The ε-SVR was compared with ARIMAX, ANN and reduced-error pruning tree models. The former outperformed all three other techniques and for all five loads. Total load showed a CV-RMSE of 15.2% compared to 22.4%, 27.2% and 22.1% for ARIMAX, reduced-error pruning tree and ANN, respectively. Liu et al. (2015) studied total electricity consumption of a campus building and an office building with energy-saving measures. They implemented a SVR using MATLAB with LIBSVM (Chang & Lin, 2013) and its FarutoUltimate (Li, 2011) toolbox. One month of hourly load data were collected and divided into three weeks for model training and one week for testing. Results showed a higher R^2 of 0.906 for the first and of 0.921 for the second building, compared to 0.822 and 0.843 for an ARIMA model. Zhang, Zhao, Zhang, Fan, and Li (2017) developed support vector regression (SVR) and multiple linear regression (MLR) in Python Scikit-Learn (Scikit-Learn, 2019"Scikit-Learn: machine learning in Python, " n.d.) environment to model and forecast the cooling load of a virtual large office building. One year of hourly cooling load was simulated with EnergyPlus (EnergyPlus, 2019"EnergyPlus, " n.d.) under Miami climate, Florida, United States. Inputs of machine algorithms included OAT, RH, WS, WD, SR and cooling loads at previous 1, 2, 3, 4 and 24 h. Input data were first selected for modeling using distance-correlation-based input method. The basic approaches were also improved with a prediction error correction according to the type of day and hour of the day. Results showed that the error correction improved the MLR forecasting accuracy with a MAPE reduction from 7.10% to 5.51%. However, it was less effective for SVR with a MAPE reduction from 5.70% to 5.66%.

2.3.4. *Artificial neural networks (ANN)*

Artificial neural networks modeling is one of the most applied data-driven BECMF methods. It is a nonlinear machine learning technique inspired from neural networks in the human brain of which it copies the information propagation process in a simplified manner. First, an information coming from a processing element (neuron) is sent with a weight (synapse) through a link (axon) to following processing elements. These combine the information received with other incoming

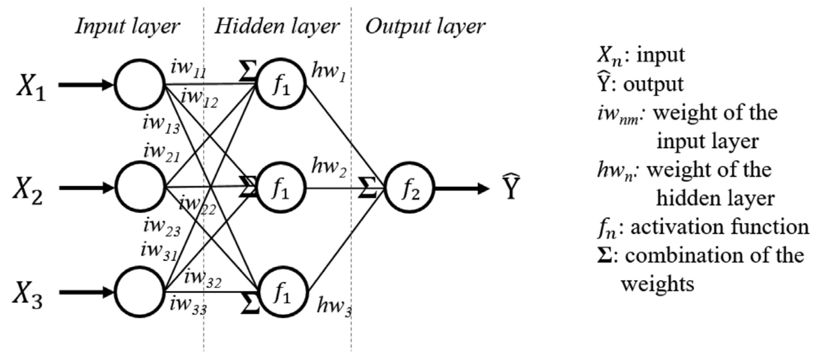

Fig. 4. Schematic of a classical three-layer ANN.

information from other neurons, using a combining function (dendrite). Finally, the combination of weighted information is sent to other receivers, depending on an activation function, also called transfer function (cell body). The process is repeated as many times as the number of layers in the network and until the model accurately fits the data (when the error rate converges) or when the maximum number of iterations is reached (with or without error convergence). The basic form of ANN is composed of three layers: an input layer used either to train the model or to get prediction from input data in the testing phase, an output layer giving the final result(s), and a hidden layer bridging between inputs and outputs (the number and structure can be modified depending on the type of ANN and the needs of the modeling) (Fig. 4). An ANN is usually designed according to three different criteria: (1) the interconnections between the different neurons of different layers – how many neurons from how many layers are communicating in what specific way; (2) the learning method – how the final error is retro-propagated in the network and how it affects the different weights (the error retro-propagation is not illustrated in Fig. 4); (3) the activation functions of each neurons based on the input (Magoulès & Zhao, 2016) (for different layers, the activation functions can be different).

ANN are highly flexible and adaptable models. They enable most forecasting problem solving including with non-linear patterns. They have been applied for short- to long-term forecasting horizon, using any time-step available and any type of input data. However, as most data-driven techniques, one of the major disadvantages of ANN application is the black-box nature of the model with no transparency in terms of physical interpretation. Moreover ANN models are subject to overfitting (Chalal et al., 2016; Massana et al., 2015). They tend to perform very well for one specific dataset but poorly if the same model is applied on another dataset (for training and testing steps for instance). The risk of overfitting increases with the degree of complexity of a model that is usually raised when higher accuracy is targeted. Solutions have been suggested to prevent this issue. A particular attention should be directed structure modification of hidden layers of ANN (Ahmad, Mourshed, & Rezgui, 2017) during the training phase. Using regularization techniques is also possible, especially with large input dataset (L'Heureux, Grolinger, Elyamany, & Capretz, 2017). For example, a pre-selection on input variables can be performed with respect to their impact on the energy consumption to reduce the amount of data processed and the complexity of the model. Indeed, Massana et al. (2015) showed that a pre-selection of the input parameters significantly affected their forecasting results. In their study, higher accuracy was obtained using OAT, occupancy, type of day and hour of the day with filtered instances, resulting in a reduced MAPE of 0.45% for a MLP compared to all non-filtered inputs with a MAPE of 23.7%.

It should be highlighted that ANN is a generic term. Therefore, different specific types of ANN exist to be used in various situations and with different level of complexity and accuracy. The review of the literature presented feed forward neural networks (FFNN) (Jovanović, Sretenović, & Živković, 2015) and back propagation neural networks (BPNN) (Wang, Wang, Li, Zhu, & Zhao, 2014), radial basis function neural networks (RBFNN) (Lee & Ko, 2009), a derivative of FFNN called extreme learning machine (ELM) (Sekhar Roy et al., 2018), adaptive network-based fuzzy interference systems (ANFIS) (Ghanbari, Abbasian-Naghneh, & Hadavandi, 2011) and multi-layer perceptron (MLP) (Amasyali & El-Gohary, 2018) which is the premise of deep learning ANN. More advanced ANN sub-classes can also be found such as recurrent neural networks (RNN) (Mocanu, Nguyen, Gibescu, & Kling, 2016) or nonlinear autoregressive model with exogenous inputs (NARX) (Mena, Rodríguez, Castilla, & Arahal, 2014) and probabilistic entropy-based neural networks (PENN) (Kwok & Lee, 2011).

Mena et al. (2014) implemented an ANN with NARX architecture for a bioclimatic university building in Almeria, Spain. One year and a half of electric load data with 1-min time-step were collected. Input parameters included the type of days, hour of the day, OAT, SR, state of several cooling and heating equipment and combined electric power demand. A full-input model was built and compared to model with limited input (in that case the solar cooling system information were removed). Building energy consumption was forecasted with 1-min-ahead, 1-h-ahead and "infinite" horizons. Kwok and Lee (2011) used a specific type of ANN called probabilistic entropy-based neural network (PENN) to forecast the cooling load of an office building in Hong-Kong. It used hourly weather data including OAT, RH, rainfall, WS, bright sunshine duration and SR. Occupancy was also considered with occupancy rate and internal load. Three models were designed: (1) only weather input parameters, (2) weather inputs and occupancy area and (3) all input parameters. The cooling load was forecasted hourly with a one week-ahead horizon. Results showed that the third model was the most accurate with CV-RMSE (95% lower and upper limits) of 11.41%–17.17% compared to the second and first models with 14.84%–30.09% and 40.38%–52.05%, respectively. Bagnasco, Fresi, Saviozzi, Silvestro, and Vinci (2015) implemented a MLP with MATLAB (MATLAB, 2019"MATLAB – MathWorks – MATLAB Simulink, " n.d.) to forecast total electricity consumption of a hospital complex in Turin, Italy. They used one year of 15-min load data and divided the year of observation into quarters to enhance the prediction force of the model. Input variables were the load of the previous day and of the same-day previous-week, the average of the previous day energy consumption, the type of day, the timestamp and the OAT. MLP was implemented for all four quarters, trained with two months and half of data and tested with fifteen days of data. Forecasting results showed a mean MAPE of 7%. Finally, Biswas, Robinson, and Fumo (2016) proposed two feed-forward artificial neural network (FFNN) with optimization of the convergence speed and accuracy and of the initialization of the algorithms, using MATLAB Neural Network Toolbox (Neural Network Toolbox, 2019"Neural Network Toolbox – MATLAB, " n.d.). These FFNN aimed to predict daily electricity consumption for a research and demonstration residential building located in Texas, USA, using the timestamp, daily mean OAT and daily SR. Data were collected for three months. Models were trained with 70% of the dataset and validated and tested with the remaining 30%. The FFNN performances were assessed with a coefficient of determination, R^2, ranging between 0.871 and 0.878.

Compared to other forecasting techniques, ANN rank among the most accurate. Ahmad et al. (2017) compared FFNN implemented with Python NeuroLab ("NeuroLab 0.3.5, Neural Network Library for Python," n.d.) to random forest decision tree (RF) built with Scikit-Learn ("Scikit-Learn: machine learning in Python, " n.d.) for hourly prediction of combined heating and cooling loads in a hotel building in Madrid, Spain. They showed that the RF actually performed better than the FFNN using all variables (RMSE of 4.66 kWh against 4.72 kWh) while the ANN performed better than the RF with a selection of relevant variables based on their impact on the energy consumption (4.60 kWh against 4.84 kWh). Zhao, Zhong, Zhang, and Su (2016) proposed a comparison of ANN, SVM and ARIMA models to forecast combined heating and cooling energy consumption in Chinese office buildings in Shanghai, Nanjing and Changsha. All models where developed using IBM SPSS Modeler (IBM SPSS Modeler, 2019"IBM SPSS Modeler, " n.d.). Results showed that the ANN performed better than the SVM and that both outperformed the ARIMA with a MAPE of 0.15%–0.11%, 0.21%–0.18% and 0.41%–0.33%, respectively (first–second month testing samples).

With the recent popularization of the technique, deep learning models defined as "computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction" (LeCun, Bengio, & Hinton, 2015) have significantly gained in interest in BECMF studies. Deep learning modeling can be regarded as a technique similar to ANN. However, while standard ANN have only three layers (one input, one hidden and one output layer), deep learning neural networks (DNN) develop more complex algorithm architecture and training schemes. Hence, the number and structure of the hidden layers are adapted depending on their function in the modeling process (Fig. 5). Moreover, the training process is not as straightforward as in conventional ANN. The definition of specific operators to provide more flexibility to the model and achieve higher accuracy.

Thus, ANN with more than three layers can be considered as a DNN. For instance, MLP as implemented in (Bagnasco et al., 2015; Massana et al., 2015; Tso & Yau, 2007) can contain three or more layers. RNN (Mocanu, Nguyen, Gibescu, et al., 2016) also fall in this class of neural networks. Besides, other examples can be found in the literature. Marino, Amarasinghe, and Manic (2016) compared a deep recurrent neural network called long short-term memory. They aimed to forecast the electricity load of a benchmark single residential building. They used the date and time of the targeted prediction together with the previous time-step of power demand. Data were collected for four years: at 1-min time-step for hour-ahead forecasting the DNN gave a RMSE of 0.667 kW; at 1-h time-step for 60-h ahead forecasting, the DNN gave a RMSE of 0.625 kW. Amber et al. (2018) compared a two-layer DNN with genetic programming (GP), ANN, SVR and MLR. ANN and DNN algorithms were developed with TensorFlow library (TensorFlow, 2019"TensorFlow, " n.d.) and SVR was developed with libSVM (Chang & Lin, 2013). Authors used an administrative university building located in London, England as a case-study. They aimed to predict the daily building electricity consumption per unit of surface using daily mean OAT, RH and WS, and a weekday index. All models were trained using three years of data and were tested with one year of data. Results showed that ANN outperformed other techniques with a MAPE of 6%. The multiple regression gave a MAPE of 8.5%, SVM gave 9% and DNN gave a MAPE of 11.15%. Mocanu, Nguyen, Gibescu, et al. (2016) proposed two DNN implemented on MATLAB ("MATLAB – MathWorks – MATLAB Simulink, " n.d.), namely conditional restricted Boltzmann machine (CRBM) and factored CRBM (FCRBM). They used an open dataset with aggregated and sub-metered active power data from a benchmark single residential housing. Energy consumption forecast was performed with different time-steps (1 min to one week) and for different forecasting horizons (15-min-ahead to one-year-ahead), using past time-steps of electric load demand. The two methods were compared with ANN, SVM and RNN. The FCRBM gave the best accuracy for all time-steps and all forecasting horizons. Fan, Wang, Gang, and Li (2019) compared various approaches and model architectures for forecasting with deep recurrent neural networks developed with R software programming ("R: The R Project for Statistical Computing, " n.d.) and Keras package (Keras Documentation, 2019"Keras Documentation, " n.d.). They aimed to forecast the cooling load demand of an educational building in Hong-Kong with half-hour time-step at 24-h-ahead prediction horizon. Input variables included OAT, RH and past time-steps of cooling load power demand, collected over one year. 70% of the dataset were used for model training and 30% were used for testing. Performances of the algorithms ranged between a CV-RMSE of 16.0% for the best performing model, a DNN with gated recurrent unit and direct inference approach, and a CV-RMSE of 38% for the least performing method, a DNN with long short-term memory and recursive inference approach. Finally, Shi et al. (Shi, Liu, & Wei, 2016) proposed the study of a type of RNN called echo state network (ESN) using neuron reservoirs instead of the classical hidden layer. It was implemented for an office building in China to predict hourly electricity consumption and using 6 different reservoir topologies. Four years of breakdown hourly load data were collected, and OAT and building occupancy were used as inputs. Models considered only working days to assess the three main building loads (lights, plugs and AC), total rooms load for four types of rooms and total building electricity demand. Errors for the whole building (CV-RMSE) were between 3.72% and 4.97% depending on the typology of reservoir.

2.4. Combined models

Combined models focus on the optimization of forecasting

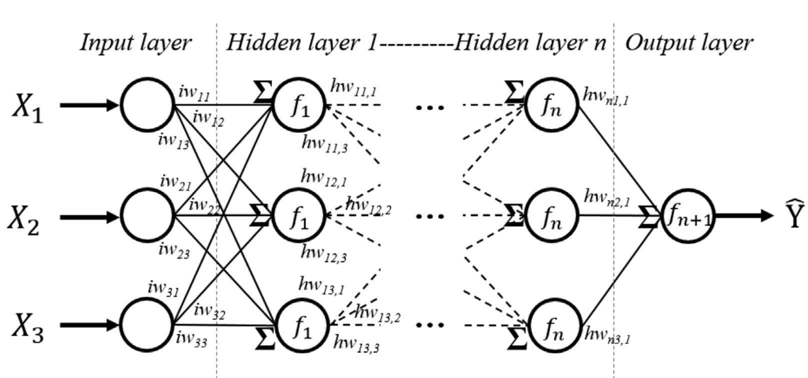

Fig. 5. *Schematic* of a Deep Neural Network with *n* hidden layers.

techniques to improve the prediction accuracy. Combined modeling framework either mixes several single algorithms together (ensemble models) or with optimization methods (improved models).

2.4.1. Ensemble models

Ensemble models are data-driven algorithms designed for forecasting applications. They use a specific framework focusing on the improvement of prediction performance and on the tradeoff of the strengths and weaknesses of predictive algorithms. The ensemble modeling framework comprises two main steps (Fan et al., 2014): (1) several sub-models called "base learners" (for homogenous ensemble models) or "base models" (for heterogeneous ensemble models) are obtained; (2) the comparison of their respective forecasting results is performed, these results are weighted depending on their accuracy and they are combined to generate the optimal output of the ensemble model.

The general ensemble modeling process can further adopt different approaches (Wang & Srinivasan, 2017). A first strategy is called homogenous modeling. It creates sub-samples from the original dataset which are processed through one specific single data-driven technique. The results obtained for each sub-sample, the base learners, are weighted based on their respective prediction performances and are combined into the ensemble model. Two additional specific paths can also be used for homogeneous modeling (Alobaidi, Chebana, & Meguid, 2018): sequential or in-series learning whose classical examples are boosting algorithms (Schapire & Singer, 1999), and which generates base learners sequentially to exploit their interdependence (Fig. 6d); parallel learning which can refer to bagging method (Breiman, 1996) is also called bootstrap aggregation method and generates base learners in parallel to exploit their independence (Fig. 6c). The former more specifically aims to reduce the variance of the estimates of each base learner, while the latter targets bias reduction. Finally, the second strategy for ensemble modeling is called heterogeneous modeling and can refer to stacking techniques (Wolpert, 1992) (Fig. 6b). It uses several different single forecasting algorithms trained on the same dataset (Step 1). The forecasting results from each base model are weighted to give the ensemble model (Step 2). To compare with the different ensemble modeling processes, Fig. 6a illustrates single techniques modeling process when a unique dataset is processed by a unique algorithm to give forecasting results.

Ensemble models have gained in interest for BECMF in the past few years. They provide better prediction accuracy than regular single models and they have been applied to various case-studies with different time-steps and types of data. However, the increasing of prediction accuracy is paid in complexity. Indeed, the framework of ensemble learning is particularly challenging to implement and requires advanced expertise in machine learning. Moreover, it is completely a black-box modeling process, and the prediction horizon in the reviewed studies has been limited to short-term forecasting.

Typical examples of homogenous ensemble learning models are improved decision trees such as random forest (RF) and boosting decision trees (BDT). They have the main advantage to correct the

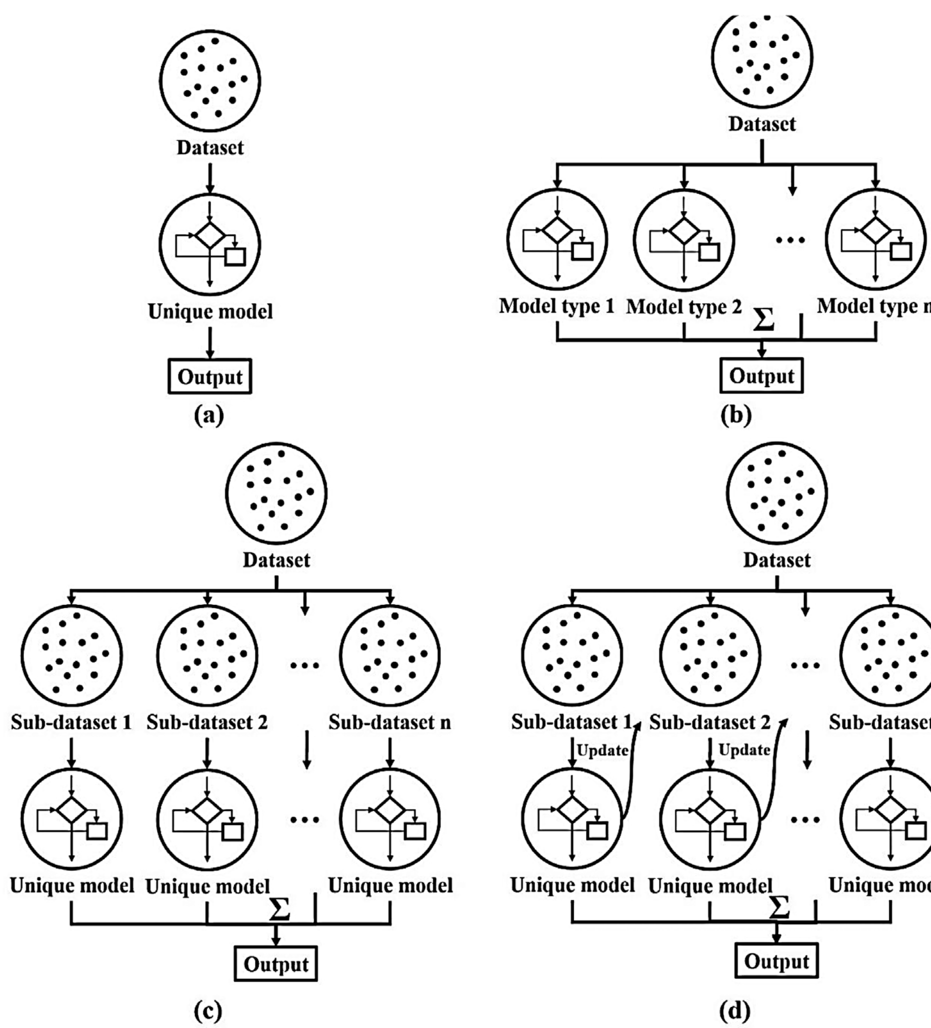

Fig. 6. Comparison between (a) single, (b) heterogeneous ensemble, (c) parallel fomogeneous ensemble and (d) sequential homogeneous ensemble models.

tendency of standard DT to overfit their training set. RF consists in a group of several decision trees whose results are aggregated into one final result (Breiman, 2001). They usually use a two-level randomization strategy. First each trained with a random subset of observations and then each tree node is divided by considering a random subset of variables. BDT, and their modified version called gradient boosted decision trees (GBDT) (Ho, 1995), are based on classification and regression tree (CART) methodology. They use boosting techniques: in the modeling process a sequence of simple decision trees are developed, with each successive tree modeling the residuals of the precedent one. The final model is a weighted additive binary tree model (Elith, Leathwick, & Hastie, 2008). Tsanas and Xifara (2012) compared an iterative reweighted least square regression method to a RF model to predict the heating and cooling load of residential buildings simulated with Ecotect tool. They created an open database including eight passive systems variables (Xifara & Tsanas, n.d.). Results showed that RF outperformed the regression, with a MAE of 0.51 kW/1.42 kW and 2.14 kW/2.21 kW, for heating/cooling loads and for both models, respectively. Wang, Wang, Zeng, Srinivasan, and Ahrentzen (2018) compared a RF to a regression tree and a SVR for the forecasting of electricity consumption in two institutional buildings, including a LEED building, in Florida, USA. Input data were OAT, DPR, RH, pressure, precipitation, WS, SR, estimated occupancy from daily operation and class schedule, time of the day, workday type and day type. Data were collected over a typical year of operation: 80% of the dataset was used for model training and 20% for testing. RF outperformed both the regression tree and SVR for both buildings. For the LEED building, RF had a MAPE of 7.75% compared to 8.04% for SVR and 8.90% for the regression tree. For the second building, RF had a MAPE of 11.93%, compared to 12.21% for SVR and 14.50% for the regression tree. Papadopoulos, Azar, Woon, & Kontokosta (2017) proposed a comparison of three different ensemble DT-based models implemented with Python Scikit-Learn (Scikit-Learn, 2019“Scikit-Learn: machine learning in Python, ” n.d.). They developed a RF, an extra randomized trees and a GBDT to forecast combined heating and cooling loads. Using an open database (Xifara & Tsanas, n.d.) they compared their forecasting method with results from Tsanas and Xifara (Tsanas & Xifara, 2012) (regression and RF models), Chou and Bui (Chou & Bui, 2014) (SVM and ensemble ANN-SVM models) and Castelli et al. (genetic programming) (Castelli, Trujillo, Vanneschi, & Popovič, 2015). The performances of the GBDT improved heating load forecasting performances by 8% to 68% for heating and by 51% to 63% for cooling load compared to the three other studies. Wang, Wang, and Srinivasan (2018) developed a homogeneous ensemble BDT using MATLAB (“MATLAB – MathWorks – MATLAB Simulink, ” n.d.). It was applied on a LEED institutional building in the University of Florida for short-term electricity demand prediction. The model used one year of time-series weather and occupancy data, together with time of the day and day type. To provide higher forecasting performance and because of the different usage period of the building over the year, the dataset was portioned in three sets for summer, fall and spring seasons. Moreover, a method called “compact” modeling was proposed and applied to BDT to measure the influence of the different features and select the most relevant ones for forecasting. Results showed that BDT and compact BDT performed better than CART, with 2.97%/4.62%/4.40%, 2.92%/4.40%/4.48% and 3.08%/5.05%/5.06%, for periods of summer/spring/fall and for the three models respectively. Nevertheless, it was also highlighted that feature selection did not significantly improve forecasting accuracy.

Ensemble models have been implemented with other techniques than based on decision trees. For instance, C. Fan et al. (2014) compared a MLR, an ARIMA, a SVR with Gaussian kernel, a RF, a MLP, a BDT, a MARS, a *k*-NN and an heterogeneous ensemble model combining all eight single models. They were used for daily energy consumption forecasting of a mixed-use (commercial center, offices and hotel) high-rise building in Hong-Kong. Data collection included one year of 15-min time-step building electricity data and one year of daily weather data. The forecasting process also included an outlier detection and elimination method. The ensemble model performed better with a MAPE of 2.32%. For comparison, the SVR was the second-best performing model with a MAPE of 3.11% and the ARIMA was the least performing technique with a MAPE of 5.45%. Alobaidi et al. (2018) developed a homogeneous MLP-FFNN-based ensemble model to forecast day-ahead mean daily household electricity use. It was implemented with a smart energy system in a French household, considering two years and nine months electricity consumption and OAT datasets. The method aimed to improve the dataset resampling in the heterogeneous ensemble modeling process with a two-step strategy to prompt diversity in the model and to better capture the different trends in the dataset. It also targeted the improvement of ensemble model generation, using MLR to combine base learners. The ensemble model was compared to a single ANN and an ANN-based boosting ensemble model, all applied for every days of a week one-by-one. The homogeneous ensemble model outperformed the two latter techniques with a mean weekly MAPE of 14.4%, 18.3% and 15.2%, respectively.

2.4.2. Improved models

In the present study improved models are defined as the combination of single models and optimization techniques. This technique has also been identified by Mat Daut et al. (2017) who referred to them as “hybrid methods”. Optimization methods can include swarm intelligence algorithms such as particle swarm optimization (PSO) (Kennedy & Eberhart, 1995) Genetic algorithms (GA) (Mitchell, 1998), a popular sub-class of evolutionary algorithms or differential evolution (DE) (Storn & Price, 1997).

Zhang et al. (2016) compared the efficiency of swarm intelligence and evolutionary algorithms by applying three techniques: a GA, a differential evolution (DE) algorithm and PSO for the optimization of SVR parameters. All three optimization techniques were implemented with R programming tool (The R Project for Statistical Computing, 2019“R: The R Project for Statistical Computing, ” n.d.) on ε-SVR and ν-SVR models separately. Single models were compared to a weighted combination of both types of SVR and optimized with DE. The different techniques were tested on an institutional building in Singapore for half-hour (10-day dataset, 8:2 training-testing ratio) and daily energy forecasting (260-day dataset, 8:2 training-testing ratio). Inputs only included the past energy demand of the building. The combination of both types of SVR and DE showed the highest forecasting accuracy for both half-hour and daily electricity consumption forecasting, with a MAPE of 3.77% and 5.84%. For ε-SVR with GA, PSO and DE optimization, the MAPE were 6.67%, 5.44% and 5.44% for half-hourly time-step and 5.93%, 5.95%, and 5.95% respectively for daily time-step. For ν-SVR with GA, PSO and DE optimization, the MAPE was 3.77% for all three techniques with half-hourly time-step and 6.37%, 6.36% and 6.36%, respectively for daily time-step. Besides, for energy consumption prediction of residential buildings, Castelli et al. (2015) used genetic programming (GP) with symbolic regression to develop an improved model. Using an opened database of benchmark residential building characteristics (Xifara and Tsanas, 2019Xifara & Tsanas, n.d.), it performed with MAE of 0.51 kW for heating load and of 1.18 kW for cooling load forecasting.

Castelli, Trujillo, Vanneschi, & Popovič (2015) compared a PSO-ANN to simple ANN and GA-ANN (GA was used for the same purpose as PSO). Two databases were used to predict hourly electricity consumption for: 1) a research building from ASHRAE dataset located in USA with four-month data collection including WS, SR, RH and OAT; 2) a campus library located in East China with 100-day data collection including estimated occupancy and daily OAT. For the former, 70% of the data were used for training and 30% for testing. For the latter, 93% of the data were used for training and the remaining 7% for testing. Principal component analysis (PCA) (Jolliffe, 2002) was applied with both case studies for relevant input data selection. PSO-ANN gave better forecasting results than GA-ANN and ANN, with MAPE of 1.6%, 1.9% and 2.2%, respectively, for the ASHRAE database. PSO-ANN also outperformed GA-ANN and ANN when applied on the Chinese library building with a MAPE of 5.9%, 7.1% and 8.0%. In an author study (K. Li et al., 2018), authors developed an optimization strategy called teaching-learning based optimization. It used evolutionary algorithms combined with BPNN to improve convergence speed and forecasting performances. A basic combined model was proposed, together with five improved versions. Models were also coupled with PCA for relevant input variables selection. Algorithms were compared with the PSO-ANN and GA-ANN from the previous study (K. Li et al., 2015). Improved teaching-learning based optimization ANN performed slightly better than the two latter methods and for both buildings.

3. Machine learning approaches for data-driven techniques

The goal of data-driven techniques in BECMF studies is to model the relationships between a combination of inputs and outputs under a specific process. The model is then used to forecast building energy consumption or power load demand. Model outputs are always known since they are the target of the whole study. However, there exist different strategies regarding the utilization of input variables and the extraction of features from an input dataset to train data-driven models. Two main approaches have been used for input variable selection in BECMF: supervised and unsupervised learning. Moreover, other tasks exists such as reinforcement and transfer learning. This section discusses all four machine learning tasks and their application in BECMF studies.

3.1. Supervised and unsupervised learning

Supervised learning can be understood as having input variables of a model that are all labeled, indicating they have been clearly identified before the modeling process. For building energy studies, it would mean that inputs used to assess the energy consumption of a building have been related to their tangible physical meaning (Fan et al., 2017). For instance, the first variable would be outdoor air temperature, the second would be occupancy and so on. Then identified variables can be cleaned and pre-processed to select the most impacting on energy consumption, or directly used as such for model training. Therefore, most data-driven applications presented in the previous section use supervised learning approach.

Unsupervised learning is the second main task of machine learning. It has been largely implemented to make full use of big data collection in buildings and its applications include data analytics, optimization, control, identification of occupants behavior and anomaly detection (Fan, Xiao, Li, & Wang, 2018; Miller, Nagy, & Schlueter, 2018). Contrariwise supervised learning, unsupervised learning uses unlabeled data to discover relevant relationships within a dataset. Thus, a predefinition of specific data types which could influence the modeling process is not implemented. More precisely, for BECMF applications, feature extraction within an input variable dataset is independent from the physical meaning of the variables. Popular techniques for unsupervised feature extraction and classification include *k*-means (Jain, 2008), self-organizing maps (SOM) (Kohonen, 1997), hierarchical clustering algorithms (Rokach & Maimon, 2005) and expectation maximization algorithms (EM) (Dempster, Laird, & Rubin, 1977). The detailed description of these classification methods is out of the scope of this review, but interested reader can refer to Wei et al. (2018) for further details. Using unsupervised approach, Tang, Kusiak, & Wei (2014) investigated the impact of input data clustering on the prediction accuracy of commercial combined heating and cooling demand with short-term forecasting horizon. They first compared several single (SVR, MLP) and ensemble models (RF, boosting tree and MLP-ensemble) to highlight that a MLP-ensemble performed best. Then they further prepared four input data pre-treatment scenarios applied on the initial supervised MLP-ensemble model, and on a season-based model and on two cluster-based models. Results showed an improvement of the forecasting accuracy using clusters, with a MAPE of 3.62%, 3.64%, 3.32% and 3.22% for the single, seasonal-based, first cluster-based and second-cluster based models respectively. Nilashi et al. (2017) implemented both EM and principal component analysis (PCA) with an adaptive network-based fuzzy inference system (ANFIS) for cooling and heating load forecasting of residential buildings. Comparisons were performed between seven forecasting techniques implemented with MATLAB (MATLAB, 2019MATLAB – MathWorks – MATLAB Simulink, " n.d.): SVR, ANFIS, ANN, CART, MLR, PCA-ANFIS and an improved model of MARS with artificial bee colony algorithm (Karaboga & Basturk, 2007). Results showed that the prediction scheme with EM + PCA + ANFIS performed better for both heating and cooling loads, with a MAPE of 1.39% and 2.45% respectively. Finally, comparing the efficiency of both supervised and unsupervised approaches, C. Fan et al. (2017) implemented seven forecasting techniques including MLR, elastic net regression, RF, gradient boosting machine, SVR, extreme gradient boosting decision tree (GBDT) and DNN, using one year of half-hour data. Five input variable datasets were prepared, based on supervised learning techniques. The first dataset included 1) seven variables (OAT, RH and five-time indicators). The other four datasets added 2) the past 24-h cooling load, OAT and RH; 3) the previous time-step of cooling load, OAT and RH; 4) the previous 24-h minimum, maximum, mean and standard deviation of the three variables; 5) the four most dominant frequencies resulting from a discrete Fourier transform and performed on the previous 24-h for each of the three time series. A sixth dataset was also prepared using an unsupervised deep auto-encoder, a DNN, considering the four previous feature extraction methods. The smallest forecasting error was obtained using extreme GBDT with the unsupervised dataset (CV-RMSE of 17.8%). On the opposite, supervised learning approach did not show evident advantages for building cooling load prediction.

3.2. Reinforcement and transfer learning

Reinforcement learning (Busoniu, Ernst, De Schutter, & Babuska, 2011) is an another approach in the field of machine learning and differs from supervised and unsupervised learning. The process is not based on feature extraction in input datasets or on data labeling. It is inspired from psychology and follows a concept of learning through rewarding. In reinforcement learning, an artificial agent figuring a decision-maker is set up in a determined environment and with a specific goal to achieve. The agent performs self-decided actions to reach the predefined goal. For each action it moves within the environment and receives a retro-fed information as a reward to let it know how far away from the final target is its position (Fig. 7). Moreover, each performed action is memorized by the agent to assess its efficiency based on the reward received. Indeed, it aims to maximize the sum of rewards over time to achieve the final state. Hence, agent must be able to learn and decide on a strategy to automatically select a next action without any intervention from a programmer. Therefore, reinforcement learning is not supervised since it relies also on the results of agent-based actions and not only on labeled input data. Reinforcement learning is not unsupervised either as the nature of the reward is already known. Some recent applications for building energy and forecasting studies can be found in the literature. Mocanu, Nguyen, Kling, & Gibescu (2016) implemented two reinforcement algorithms, namely Q-learning and state-action-reward-state-action (SARSA) algorithms, with an unsupervised deep belief network (DBN, a type of DNN) in MATLAB (MATLAB, 2019"MATLAB – MathWorks – MATLAB Simulink, " n.d.). They used seven years of hourly data to forecast energy consumption in a smart-grid context. The database was divided between five different building types and five scenarios were implemented for hour-ahead, day-ahead, week-ahead and month-ahead forecasting with hourly time-step, and year-ahead with weekly time-step. Among the tested model, the Q-learning with DBN obtained the highest accuracy for every scenario and

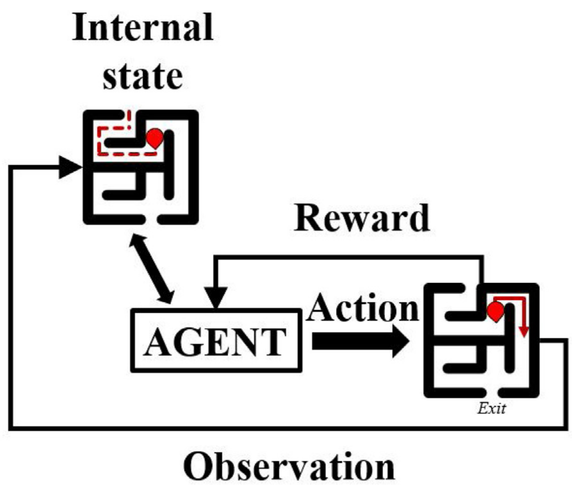

Fig. 7. Reinforcement learning modeling process with reward-based decision making from an artificial agent.

for both two transfer learning strategies. It was followed by DBN + SARSA algorithm, while reinforcement learning algorithms alone performed less accurately.

Furthermore, in this study the particularity of the training method based on reinforcement learning algorithms lied in training models with a dataset from a specific building type to forecast energy consumption for another building with different characteristics. This method is called transfer learning (Pan & Yang, 2010). In the specific case of BECMF, it aims to use and adapt data from specific buildings to train forecasting models implemented for energy demand prediction in other different buildings. For instance, Mocanu, Nguyen, Kling, et al. (2016) used commercial building data as a training set to predict the energy consumption of residential buildings. They also used a similar process to train a model using data for residential buildings without electric heating to predict energy consumption of residential buildings with electric heating. In another study, Ribeiro, Grolinger, ElYamany, Higashino, & Capretz (2018) developed a specific approach for cross-building (transfer learning) building energy forecasting using seasonal and trend adjustment. They selected a case study of four different schools with relatively similar but different energy behavior and climate locations. They proposed two modeling schemes based on their transfer learning method: 1) a training set of one month of data from the target building coupled with twelve months of data collected on the other three buildings; 2) a training set of twelve months of data from the target building coupled with twelve months of data collected on the other three buildings. These schemes were compared with a classical supervised machine learning approach with 3) one month and 4) twelve months of data from one building to forecast the next month same-building energy demand; (5) a training set combining one month from the target building and twelve months of data from the other three. All five schemes were tested with both a SVR and a MLP. Results highlighted the efficiency of the proposed transfer learning method over classical supervised learning and for data-driven both techniques. Nevertheless, it should be noted that the fourth training scheme also produced good forecasted performances almost equivalent to these of the transfer learning method.

4. Input data for data-driven techniques

4.1. Data characteristics

Input data are the driver of all approaches and techniques in the reviewed studies. Input datasets have different characteristics with a direct impact on the modeling and forecasting accuracy. First is the origin of data classified in three main categories with real, simulated and benchmark data accounting for 64%, 20% and 16% of the studied in the present review work. Real data are directly collected from billings, energy meters, environment sensors and onsite surveys. Simulated data are extracted from physics-based models of existing or unexisting buildings, using tools such as EnergyPlus ("EnergyPlus, " n.d.), TRNSYS (TRNSYS, 2019"TRNSYS: Transient System Simulation Tool, " n.d.), DeST ("DeST simulation software, " n.d.), Ecotect ("Ecotect Analysis | Autodesk Knowledge Network, " n.d.)or eQuest ("eQUEST, " n.d.). Benchmark data come from publicly-available datasets provided for researchers to compare forecasting algorithms performances. Benchmark databases used in reviewed studies have been summarized in Table 2.

The features in the dataset can be divided into six main groups: 1) weather data grouping all data related to outdoor conditions; 2) indoor environment to characterize the building indoor conditions; 3) occupancy and occupants behavior; 4) time indicators that deliver information on the building operation and its energy behavior; 5) past time-steps that account for the potential impact of past events on the current and predicted states of the building energy; 6) building characteristics with information on the building passive and active systems. A more detailed list of the different types of input variables found in reviewed studies and falling under these six main categories is provided in Table 3. The number of studies referring to these data and the corresponding techniques implemented are also indicated. From the analysis of this table comes out the predominant usage of specific categories of data. Outdoor air temperature (OAT), outdoor relative humidity (RH) and solar radiations (SR) are considered in thirty-two, nineteen and eighteen different studies respectively. Indeed, these parameters are easily accessible through various open-access or charging weather databases platform ("Iowa Environmental Mesonet, " n.d.; "Meteonorm: Irradiation data for every place on Earth, " n.d.). Moreover, their impact on building energy behavior is well known. Building and equipment characteristics information are crucial as well to accurately model building energy consumption. They can be accessed through onsite surveys, design-related documents or energy standards. On the opposite, some well-identified energy drivers were less considered, such as building occupancy: real occupancy data have been reported in only seven reviewed studies. As a matter of fact challenges lying in occupancy measurements (Yang, Santamouris, & Lee, 2016) often lead to prefer the use of assumed occupancy schedules (used in seven reviewed studies). Thus, because of data availability issues time-related parameters such as the type of day, day of the week indexes and the time of the day are considered to replace other time-dependent measurements such as weather information, occupancy and usages or equipment triggering. Similarly, past load demand data points have been used in twenty-one reviewed articles. Indeed, past load demand provides information on past energy behavior related to time periods, building operation conditions and events similar to the future states of the building energy demand, but for which specific data are unavailable. Finally, it should be mentioned that other parameters were much less used because of their limited impact on building energy consumption such as barometric pressure, cloud coverage or evaporation. However, some of these features (indoor environment measurements or CO_2 levels for instance) could be relevant when considering building comfort which impacts building energy consumption (Allab, Pellegrino, Guo, Nefzaoui, & Kindinis, 2017).

The third input dataset characteristic is the granularity of the time-series. Different time-steps may firstly relate to the need of the studies. Indeed, using a very small time-step such as 1-min provides information on very short and specific events in buildings energy demand patterns. However, such precise information induce a very high variability of the energy demand time series and therefore brings challenges and complexity to obtain accurate forecasting (Mena et al., 2014). On the opposite, a larger granularity, such as weekly or monthly reporting provides information on building design features (Tsanas & Xifara, 2012) and socio-economic related aspects (Yu et al., 2010). However, large

Table 2
Description of benchmark databases used in reviewed studies.

Database name	Building type(s)	Number of buildings/ appliances	Type of data	End-use(s)	Data collection scale	Other	Time-step	Timeframe	Referring studies	Link to the database
BGE: Baltimore Gas and Electricity company	Residential & commercial	5 buildings	Load profiles	Electricity	Building	N/S	1 h	7 years	Mocanu, Nguyen, Kling, et al. (2016)	https://supplier.bge.com/ https://supplier.bge.com/documents/index.asp#electric
Energy Efficiency Dataset	Residential	768 simulated buildings	Benchmark passive system data	Heating & cooling	Building	N/S	N/S	N/S	Tsanas and Xifara (2012)	https://archive.ics.uci.edu/ml/datasets/energy + efficiency
ICER: Irish Commission for Energy Regulation dataset	Residential & commercial	> 6000 buildings	Time series	Electricity, natural gas, water	Building	N/S	1 day	19 months	Valgaev and Kupzog (2016); Valgaev, Kupzog, and Schmeck (2017)	http://www.ucd.i.e./issda/data/commissionforenergyregulationcer/ https://github.com/wwzjustin/CER-Smart-Meter-Project-by-Irish-Social-Science-Data-Archive
iHEPCDS: individual Household Electric Power Consumption Data Set	Residential	1 building	Time series	Electricity	Building & sub-meters	N/S	1 min	4 years	Marino et al. (2016); Mocanu, Nguyen, Gibescu, et al. (2016)	http://archive.ics.uci.edu/ml/datasets/Individual + household + electric + power + consumption
The Great Building Energy Predictor Shootout I	Commercial	1 building	Time series	Electricity, hot water and cold water	Building	OAT, DBT, SR, HR, WS	1 h	6 months	Fan et al. (2019); Li et al. (2015)	N/S

Table 3
Description of the different types of input data in the reviewed studies.

Main input data type	Specific input data	Techniques and reference using the data	Number of studies
Weather/outdoor environment	Outdoor air temperature (OAT)	*AR*: (Fu et al., 2015; Yun et al., 2012) *Regression*: (Amber et al., 2017, 2018; Dagnely et al., 2015; Dong et al., 2016; Fan et al., 2014, 2017; Massa Gray & Schmidt, 2018; Massana et al., 2015; Yun et al., 2012; Zhang et al., 2017) *k-NN*: (Fan et al., 2014; Ma et al., 2017) *DT*: (Fu et al., 2015; Wang, Wang, & Srinivasan, 2018; Wang, Wang, Zeng, et al., 2018; Yu et al., 2010) *SVM*: (Amber et al., 2018; Dagnely et al., 2015; Dong et al., 2016; Fan et al., 2014, 2017; Fu et al., 2015; Li et al., 2009; Massana et al., 2015; Paudel et al., 2017; Ribeiro et al., 2018; Tang et al., 2014; Wang, Wang, Zeng, et al., 2018; Zhao et al., 2016) *ANN*: (Ahmad et al., 2017; Alobaidi et al., 2018; Amber et al., 2018; Bagnasco et al., 2015; Biswas et al., 2016; Dong et al., 2016; Fan et al., 2014; Fu et al., 2015; Kwok & Lee, 2011; Li et al., 2009, 2015; Massana et al., 2015; Mena et al., 2014; Neto & Fiorelli, 2008; Ribeiro et al., 2018; Tang et al., 2014; Yun et al., 2012; Zhao et al., 2016) *DNN*: (Amber et al., 2018; Fan et al., 2019, 2017; Shi et al., 2016) *Ensemble*: (Ahmad et al., 2017; Alobaidi et al., 2018; Fan et al., 2014, 2017; Tang et al., 2014; Wang, Wang, & Srinivasan, 2018; Wang, Wang, Zeng, et al., 2018) *Improved*: (Dong et al., 2016; Li et al., 2015, 2018) *Hybrid*: (Collinge et al., 2016; Dong et al., 2016; Massa Gray & Schmidt, 2018)	32
	Dew point temperature (DPT)	*Regression*: (Fan et al., 2014) *k-NN*: (Fan et al., 2014) *DT*: (Fu et al., 2015; Wang, Wang, & Srinivasan, 2018; Wang, Wang, Zeng, et al., 2018) *SVM*: (Fan et al., 2018; Fu et al., 2015; Ribeiro et al., 2018; Wang, Wang, Zeng, et al., 2018) *ANN*: (Ahmad et al., 2017; Fan et al., 2014; Fu et al., 2015; Ribeiro et al., 2018) *Ensemble*: (Ahmad et al., 2017; Fan et al., 2014; Wang, Wang, & Srinivasan, 2018; Wang, Wang, Zeng, et al., 2018)	7
	Outdoor relative humidity (RH)	*Regression*: (Amber et al., 2017, 2018; Fan et al., 2014, 2017; Massa Gray & Schmidt, 2018; Massana et al., 2015; Yun et al., 2012; Zhang et al., 2017) *k-NN*: (Fan et al., 2014) *DT*: (Wang, Wang, & Srinivasan, 2018; Wang, Wang, Zeng, et al., 2018) *SVM*: (Amber et al., 2018; Fan et al., 2014, 2017; Li et al., 2009; Massana et al., 2015; Ribeiro et al., 2018; Tang et al., 2014; Wang, Wang, Zeng, et al., 2018) *ANN*: (Ahmad et al., 2017; Amber et al., 2018; Fan et al., 2014; Kwok & Lee, 2011; Li et al., 2015; Li et al., 2009; Massana et al., 2015; Neto & Fiorelli, 2008; Ribeiro et al., 2018; Tang et al., 2014; Yun et al., 2012) *DNN*: (Amber et al., 2018; Fan et al., 2019, 2017) *Ensemble*: (Ahmad et al., 2017; Fan et al., 2014, 2017; Tang et al., 2014; Wang, Wang, & Srinivasan, 2018; Wang, Wang, Zeng, et al., 2018) *Improved*: (Li et al., 2015, 2018)	19
	Wind speed (WS)	*Regression*: (Fan et al., 2014; Yun et al., 2012; Zhang et al., 2017) *k-NN*: (Fan et al., 2014) *DT*: (Wang, Wang, & Srinivasan, 2018; Wang, Wang, Zeng, et al., 2018) *SVM*: (Amber et al., 2018; Fan et al., 2014; Tang et al., 2014; Wang, Wang, Zeng, et al., 2018) *ANN*: (Ahmad et al., 2017; Amber et al., 2018; Fan et al., 2014; Kwok & Lee, 2011; Li et al., 2015; Tang et al., 2014; Yun et al., 2012) *Ensemble*: (Ahmad et al., 2017; Fan et al., 2014; Tang et al., 2014; Wang, Wang, & Srinivasan, 2018; Wang, Wang, Zeng, et al., 2018) *Improved*: (Li et al., 2015, 2018)	11
	Wind direction (WD)	*Regression*: (Zhang et al., 2017) *k-NN*: (Fan et al., 2014) *DT*: (Wang, Wang, & Srinivasan, 2018) *SVM*: (Tang et al., 2014) *ANN*: (Tang et al., 2014) *Ensemble*: (Tang et al., 2014; Wang, Wang, Zeng, et al., 2018)	4
	Rain level/rainfalls	*Regression*: (Fan et al., 2014) *k-NN*: (Fan et al., 2014) *DT*: (Wang, Wang, & Srinivasan, 2018; Wang, Wang, Zeng, et al., 2018) *SVM*: (Fan et al., 2014; Wang, Wang, Zeng, et al., 2018) *ANN*: (Fan et al., 2014; Kwok & Lee, 2011) *Ensemble*: (Fan et al., 2014; Wang, Wang, & Srinivasan, 2018; Wang, Wang, Zeng, et al., 2018)	5
	Solar radiation (SR)	*Regression*: (Amber et al., 2017; Dagnely et al., 2015; Dong et al., 2016; Fan et al., 2014; Massana et al., 2015; Yun et al., 2012; Zhang et al., 2017) *k-NN*: (Fan et al., 2014) *DT*: (Wang, Wang, & Srinivasan, 2018; Wang, Wang, Zeng, et al., 2018) *SVM*: (Dagnely et al., 2015; Dong et al., 2016; Fan et al., 2014; Li et al., 2009; Massana et al., 2015; Paudel et al., 2017; Tang et al., 2014; Wang, Wang, Zeng, et al., 2018)	18

(*continued on next page*)

Table 3 (*continued*)

Main input data type	Specific input data	Techniques and reference using the data	Number of studies
		ANN: (Biswas et al., 2016; Dong et al., 2016; Fan et al., 2014; Kwok & Lee, 2011; Li et al., 2009, 2015; Massana et al., 2015; Mena et al., 2014; Neto & Fiorelli, 2008; Tang et al., 2014; Yun et al., 2012) *Ensemble*: (Fan et al., 2014; Tang et al., 2014; Wang, Wang, & Srinivasan, 2018; Wang, Wang, Zeng, et al., 2018) *Improved*: (Dong et al., 2016; Li et al., 2018) *Hybrid*: (Dong et al., 2016)	
	Solar gains	*SVM*: (Paudel et al., 2017)	1
	Bright sunshine duration	*ANN*: (Kwok & Lee, 2011)	1
	Cloud coverage	*Regression*: (Fan et al., 2014) *k-NN*: (Fan et al., 2014) *SVM*: (Fan et al., 2014) *ANN*: (Fan et al., 2014) *Ensemble*: (Fan et al., 2014)	1
	Evaporation	*Regression*: (Fan et al., 2014) *k-NN*: (Fan et al., 2014) *SVM*: (Fan et al., 2014) *ANN*: (Fan et al., 2014) *Ensemble*: (Fan et al., 2014)	1
	CO_2	*SVM*: (Tang et al., 2014) *ANN*: (Tang et al., 2014) *Ensemble*: (Tang et al., 2014)	1
	Barometric pressure	*Regression*: (Fan et al., 2014) *k-NN*: (Fan et al., 2014) *DT*: (Wang, Wang, & Srinivasan, 2018; Wang, Wang, Zeng, et al., 2018) *SVM*: (Fan et al., 2014) *ANN*: (Fan et al., 2014; Tang et al., 2014) *Ensemble*: (Fan et al., 2014; Tang et al., 2014; Wang, Wang, & Srinivasan, 2018; Wang, Wang, Zeng, et al., 2018)	5
	Weather type/category	*k-NN*: (Ma et al., 2017)	1
	Statistical data	*Physics-based*: (Ma et al., 2017; Massa Gray & Schmidt, 2018; Neto & Fiorelli, 2008) *Hybrid*: (Siddharth et al., 2011)	4
Indoor environment	Indoor air temperature (IAT)	*Regression*: (Massana et al., 2015) *SVM*: (Massana et al., 2015) *ANN*: (Massana et al., 2015)	1
	Indoor relative humidity	*Regression*: (Massana et al., 2015) *SVM*: (Massana et al., 2015) *ANN*: (Massana et al., 2015)	1
	Indoor luminosity level	*Regression*: (Massana et al., 2015) *SVM*: (Massana et al., 2015) *ANN*: (Massana et al., 2015)	1
Occupancy	Occupants number/counting (real data)	*Regression*: (Yun et al., 2012) *DT*: (Wang, Wang, & Srinivasan, 2018; Yu et al., 2010) *SVM*: (Paudel et al., 2017) *ANN*: (Ahmad et al., 2017; Kwok & Lee, 2011; Yun et al., 2012) *DNN*: (Shi et al., 2016) *Ensemble*: (Ahmad et al., 2017; Wang, Wang, & Srinivasan, 2018)	7
	Occupancy design data/ estimated data	*AR*: (Newsham & Birt, 2010) *Regression*: (Massana et al., 2015) *DT*: (Wang, Wang, Zeng, et al., 2018) *SVM*: (Massana et al., 2015; Wang, Wang, Zeng, et al., 2018) *ANN*: (Ahmad et al., 2017; Li et al., 2015; Massana et al., 2015) *Ensemble*: (Ahmad et al., 2017; Wang, Wang, Zeng, et al., 2018) *Improved*: (Li et al., 2015, 2018)	7
	Occupancy status	*Regression*: (Dagnely et al., 2015) *SVM*: (Dagnely et al., 2015)	2
Time-related indicators	Time periods	AR: (Yun et al., 2012) Regression: (Lachut et al., 2014) *k*-NN: (Lachut et al., 2014) SVM: (Lachut et al., 2014)	2
	Timestamp	*ANN*: (Biswas et al., 2016)	1
	Time of the day	*AR*: (Yun et al., 2012) *Regression*: (Fan et al., 2014; Lachut et al., 2014; Massa Gray & Schmidt, 2018; Massana et al., 2015) *k-NN*: (Fan et al., 2014; Lachut et al., 2014) *DT*: (Wang, Wang, & Srinivasan, 2018; Wang, Wang, Zeng, et al., 2018) *SVM*: (Fan et al., 2014, 2017; Lachut et al., 2014; Massana et al., 2015; Wang, Wang, Zeng, et al., 2018; Zhao et al., 2016) *ANN*: (Ahmad et al., 2017; Bagnasco et al., 2015; Fan et al., 2014; Massana et al., 2015; Zhao et al., 2016) *DNN*: (Fan et al., 2017; Marino et al., 2016) *Ensemble*: (Ahmad et al., 2017; Fan et al., 2014, 2017; Wang, Wang, & Srinivasan, 2018; Wang, Wang, Zeng, et al., 2018) *Hybrid*: (Massa Gray & Schmidt, 2018)	12

(*continued on next page*)

Table 3 (*continued*)

Main input data type	Specific input data	Techniques and reference using the data	Number of studies
	Day of the week	*Regression*: (Amber et al., 2018; Amber et al., 2017; Fan et al., 2014; Lachut et al., 2014; Massa Gray & Schmidt, 2018; Massana et al., 2015) *k-NN*: (Fan et al., 2014; Lachut et al., 2014) *DT*: (Wang, Wang, & Srinivasan, 2018) *SVM*: (Amber et al., 2018; Fan et al., 2014; Lachut et al., 2014; Massana et al., 2015; Ribeiro et al., 2018; Wang, Wang, Zeng, et al., 2018) *ANN*: (Ahmad et al., 2017; Fan et al., 2014; Lachut et al., 2014; Massana et al., 2015; Ribeiro et al., 2018; Wang, Wang, Zeng, et al., 2018) *DNN*: (Amber et al., 2018; Fan et al., 2017; Marino et al., 2016) *Ensemble*: (Fan et al., 2014, 2017; Wang, Wang, Zeng, et al., 2018) *Hybrid*: (Massa Gray & Schmidt, 2018)	13
	Type of day	*Regression*: (Dagnely et al., 2015; Fan et al., 2014, 2017; Massana et al., 2015) *k-NN*: (Fan et al., 2014; Ma et al., 2017) *DT*: (Fu et al., 2015; Wang, Wang, & Srinivasan, 2018; Wang, Wang, Zeng, et al., 2018) *SVM*: (Dagnely et al., 2015; Fan et al., 2014, 2017; Fu et al., 2015; Massana et al., 2015; Paudel et al., 2017; Wang, Wang, Zeng, et al., 2018; Zhao et al., 2016) *ANN*: (Bagnasco et al., 2015; Fan et al., 2014; Fu et al., 2015; Massana et al., 2015; Mena et al., 2014; Neto & Fiorelli, 2008; Zhao et al., 2016) *Ensemble*: (Fan et al., 2014, 2017; Wang, Wang, & Srinivasan, 2018; Wang, Wang, Zeng, et al., 2018)	13
	Specific day indicator	*Regression*: (Fan et al., 2014, 2017; Massana et al., 2015)	2
	Month of the year	*k-NN*: (Fan et al., 2014) *SVM*: (Fan et al., 2014, 2017; Massana et al., 2015; Ribeiro et al., 2018) *ANN*: (Ahmad et al., 2017; Fan et al., 2014; Massana et al., 2015; Ribeiro et al., 2018) *DNN*: (Fan et al., 2017) *Ensemble*: (Ahmad et al., 2017; Fan et al., 2014, 2017; Wang, Wang, & Srinivasan, 2018; Wang, Wang, Zeng, et al., 2018)	8
	Year	*SVM*: (Ribeiro et al., 2018) *ANN*: (Ribeiro et al., 2018)	1
Past time-steps/data points	Previous power demand/ energy consumption	*AR*: (Dagnely et al., 2015; Fan et al., 2014; Fu et al., 2015; Lachut et al., 2014; Liu et al., 2015; Newsham & Birt, 2010; Yun et al., 2012; Zhao et al., 2016) *Regression*: (Dong et al., 2016; Fan et al., 2014; Lachut et al., 2014) *k-NN*: (Lachut et al., 2014; Ma et al., 2017; Valgaev & Kupzog, 2016; Wahid & Kim, 2016) *SVM*: (Dong et al., 2016; Fan et al., 2017; Lachut et al., 2014; Liu et al., 2015; Mocanu, Nguyen, Gibescu, et al., 2016; Chaobo Zhang et al., 2017) *ANN*: (Alobaidi et al., 2018; Bagnasco et al., 2015; Dong et al., 2016; Kwok & Lee, 2011; Mena et al., 2014; Mocanu, Nguyen, Gibescu, et al., 2016; Yun et al., 2012) *DNN*: (Fan et al., 2019; Fan et al., 2017; Mocanu, Nguyen, Gibescu, et al., 2016; Mocanu, Nguyen, Kling, et al., 2016) *Ensemble*: (Alobaidi et al., 2018; Fan et al., 2017; Zhang et al., 2016) *Improved*: (Dong et al., 2016; Zhang et al., 2016) *Hybrid*: (Dong et al., 2016)	21
	Previous OAT	*Regression*: (Fan et al., 2017) *SVM*: (Fan et al., 2017; Paudel et al., 2017) *DNN*: (Fan et al., 2017) *Ensemble*: (Fan et al., 2017)	2
	Previous RH	*Regression*: (Fan et al., 2017) *SVM*: (Fan et al., 2017) *DNN*: (Fan et al., 2017) *Ensemble*: (Fan et al., 2017)	1
	Previous SR	*SVM*: (Paudel et al., 2017)	1
	Previous solar gains	*SVM*: (Paudel et al., 2017)	1
Mathematical characteristics	Minimum, maximum and/or mean of time series	*Regression*: (Fan et al., 2017) *SVM*: (Fan et al., 2017; Ribeiro et al., 2018) *ANN*: (Ribeiro et al., 2018) *DNN*: (Fan et al., 2017) *Ensemble*: (Fan et al., 2017)	2
	Fourier transform	*Regression*: (Fan et al., 2017) *SVM*: (Fan et al., 2017) *DNN*: (Fan et al., 2017) *Ensemble*: (Fan et al., 2017)	1
Deep learning-based time series		*Regression*: (Fan et al., 2017) *SVM*: (Fan et al., 2017) *DNN*: (Fan et al., 2017) *Ensemble*: (Fan et al., 2017)	1
Building characteristics and operation information	Passive system	*Regression*: (Amber et al., 2017; Chou & Bui, 2014; Dong et al., 2016; Massa Gray & Schmidt, 2018; Nilashi et al., 2017; Pulido-Arcas et al., 2016; Sekhar Roy et al., 2018; Tsanas & Xifara, 2012; Tso & Yau, 2007) *DT*: (Chou & Bui, 2014; Nilashi et al., 2017; Tso & Yau, 2007; Yu et al., 2010) *SVM*: (Chou & Bui, 2014; Dong et al., 2016; Nilashi et al., 2017) *ANN*: (Chou & Bui, 2014; Dong et al., 2016; Nilashi et al., 2017; Sekhar Roy et al.,	14

(*continued on next page*)

Table 3 (*continued*)

Main input data type	Specific input data	Techniques and reference using the data	Number of studies
		2018; Tso & Yau, 2007) *Ensemble*: (Chou & Bui, 2014; Papadopoulos et al., 2017; Tsanas & Xifara, 2012) *Improved*: (Castelli et al., 2015; Dong et al., 2016; Nilashi et al., 2017) *Physics-based*: (Ma et al., 2017; Massa Gray & Schmidt, 2018; Neto & Fiorelli, 2008) *Hybrid*: (Dong et al., 2016; Massa Gray & Schmidt, 2018; Siddharth et al., 2011)	
	Active systems	*Regression*: (Pulido-Arcas et al., 2016) *SVM*: (Tang et al., 2014) *ANN*: (Mena et al., 2014; Tang et al., 2014) *Ensemble*: (Tang et al., 2014) *Physics-based*: (Ma et al., 2017; Massa Gray & Schmidt, 2018; Neto & Fiorelli, 2008) *Hybrid*: (Collinge et al., 2016; Dong et al., 2016; Massa Gray & Schmidt, 2018; Siddharth et al., 2011)	9

time-steps are not suitable for forecasting applications related to day-to-day building energy management. In the reviewed papers, the times-step ranged from 1-min to annual with the following repartition: (1) three studies with 1-min time-step, (2) one study with 5-min time-step, (3) three studies with 15-min time-step, (4) three studies with half-hour time-step, (5) twenty-eight studies with hourly time-step, (6) five studies with daily time-step, (7) three studies with weekly time-step and (8) nine studies with annual time-step. The details of the studies with the corresponding time-steps and other characteristics is available in Appendix A.

Finally, the fourth characteristics of input dataset is the amount of data used for training, validation and testing of the forecasting algorithms. Among the reviewed studies most of the database contained between one and six months of data (45%). 7% (3 studies) used less than a month of data, 7% of the database used between 6 months and 1 year of data, 24% used between 1 year and 2 years and 17% used more than 2 years of data. Moreover, it should be noted that specific studies such as Yu et al. (2010) or those referring to Tsanas and Xifara (2012)'s database (Xifara and Tsanas, 2019Xifara & Tsanas, n.d.) did not used time series as inputs of their models. They referred to the amount of data they used as test cases or sets of data but without timeframe indication. Training, validation and testing ratios were investigated. Most datasets in reviewed studies, with a share of 65%, used between 50% and 90% of their data for training or training and validation combined (therefore between 10% and 50% for testing). Then 20% of the datasets were split with more than 90% dedicated to training. Only in (Massa Gray & Schmidt, 2018) were more data used for testing than for training (10% of the total number of datasets) and in (Ribeiro et al., 2018; D. Zhao et al., 2016) were the data divided with a 50%–50% training–testing ratio. Frequently used ratios were 70%–30% in six studies and 80%–20% and 75%–25% in five studies. Finally, when the validation step was dissociated from the training step most data were dedicated to algorithm training with a training–validation–testing ratio of 70%–15%–15% in (Fan et al., 2017), 62%–17%–21% in (Mena et al., 2014) and 80%–10%–10% in (Chaobo Zhang et al., 2017). This highlights that if the validation step has a different purpose than the training step and is supposed to use a different dataset, it is not always clearly dissociated or even mentioned at all.

4.2. Data pre-processing

As part of the modeling process, input data pre-processing is a very important. It involves a verification of the input data quality and eventually an optimization of the types of inputs, time frames and time-steps selected. Hence many reviewed studies used common methods or developed specific ones for data pre-processing, as it directly impacts on the forecasting results, their accuracy and reliability.

Data pre-processing relies on two sub-processes: data cleaning and input data selection. Data cleaning is a mandatory step to remove all poor-quality information such as missing data, monitoring issues and outliers that depict very unusual energy behaviors. It can be done manually or automatically. For instance, Fan et al. (2014) who developed an automated outlier detection method.

Input data selection is not mandatory. It aims to select specific combinations of inputs to retrieve the most influential energy drivers in order to enhance forecasting performances and reduce calculations complexity. Several approaches have been highlighted during the review work. Input data selection can first relate to the selection of an adapted forecasting time-step (Mocanu, Nguyen, Gibescu, et al., 2016; Mocanu, Nguyen, Kling, et al., 2016) or an adapted training–validation–testing ratio (Massa Gray & Schmidt, 2018; Wahid ö Kim, 2016). Also, a common method is the manual selection of different combinations of inputs (Dagnely et al., 2015; Fan et al., 2017; Neto & Fiorelli, 2008; Yun et al., 2012) and the comparison of the forecasting results. Indeed, some parameters may not have any direct impact on building energy consumption. Then, to use them for model training can result in lower forecasting accuracy and in overfitting the models. However, it should be highlighted that in relevant data pre-selection is not always effective (Massa Gray & Schmidt, 2018; Wang, Wang, ö Srinivasan, 2018). Input dataset can also be pre-processed using original dataset manually divided into different periods such as weekdays/weekends (Newsham & Birt, 2010), days with specific types of weather (Ma et al., 2017; Mena et al., 2014) or seasons (Tang et al., 2014). By doing so, specific building operating conditions are isolated and sub-models can be developed for each or specific periods depending on the focus of the studies. The same idea can be achieved with clustering algorithms that automatically identify the trends in the building energy behavior. Extracted trends can then be associated with different usages or types of days (Tang et al., 2014). These algorithms can be either supervised when the clustering is based on user-defined features, or unsupervised if the features are extracted by using mathematical operators and metrics (Toffanin, 2016). Then, it relates to machine learning tasks described in the previous section. Finally, a pre-selection of features can be implemented, using sensitivity analyses (Kristensen & Petersen, 2016), principal component analysis (PCA) (K. Li et al., 2018; Nilashi et al., 2017) or other specific methods (Deb & Lee, 2018; Massana et al., 2015; Paudel et al., 2017; Wang, Wang, Zeng, et al., 2018; Chaobo Zhang et al., 2017).

5. Discussions

5.1. Building energy modeling and forecasting targets

5.1.1. Building typologies

Reviewed studies have focused on a variety of building typologies. These typologies have been classified into three main categories: commercial buildings including educational buildings, residential buildings and mixed usages buildings representing 66%, 30% and 4% of the buildings in reviewed studies. Table 4 provides further description of the different building typologies.

Table 4
Summary of the different building types in the reviewed studies.

Main building typology	Specific building type	Number of studies considering the specific building types	
Commercial	Hotel	1	40%
	Hospital	1	
	Office (real)	10	
	Office (simulated)	5	
	Research	2	
	Library	2	
	N/S	2	
Educational	Academic (classrooms & laboratories)	7	26%
	Administrative	3	
	Institutional (classrooms, offices and laboratories)	1	
	Research center (offices and laboratories)	4	
Residential	Single family housings (real)	5	30%
	Single family housings (simulated)	1	
	Multifamily building (real)	2	
	Multifamily building (simulated)	7	
	Research/demonstration	1	
	N/S	1	
Mixed-use	Residential multifamily building and commercial office (simulated)	1	4%
	Mixed-use (commercial center, office, hotel)	1	

Therefore, it clearly appears a lack of studies on residential buildings and mixed-use buildings. The over-representation of educational buildings in probably due to data availability since campus buildings can be more easily instrumented and monitored for research purposes. On the opposite, sensor-based data for residential building are more difficult to obtain. Indeed, half of residential buildings considered were either simulated buildings using benchmark dataset (Xifara & Tsanas, n.d.) or unoccupied residential buildings used as research demonstration (Biswas et al., 2016). This clearly highlights an insufficient number of monitored residential buildings and a lack of data for this building typology. Another reason could also be the complexity of residential energy demand forecasting. Because of the smaller size of residential buildings, the relatively small number of energy consuming appliances and the complexity to account for occupants' behavior in energy models, individual dwelling energy demand is more difficult to assess than for commercial buildings. Occupant's behavior has a significant impact on building energy consumption (Pisello & Asdrubali, 2014) and a higher variability in residential buildings than in commercial or large office buildings (Xu et al., 2012). Furthermore, advanced occupant behavior modeling has been lacking in BECMF studies as reported in Table 3 and real occupancy data are hardly accessible. Thus, they are replaced by predefined occupancy scenarios resulting in even larger uncertainties in building energy forecasting (Azar & Menassa, 2012). Despite obvious challenges accurate residential and mixed-used energy modeling and forecasting is needed. Indeed, residential energy consumption represented 25.7% of the European final energy consumption in 2016 against 13.5% for commercial buildings ("European Environment Agency – Final energy consumption by sector and fuel, " n.d.). Therefore, it holds a large share of energy consumption with large potential energy savings.

Regarding lack of studies on mixed-use buildings, similar problems as those faced for residential energy exist. Furthermore, combining different building types induces a larger diversity of appliances, behaviors and demand profiles which increases the modeling complexity (Choi, Cho, & Kim, 2012). Consequently, addressing such cases requires an even larger amount of data which are not easily available. When unavailable, data are replaced by assumptions at the expense of realistic case studies (Valgaev & Kupzog, 2016). Nevertheless, mixed-use building energy studies are essential as well since this building typology is gaining ground in some energy-consuming and urbanized countries (Choi et al., 2012; Woo and Cho, 2018). Hence, future studies should 1) provide residential buildings case studies with integration of retrofit impact assessment together with enhanced human behavior capturing and 2) focus on providing a better mixed-use buildings understanding.

5.1.2. Energy end-uses

Most of reviewed studies (52%) focus on overall energy forecasting while a smaller part (46%) focus on cooling and heating load demand prediction (separated or combined) (details on studies targeting heating, cooling or combined cooling and heating loads are provided in Table 5). Then only 4% of the reviewed studies targeted other loads. Newsham and Birt (2010) assessed "occupancy-related" loads with combined lightings and plugs electricity demand. Shi et al. (2016) proposed forecasting models for lightings, AC and plug loads separately and combined at the building scale. Therefore, there is a lack of studies on other loads than thermal loads and total building energy demand. However, these loads such as lighting and plug loads represent more than 19% of residential energy consumption in Europe ("Energy consumption in households – Statistics Explained, " n.d.). Consequently, they hold a significant share of energy demand and of consequent potential energy savings (Ghadi, Rasul, & Khan, 2017). Even more so that both lighting and equipment on plugs are a significant internal heat source in buildings and they directly impact on cooling load demand (Dong et al., 2016).

The main reason for this gap could be related to the "occupancy-based" nature of lighting and plugs energy consumption (Newsham & Birt, 2010). Indeed, energy standards require specific amounts of lighting for optimal operating conditions in offices and for activities in residential buildings (ASHRAE, 2013). Thus, for obvious energy conservation measures, lighting might be automated to detect occupancy (Kandasamy, Karunagaran, Spanos, Tseng, & Soong, 2018) or at least for non-zero occupancy. Similarly, most equipment usage such as office, electronic and cooking equipment usage are also occupancy related. Moreover, as highlighted in Section 4.1 few occupancy and behavior-related data have been used in the reviewed studies, which induces a lack of information for other building loads forecasting. Hence, further studies should focus on lighting and plug energy demand with a better accounting for occupancy and occupants' behavior information.

5.1.3. Forecasting horizon

Forecasting horizons can be divided between short-term, medium-term and long-term (Mocanu, Nguyen, Kling, et al., 2016; Yalcinoz and Eminoglu, 2005). They aim at different purposes for energy management and savings. Short-term horizon can be defined as forecasting from the next minute to the next week. It is essential for real-time management of building energy systems (Fan et al., 2019) such as HVAC systems or to manage local energy generation, storage and provision (Bouzerdoum, Mellit, & Massi Pavan, 2013). It represents 41% of the simulations in studies covered by the present review. Medium-term

Table 5
Description of the different end-uses targeted in the reviewed studies.

End-use	Techniques implemented	Number of studies
Overall energy	*AR*: (Dagnely et al., 2015; Fan et al., 2014; Fu et al., 2015; Lachut et al., 2014; Liu et al., 2015) *Regression*: (Amber et al., 2017, 2018; Dagnely et al., 2015; Dong et al., 2016; Fan et al., 2014; Lachut et al., 2014; Massana et al., 2015; Pulido-Arcas et al., 2016) *k-NN*: (Fan et al., 2014; Lachut et al., 2014; Valgaev & Kupzog, 2016; Wahid & Kim, 2016) *DT*: (Fu et al., 2015; Tso & Yau, 2007; Wang, Wang, & Srinivasan, 2018; Wang, Wang, Zeng, et al., 2018) *SVM*: (Amber et al., 2018; Dagnely et al., 2015; Dong et al., 2016; Fan et al., 2014; Fu et al., 2015; Lachut et al., 2014; Liu et al., 2015; Massana et al., 2015; Mocanu, Nguyen, Gibescu, et al., 2016; Ribeiro et al., 2018) *ANN*: (Alobaidi et al., 2018; Amber et al., 2018; Bagnasco et al., 2015; Biswas et al., 2016; Dong et al., 2016; Fan et al., 2014; Fu et al., 2015; Li et al., 2015; Massana et al., 2015; Mena et al., 2014; Mocanu, Nguyen, Gibescu, et al., 2016; Neto & Fiorelli, 2008; Ribeiro et al., 2018; Tso & Yau, 2007) *DNN*: (Amber et al., 2018; Marino et al., 2016; Mocanu, Nguyen, Gibescu, et al., 2016; Shi et al., 2016) *Ensemble*: (Alobaidi et al., 2018; Fan et al., 2014; Wang et al., 2018a; Wang, Wang, Zeng, et al., 2018) *Improved*: (Dong et al., 2016; Li et al., 2015, 2018; Zhang et al., 2016) *Hybrid*: (Dong et al., 2016; Siddharth et al., 2011)	28
Cooling load	*AR*: (Yun et al., 2012) *Regression*: (Chou & Bui, 2014; Fan et al., 2017; Nilashi et al., 2017; Sekhar Roy et al., 2018; Tsanas & Xifara, 2012; Yun et al., 2012; Chaobo Zhang et al., 2017) *DT*: (Chou & Bui, 2014; Nilashi et al., 2017) *SVM*: (Chou & Bui, 2014; Fan et al., 2017; Li et al., 2009; Nilashi et al., 2017; Zhang et al., 2017) *ANN*: (Chou & Bui, 2014; Kwok & Lee, 2011; Li et al., 2009; Nilashi et al., 2017; Sekhar Roy et al., 2018; Yun et al., 2012) *DNN*: (Fan et al., 2019; Fan et al., 2017) *Ensemble*: (Chou & Bui, 2014; Fan et al., 2017; Papadopoulos et al., 2017; Sekhar Roy et al., 2018; Tsanas & Xifara, 2012) *Improved*: (Castelli et al., 2015; Nilashi et al., 2017)	12
Heating load	*AR*: (Yun et al., 2012) *Regression*: (Chou & Bui, 2014; Nilashi et al., 2017; Sekhar Roy et al., 2018; Tsanas & Xifara, 2012; Yun et al., 2012) *DT*: (Chou & Bui, 2014; Nilashi et al., 2017) *SVM*: (Chou & Bui, 2014; Nilashi et al., 2017) *ANN*: (Chou & Bui, 2014; Nilashi et al., 2017; Sekhar Roy et al., 2018; Yun et al., 2012) *Ensemble*: (Chou & Bui, 2014; Papadopoulos et al., 2017; Tsanas & Xifara, 2012) *Improved*: (Castelli et al., 2015; Nilashi et al., 2017)	7
Combined heating and cooling loads	*AR*: (Zhao et al., 2016) *Regression*: (Massa Gray & Schmidt, 2018) *k-NN*: (Ma et al., 2017) *SVM*: (Paudel et al., 2017; Tang et al., 2014; Zhao et al., 2016) *ANN*: (Ahmad et al., 2017; Tang et al., 2014; Zhao et al., 2016) *Ensemble*: (Ahmad et al., 2017; Tang et al., 2014) *Hybrid*: (Collinge et al., 2016; Massa Gray & Schmidt, 2018)	7
Other loads	*AR*: (Newsham & Birt, 2010) *DNN*: (Shi et al., 2016)	2

prediction of energy consumption, from one week to several months ahead, focuses on energy storage systems management and maintenance planning of building equipment (Rahman, Srikumar, & Smith, 2018). This horizon is considered in 35% of reported studies.

Finally, long-term horizon provides information on the next year and over which is used for design and planning tasks (Rahman et al., 2018). Thus, as much as short- and medium-term predictions, it is essential, to serve long-term sustainability strategies in built environment. However, long-term forecasting is targeted by less than 25% of reviewed studies. This is probably due to data availability problems. Indeed, the longer the forecasting horizon the larger the diversity of demand patterns (Mocanu, Nguyen, Gibescu, et al., 2016). Accounting for a larger diversity requires more data, collected over longer periods of time. However, as explained in Section 4.1 a majority (59%) of studies relied on less than one-year measurement campaigns while only 17% used data collected for more than two years and thus rely on relevant training sets for long-term prediction. In addition, building energy systems can exhibit nonlinear behaviors (Li & Wen, 2015). If nonlinearities can be handled by most forecasting algorithms on a short-term basis, it becomes much harder for long-term forecasting horizon. Finally, the forecasting time-step also has a non-negligible impact (Mena et al., 2014). Among the long-term forecasting studies, 53% used annual time-step, compared to 12% for weekly and daily time-step, 18% for hourly time-step and 6% for 1-min time-step. Thus, the smaller the time-step the more challenging the long-term forecasting.

5.2. Building energy modeling and forecasting data-driven methods

In the present review, different building energy modeling and forecasting methods have been presented and described, focusing on data-driven techniques. These algorithms, even for basic methods, can achieve relatively high forecasting accuracy while requiring less expertise regarding the various building energy behavior characteristics than traditional physics-based modeling process (Ma et al., 2017; Neto ö Fiorelli, 2008). Thus, they are currently the main research focus in BECMF (Ahmad et al., 2018; Amasyali ö El-Gohary, 2018; Deb et al., 2017; Mat Daut et al., 2017; Wang & Srinivasan, 2017; Wei et al., 2018; Yildiz et al., 2017).

Among the techniques described, classical approaches with autoregressive and regression models are quite popular because of their relative implementation simplicity and good forecasting performance. They are often used as a comparison basis for the implementation of more advanced algorithms (Fan et al., 2014, 2017). Classification-based methods with DT and *k*-NN are intuitive and of significant prediction force (Chou & Bui, 2014; Ma et al., 2017; Wahid & Kim, 2016). SVM and ANN are among the best performing and the most implemented data-driven single techniques for building energy forecasting studies, as highlighted in Table 1. They can be used as a support tool for more advanced modeling process such as ensemble (Alobaidi et al., 2018) and improved models (F. Zhang et al., 2016). Furthermore, recent machine learning developments have been implemented for BECMF with deep neural networks (Amber et al., 2018), unsupervised learning

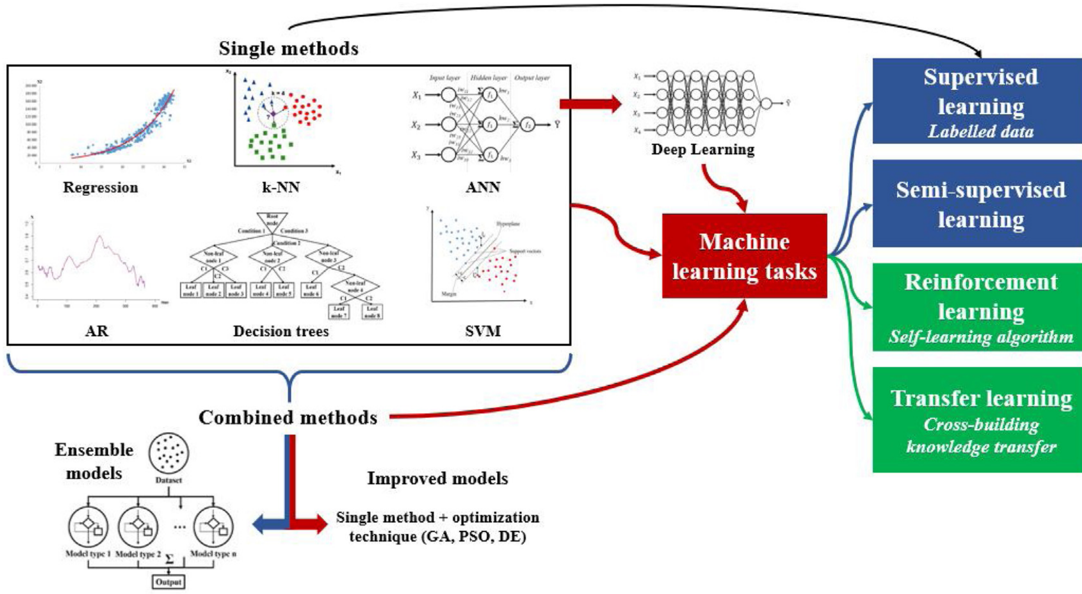

Fig. 8. Illustration of data-driven methods covered in the review and the range of machine learning improvements available.

(Fan et al., 2017), reinforcement learning (Mocanu, Nguyen, Kling, et al., 2016) and transfer learning (Ribeiro et al., 2018) (Fig. 8). Such improvements in machine learning based forecasting algorithms lead the way to less operator-dependent and more versatile algorithms in terms of data usage, with much higher prediction accuracy.

Also, machine learning techniques benefit from widespread modeling tools, libraries and packages available that encompass various pre-embedded functions which makes implementation easier. Some of the most used are Python ("Python, " n.d.), R programming (The R Project for Statistical Computing, 2019"R: The R Project for Statistical Computing, " n.d.), MATLAB ("MATLAB – MathWorks – MATLAB Simulink, " n.d.), IBM-SPSS Modeler ("IBM SPSS Modeler, " n.d.) and Statistics ("IBM SPSS Statistics, " n.d.), Weka ("Weka 3 – Data Mining with Open Source Machine Learning Software in Java, " n.d.) and mySVM software (mySVM, 2019"mySVM – TU Dortmund, " n.d.). Details on the different packages used for machine learning techniques are provided in Table 6.

5.3. Limitations of data-driven techniques: toward grey-box modeling

Despite great flexibility and good forecasting performances, data-driven algorithms show several limitations. First, they rely on large quantities of data that must be representative of the different operating conditions of the building. Otherwise, they would only capture specific patterns lack generality. This is a common problem in machine learning techniques, with overfitting (Chalal et al., 2016), and the reason why a particular attention is to be paid to training, validation and testing data samples independence. Nevertheless, such constraint is often limited by manual data pre-processing as presented in Section 4.1, and the lack of information and data availability on important energy drivers (Section 4.2). The first problem can be tackled through more advanced or different machine learning techniques with unsupervised, reinforcement and transfer learning (Sections 3.1 and 3.2). The latter cabs be counter-balanced by optimizing input data, using time-related parameters for example instead of physical variables.

Nevertheless, data-driven approaches remain completely black-box methods. Contrarily to physics-based models, also called white-box models, they do not provide transparency on the link between inputs and the final forecasted building energy consumption. However, physics-based modeling is also a very complex process, requiring advanced knowledge and information on building to be modeled with high uncertainties among key energy-drivers. Therefore, hybrid techniques combining both white- and black-box models a have been the focus of recent studies.

Hybrid modeling presents two main orientations. A data-driven method is used to optimize specific parameters of a white-box models. For instance, Siddharth, Ramakrishna, Geetha, and Sivasubramaniam (2011) used a genetic algorithm to quickly and realistically create sets of specific input parameters identified as key energy-drivers for a white-box model. They aimed to assess hourly total building energy consumption over a year. Then, for satisfactory results a non-linear regression model was implemented between the selected system variables and the annual energy consumption. It showed very satisfactory coefficients of determination. In the case of (Massa Gray & Schmidt, 2018), a Gaussian process was combined with a RC-lumped model (Resistance Capacitance) to predict and adjust error of the physics-based model. It showed higher forecasting performances than with the Gaussian process or the RC-model alone. Another way to combine data-driven and physics-based models consists in replacing parts of the physics-based model with machine learning algorithms, for energy equipment load demand simulation for instance. This is the case for (Collinge, DeBlois, Landis, Schaefer, & Bilec, 2016) who used sequential linear regression to assess cooling and heating loads of an HVAC system set in an EnergyPlus (EnergyPlus, 2019"EnergyPlus, " n.d.) physics-based environment. Similarly, Dong et al. (2016) proposed an hybrid strategy to predict the total electricity consumption of residential buildings by dividing the electric loads between AC and non-AC consumption. Non-AC electricity consumption was forecasted using a LS-SVM model, from which internal heat gain variations were deduced. Heat gains together with weather information they were input in a 2R-1C lumped model to calculate the different building zones temperature. Zone temperatures results were input in an AC regression model to further AC cooling power consumption. Finally, both data-driven-based non-AC electricity consumption and hybrid-based AC electricity consumption were summed up to forecast the total building electricity consumption. A comparison between the grey-box model and data-driven algorithms such as FFNN, SVR, LS-SVM, Gaussian mixture model and Gaussian process regression showed a significant improvement of forecasting performances with the proposed hybrid model.

5.4. Occupants' behavior impact on building energy efficiency

In spite of the significant progress in data-driven modeling in recent

Table 6
Summary of the software and packages to develop data-driven studies in reviewed studies.

Software and packages		Referring studies
IBM SPSS (“IBM SPSS Modeler, ” n.d.; “IBM SPSS Statistics, ” n.d.)		*AR*: (Newsham & Birt, 2010; Zhao et al., 2016) *Regression*: (Amber et al., 2017; Chou & Bui, 2014) *DT*: (Chou & Bui, 2014) *SVM*: (Chou & Bui, 2014; Zhao et al., 2016) *ANN*: (Chou & Bui, 2014; Zhao et al., 2016) *Ensemble*: (Chou & Bui, 2014) *Improved*: (Li et al., 2018)
MATLAB (“MATLAB – MathWorks – MATLAB Simulink, ” n.d.)	N/S	*Regression*: (Nilashi et al., 2017) *k-NN*: (Wahid & Kim, 2016) *DT*: (Nilashi et al., 2017; Wang, Wang, & Srinivasan, 2018) *SVM*: (Nilashi et al., 2017; Wang, Wang, & Srinivasan, 2018) *ANN*: (Bagnasco et al., 2015; Li et al., 2009; Nilashi et al., 2017) *DNN*: (Mocanu, Nguyen, Kling, et al., 2016) *Ensemble*: (Wang, Wang, & Srinivasan, 2018) *Improved*: (Nilashi et al., 2017) *Hybrid*: (Massa Gray & Schmidt, 2018)
	LibSVM (Chang & Lin, 2013)	*SVM*: (Amber et al., 2018; Dong et al., 2016; Mocanu, Nguyen, Gibescu, et al., 2016; Paudel et al., 2017)
	LibSVM + FaroUltimate (Li, 2011)	*SVM*: (Liu et al., 2015)
	Neural Network Toolbox (“Neural Network Toolbox – MATLAB, ” n.d.)	*ANN*: (Biswas et al., 2016)
mySVM software (“mySVM – TU Dortmund,” n.d.)		*SVM*: (Li et al., 2009)
Python programming	Scikit-Learn package (“Scikit-Learn: machine learning in Python, ” n.d.)	*Regression*: (Zhang et al., 2017) *SVM*: (Dagnely et al., 2015; Chaobo Zhang et al., 2017) *Ensemble*: (Ahmad et al., 2017)
	NeuroLab (“NeuroLab 0.3.5, Neural Network Library for Python, ” n.d.)	*ANN*: (Ahmad et al., 2017)
	StatsModel (“StatsModels: Statistics in Python — statsmodels 0.9.0 documentation, ” n.d.)	*Regression*: (Dagnely et al., 2015)
	TensorFlow (“TensorFlow, ” n.d.)	*Regression*: (Amber et al., 2018) *SVM*: (Amber et al., 2018) *ANN*: (Amber et al., 2018) *DNN*: (Amber et al., 2018)
R programming (“R: The R Project for Statistical Computing, ” n.d.)	N/S	*Ensemble*: (Zhang et al., 2016) *Improved*: (Zhang et al., 2016)
	Keras package (“Keras Documentation, ” n.d.)	*DNN*: (Fan et al., 2019)
Weka software (“Weka 3 – Data Mining with Open Source Machine Learning Software in Java, ” n.d.)		*Regression*: (Massana et al., 2015) *k-NN*: (Wahid & Kim, 2016) *DT*: (Yu et al., 2010) *SVR*: (Massana et al., 2015) *ANN*: (Massana et al., 2015)

years, a large gap remains when trying to accurately account for occupancy and human behavior impact. Indeed, the review work has highlighted a lack of real occupancy data used in BECMF studies, often replaced by theoretical occupancy scenarios and resulting in large modeling uncertainties (Azar & Menassa, 2012). Furthermore, even when available, most occupants-related data only considered occupancy schedules (Table 3). This problem is partly due to the scarcity of residential and mixed-use buildings case studies. Indeed, as one of the “key factor influencing energy consumption in buildings” (Pisello & Asdrubali, 2014), accuracy can only be achieved by accessing detailed occupant-related data such as occupancy but also socio-economic data (Sütterlin, Brunner, & Siegrist, 2011; Tso ö Yau, 2007), behavior understanding, equipment usages and social interactions (Peschiera & Taylor, 2012).

Thus, complex occupants' behavior modeling has been integrated in some building energy forecasting studies. Simple approaches have been implemented to assess the general behavior of occupants (Zhang, Cao, & Romagnoli, 2018) and it's negative impact on building thermal loads (Ferracuti et al., 2017). Similarly, methods have been developed for the evaluation of consumers' energy efficiency and energy-savings behavior, to assess their impact on cooling load forecasting (Spandagos & Ng, 2018). The evaluation of peer networks (Peschiera & Taylor, 2012) and behavioral modifications on the energy consumption have been investigated as well (Xu et al., 2012). It highlighted specific incentives on energy efficient behaviors and that social networking within buildings and communities could have a significant impact on energy savings, comparable to typical retrofit actions (Pisello & Asdrubali, 2014).

6. Conclusions

We identified in this paper the main building energy consumption modeling and forecasting techniques and specifically reviewed data-driven methods. We covered approaches from the most conventional to the most recent research efforts on the topic. Six single techniques have been introduced with autoregressive models, statistical regressions, *k* nearest neighbors, decision trees, support vector machines, artificial neural networks, and two combined approaches: ensemble and improved models. Furthermore, we examined different machine learning approaches commonly used in the field including supervised, unsupervised, reinforcement and transfer learning. We presented the basic concepts and illustrated them through different recent studies. Peculiar attention was given to input data characteristics (i.e. origin, inputs types, time-series time-step, amount of data and the training–validation–testing ratio) and pre-processing methods. Finally, research gaps and future research directions are identified. Although data-driven methods offer a very wide range of tools to model and forecast buildings energy consumption that can adapt to many different situations, depending on the types of buildings, available data, modeling purpose, required accuracy and forecasting horizons, a universal protocol that can tackle the variety of problems faced is still lacking and a tradeoff, accounting for each problem constraints, is often to be made. In addition, several specific points still require particular attention such as long-term energy consumption forecasting, black-box data-driven techniques enhancement by hybridization with physical models, the accurate and realistic accounting for occupancy and occupants' behavior as well as real use cases of residential or mixed-use buildings.

Declaration of interest

None.

Acknowledgment

The study has been supported by the National Key R&D Program of China (Grant No. 2017YFC0704200).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.scs.2019.101533.

References

Ahmad, M. W., Mourshed, M., & Rezgui, Y. (2017). Trees vs neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and Buildings, 147*, 77–89. https://doi.org/10.1016/j.enbuild.2017.04.038.
Ahmad, T., Chen, H., Guo, Y., & Wang, J. (2018). A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review. *Energy and Buildings, 165*, 301–320. https://doi.org/10.1016/J.ENBUILD.2018.01.017.
Allab, Y., Pellegrino, M., Guo, X., Nefzaoui, E., & Kindinis, A. (2017). Energy and comfort assessment in educational building: Case study in a French university campus. *Energy and Buildings, 143*, 202–219. https://doi.org/10.1016/J.ENBUILD.2016.11.028.
Alobaidi, M. H., Chebana, F., & Meguid, M. A. (2018). Robust ensemble learning framework for day-ahead forecasting of household based energy consumption. *Applied Energy, 212*, 997–1012. https://doi.org/10.1016/J.APENERGY.2017.12.054.
Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews, 81*, 1192–1205. https://doi.org/10.1016/J.RSER.2017.04.095.
Amber, K. P., Ahmad, R., Aslam, M. W., Kousar, A., Usman, M., & Khan, M. S. (2018). Intelligent techniques for forecasting electricity consumption of buildings. *Energy, 157*, 886–893. https://doi.org/10.1016/j.energy.2018.05.155.
Amber, K. P., Aslam, M. W., Mahmood, A., Kousar, A., Younis, M. Y., Akbar, B., & … Hussain, S. K. (2017). Energy consumption forecasting for university sector buildings. *Energies, 10*(10), 1–18. https://doi.org/10.3390/en10101579.
ASHRAE (2009). *2009 Ashrae handbook fundamentals I-P edition – Energy estimating and modeling methods.* Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers.
ASHRAE (2013). *Standard 90. 1-2013 – Energy standard for buildings except low-rise residential buildings (SI edition).* Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers.
Azar, E., & Menassa, C. C. (2012). Agent-based modeling of occupants and their impact on energy use in commercial buildings. *Journal of Computing in Civil Engineering, 26*(4), 506–518. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000158.
Bagnasco, A., Fresi, F., Saviozzi, M., Silvestro, F., & Vinci, A. (2015). Electrical consumption forecasting in hospital facilities: An application case. *Energy and Buildings, 103*, 261–270.
Bishop, C. M. (2006). *Pattern recognition and machine learning.* New York: Springer.
Biswas, M. A. R., Robinson, M. D., & Fumo, N. (2016). Prediction of residential building energy consumption: A neural network approach. *Energy, 117*, 84–92. https://doi.org/10.1016/j.energy.2016.10.066.
Bourdeau, M., Guo, X., & Nefzaoui, E. (2018). Buildings energy consumption generation gap: A post-occupancy assessment in a case study of three higher education buildings. *Energy and Buildings, 159*. https://doi.org/10.1016/j.enbuild.2017.11.062.
Bouzerdoum, M., Mellit, A., & Massi Pavan, A. (2013). A hybrid model (SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. *Solar Energy, 98*, 226–235. https://doi.org/10.1016/J.SOLENER.2013.10.002.
Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time series analysis: Forecasting and control.* John Wiley.
Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123–140. https://doi.org/10.1023/A:1018054314350.
Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.
Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees.* Chapman & Hall/CRC.
Busoniu, L., Ernst, D., De Schutter, B., & Babuska, R. (2011). Approximate reinforcement learning: An overview. *2011 IEEE symposium on adaptive dynamic programming and reinforcement learning (ADPRL)*, 1–8. https://doi.org/10.1109/ADPRL.2011.5967353 IEEE.
Castelli, M., Trujillo, L., Vanneschi, L., & Popovič, A. (2015). Prediction of energy performance of residential buildings: A genetic programming approach. *Energy and Buildings, 102*, 67–74. https://doi.org/10.1016/j.enbuild.2015.05.013.
Chalal, M. L., Benachir, M., White, M., & Shrahily, R. (2016). Energy planning and forecasting approaches for supporting physical improvement strategies in the building sector: A review. *Renewable and Sustainable Energy Reviews, 64*, 761–776. https://doi.org/10.1016/j.rser.2016.06.040.
Chang, C., & Lin, C. (2013). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST), 2*, 1–39. https://doi.org/10.1145/1961189.1961199.
Chen, X., & Yang, H. (2018). Integrated energy performance optimization of a passively designed high-rise residential building in different climatic zones of China. *Applied Energy, 215*, 145–158. https://doi.org/10.1016/J.APENERGY.2018.01.099.
Choi, I. Y., Cho, S. H., & Kim, J. T. (2012). Energy consumption characteristics of high-rise apartment buildings according to building shape and mixed-use development. *Energy and Buildings, 46*, 123–131. https://doi.org/10.1016/J.ENBUILD.2011.10.038.
Chou, J.-S., & Bui, D.-K. (2014). Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. *Energy and Buildings, 82*, 437–446. https://doi.org/10.1016/J.ENBUILD.2014.07.036.
Collinge, W. O., DeBlois, J. C., Landis, A. E., Schaefer, L. A., & Bilec, M. M. (2016). Hybrid dynamic-empirical building energy modeling approach for an existing campus building. *Journal of Architectural Engineering, 22*(1), 04015010. https://doi.org/10.1061/(ASCE)AE.1943-5568.0000183.
Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*, 273–297.
Dagnely, P., Ruette, T., Tourwé, T., & Tsiporkova, E. (2015). Predicting hourly energy consumption. Can you beat an autoregressive model? *Proceeding of the 24th annual machine learning conference of Belgium and the Netherlands*.
Deb, C., & Lee, S. E. (2018). Determining key variables influencing energy consumption in office buildings through cluster analysis of pre- and post-retrofit building data. *Energy and Buildings, 159*, 228–245. https://doi.org/10.1016/J.ENBUILD.2017.11.007.
Deb, C., Zhang, F., Yang, J., Lee, S. E., & Shah, K. W. (2017). A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews, 74*, 902–924. https://doi.org/10.1016/j.rser.2017.02.085.
Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological), 39*(1), 1–38.
Deng, J. (1989). Introduction to grey system theory. *The Journal of Grey System, 1*, 1–24 Retrieved from http://www.researchinformation.co.uk/grey/IntroGreySysTheory.pdf.
DeST simulation software. (n.d.). Retrieved from http://dest.tsinghua.edu.cn/chinese/default1.asp?accessdenied=%2Fchinese%2Fdefault.asp.
Dong, B., Li, Z., Rahman, S. M. M., & Vega, R. (2016). A hybrid model approach for forecasting future residential electricity consumption. *Energy and Buildings, 117*, 341–351. https://doi.org/10.1016/j.enbuild.2015.09.033.
Ecotect Analysis|Autodesk Knowledge Network. (n.d.). Retrieved from https://knowledge.autodesk.com/support/ecotect-analysis?sort=score.
Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology, 77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x.
Energy consumption in households – Statistics explained. (n.d.). Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php/Energy_consumption_in_households.
EnergyPlus. (n.d.). Retrieved from https://energyplus.net/.
eQUEST. (n.d.). Retrieved from http://www.doe2.com/equest/.
European Environment Agency – final energy consumption by sector and fuel. (n.d.). Retrieved from https://www.eea.europa.eu/data-and-maps/indicators/final-energy-consumption-by-sector-9/assessment-4.
Fan, C., Wang, J., Gang, W., & Li, S. (2019). Assessment of deep recurrent neural network-based strategies for short-term building energy predictions. *Applied Energy, 236*(2018, July), 700–710. https://doi.org/10.1016/j.apenergy.2018.12.004.
Fan, C., Xiao, F., Li, Z., & Wang, J. (2018). Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: A review. *Energy and Buildings, 159*, 296–308. https://doi.org/10.1016/J.ENBUILD.2017.11.008.
Fan, C., Xiao, F., & Wang, S. (2014). Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. *Applied Energy, 127*, 1–10. https://doi.org/10.1016/J.APENERGY.2014.04.016.
Fan, C., Xiao, F., & Zhao, Y. (2017). A short-term building cooling load prediction method using deep learning algorithms. *Applied Energy, 195*, 222–233. https://doi.org/10.1016/J.APENERGY.2017.03.064.
Ferracuti, F., Fonti, A., Ciabattoni, L., Pizzuti, S., Arteconi, A., Helsen, L., & Comodi, G. (2017). Data-driven models for short-term thermal behaviour prediction in real buildings. *Applied Energy, 204*, 1375–1387. https://doi.org/10.1016/J.APENERGY.2017.05.015.
Fix, E., & Hodges, J. L., Jr. (1951). *Discriminatory analysis – Nonparametric discrimination: Consistency properties*.
Foucquier, A., Robert, S., Suard, F., Stéphan, L., & Jay, A. (2013). State of the art in building modelling and energy performances prediction: A review. *Renewable and Sustainable Energy Reviews, 23*, 272–288. https://doi.org/10.1016/J.RSER.2013.03.004.
Friedman, J. H. (1991). Multivariate adaptative regression splines. *The Annals of Statistics, 19*(1), 1–67.
Fu, Y., Li, Z., Zhang, H., & Xu, P. (2015). Using support vector machine to predict next day electricity load of public buildings with sub-metering devices. *Procedia Engineering,*

121, 1016–1022. https://doi.org/10.1016/J.PROENG.2015.09.097.
Fumo, N. (2014). A review on the basics of building energy estimation. *Renewable and Sustainable Energy Reviews, 31*, 53–60. https://doi.org/10.1016/J.RSER.2013.11.040.
Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2013). *Bayesian data analysis.* New York: Chapman and Hall/CRChttps://doi.org/10.1201/b16018.
Ghadi, Y. Y., Rasul, M. G., & Khan, M. M. K. (2017). Energy savings by fuzzy base control of occupancy concentration in institutional buildings. *Energy Procedia, 105*, 2850–2858. https://doi.org/10.1016/J.EGYPRO.2017.03.628.
Ghanbari, A., Abbasian-Naghneh, S., & Hadavandi, E. (2011). An intelligent load forecasting expert system by integration of ant colony optimization, genetic algorithms and fuzzy logic. *2011 IEEE symposium on computational intelligence and data mining (CIDM),* 246–251. https://doi.org/10.1109/CIDM.2011.5949432 IEEE.
Global Energy Statistical Yearbook, 2017|World Energy Statistics|Enerdata. (2017). Retrieved from https://yearbook.enerdata.net/.
Ho, T. K. (1995). Random decision forests. *Proceedings of 3rd international conference on document analysis and recognition, Vol. 1,* 278–282. https://doi.org/10.1109/ICDAR.1995.598994 IEEE Comput. Soc. Press.
Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679–688. https://doi.org/10.1016/J.IJFORECAST.2006.03.001.
IBM SPSS Modeler. (n.d.). Retrieved from https://www.ibm.com/products/spss-modeler.
IBM SPSS Statistics. (n.d.). Retrieved from https://www.ibm.com/products/spss-statistics.
International Energy Agency (2016). *Tracking clean energy progress 2016.* Retrieved from http://www.iea.org/publications/freepublications/publication/TrackingCleanEnergyProgress2016.pdf.
International Energy Agency (2017). *Tracking clean energy progress 2017 informing energy sector transformations.* Retrieved from http://www.iea.org/publications/freepublications/publication/TrackingCleanEnergyProgress2017.pdf.
Iowa Environmental Mesonet. (n.d.). Retrieved from https://mesonet.agron.iastate.edu/.
Jain, A. K. (2008). Data clustering: 50 years beyond *k*-means. *19th international conference on pattern recognition (ICPR)* (pp. 1–33).
Jeong, K., Koo, C., & Hong, T. (2014). An estimation model for determining the annual energy cost budget in educational facilities using SARIMA (seasonal autoregressive integrated moving average) and ANN (artificial neural network). *Energy, 71*, 71–79. https://doi.org/10.1016/j.energy.2014.04.027.
Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). Springer, Ed.
Jovanović, R.Ž., Sretenović, A. A., & Živković, B. D. (2015). Ensemble of various neural networks for prediction of heating energy consumption. *Energy and Buildings, 94*, 189–199.
Kandasamy, N. K., Karunagaran, G., Spanos, C., Tseng, K. J., & Soong, B. H. (2018). Smart lighting system using ANN-IMC for personalized lighting control and daylight harvesting. *Building and Environment, 139*(April), 170–180. https://doi.org/10.1016/j.buildenv.2018.05.005.
Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm. *Journal of Global Optimization, 39*(3), 459–471. https://doi.org/10.1007/s10898-007-9149-x.
Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics, 29*(2), 119–127.
Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95 – International conference on neural networks, Vol. 4,* 1942–1948. https://doi.org/10.1109/ICNN.1995.488968 IEEE.
Keras Documentation. (n.d.). Retrieved from https://keras.io/.
Kohonen, T. (1997). Exploration of very large databases by self-organizing maps. *Proceedings of international conference on neural networks (ICNN'97), Vol. 1,* 1–6. https://doi.org/10.1109/ICNN.1997.611622 IEEE.
Kolodner, J. (2014). *Case-based reasoning.* Elsevier Science.
Kristensen, M. H., & Petersen, S. (2016). Choosing the appropriate sensitivity analysis method for building energy model-based investigations. *Energy and Buildings, 130*, 166–176. https://doi.org/10.1016/j.enbuild.2016.08.038.
Kwok, S. S. K., & Lee, E. W. M. (2011). A study of the importance of occupancy to building cooling load in prediction by intelligent approach. *Energy Conversion and Management, 52*(7), 2555–2564.
Kylili, A., & Fokaides, P. A. (2015). European smart cities: The role of zero energy buildings. *Sustainable Cities and Society, 15*, 86–95. https://doi.org/10.1016/J.SCS.2014.12.003.
L'Heureux, A., Grolinger, K., Elyamany, H. F., & Capretz, M. A. M. (2017). Machine learning with big data: Challenges and approaches. *IEEE Access, 5*, 7776–7797. https://doi.org/10.1109/ACCESS.2017.2696365.
Lachut, D., Banerjee, N., & Rollins, S. (2014). Predictability of energy use in homes. *International green computing conference,* 1–10. https://doi.org/10.1109/IGCC.2014.7039146 IEEE.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539.
Lee, C.-M., & Ko, C.-N. (2009). Time series prediction using RBF neural networks with a nonlinear time-varying evolution PSO algorithm. *Neurocomputing, 73*(1–3), 449–460. https://doi.org/10.1016/j.neucom.2009.07.005.
Li, K., Hu, C., Liu, G., & Xue, W. (2015). Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis. *Energy and Buildings, 108*, 106–113. https://doi.org/10.1016/J.ENBUILD.2015.09.002.
Li, K., Xie, X., Xue, W., Dai, X., Chen, X., & Yang, X. (2018). A hybrid teaching-learning artificial neural network for building electrical energy consumption prediction. *Energy and Buildings, 174*, 323–334. https://doi.org/10.1016/j.enbuild.2018.06.017.
Li, X., Lu, J., Ding, L., Xu, G., & Li, J. (2009). Building cooling load forecasting model based on LS-SVM. *2009 Asia-Pacific conference on information processing,* 55–58. https://doi.org/10.1109/APCIP.2009.22 IEEE.
Li, X., & Wen, J. (2015). Building energy forecasting using system identification based on system characteristics test. *2015 workshop on modeling and simulation of cyber-physical energy systems (MSCPES),* 1–6. https://doi.org/10.1109/MSCPES.2015.7115401.
Li, Y. (2011). *LIBSVM-FarutoUltimate: A toolbox with implements for support vector machines based on libsvm.* Retrieved from http://www.matlabsky.com/.
Liu, D., Chen, Q., & Mori, K. (2015). Time series forecasting method of building energy consumption using support vector regression. *2015 ieee international conference on information and automation,* 1628–1632. https://doi.org/10.1109/ICInfA.2015.7279546 IEEE.
Liu, X., Marnay, C., Feng, W., Zhou, N., & Karali, N. (2017). *A review of the American Recovery and Reinvestment Act Smart Grid Projects and their implications for China.*
Ma, Z., Song, J., & Zhang, J. (2017). Energy consumption prediction of air-conditioning systems in buildings by selecting similar days based on combined weights. *Energy and Buildings, 151*, 157–166. https://doi.org/10.1016/j.enbuild.2017.06.053.
Magoulès, F., & Zhao, H.-X. (2016). *Data mining and machine learning in building energy analysis.* Wiley-ISTE.
Marino, D. L., Amarasinghe, K., & Manic, M. (2016). Building energy load forecasting using Deep Neural Networks. *IECON 2016 – 42nd annual conference of the IEEE Industrial Electronics Society,* 7046–7051. https://doi.org/10.1109/IECON.2016.7793413 IEEE.
Massa Gray, F., & Schmidt, M. (2018). A hybrid approach to thermal building modelling using a combination of Gaussian processes and grey-box models. *Energy and Buildings, 165*, 56–63. https://doi.org/10.1016/J.ENBUILD.2018.01.039.
Massana, J., Pous, C., Burgas, L., Melendez, J., & Colomer, J. (2015). Short-term load forecasting in a non-residential building contrasting models and attributes. *Energy and Buildings, 92*, 322–330. https://doi.org/10.1016/J.ENBUILD.2015.02.007.
Mat Daut, M. A., Hassan, M. Y., Abdullah, H., Rahman, H. A., Abdullah, M. P., & Hussin, F. (2017). Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review. *Renewable and Sustainable Energy Reviews, 70*, 1108–1118. https://doi.org/10.1016/J.RSER.2016.12.015.
MATLAB – MathWorks – MATLAB Simulink. (n.d.). Retrieved from https://www.mathworks.com/products/matlab.html.
Mena, R., Rodríguez, F., Castilla, M., & Arahal, M. R. (2014). A prediction model based on neural networks for the energy consumption of a bioclimatic building. *Energy and Buildings, 82*, 142–155.
Meteonorm: Irradiation data for every place on Earth. (n.d.). Retrieved from https://www.meteonorm.com/.
Miller, C., Nagy, Z., & Schlueter, A. (2018). A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings. *Renewable and Sustainable Energy Reviews, 81*(2016, December), 1365–1377. https://doi.org/10.1016/j.rser.2017.05.124.
Mitchell, M. (1998). *An introduction to genetic algorithms.* Bradford Book – The MIT Press.
Mocanu, E., Nguyen, P. H., Gibescu, M., & Kling, W. L. (2016a). Deep learning for estimating building energy consumption. *Sustainable Energy, Grids and Networks, 6*, 91–99. https://doi.org/10.1016/j.segan.2016.02.005.
Mocanu, E., Nguyen, P. H., Kling, W. L., & Gibescu, M. (2016b). Unsupervised energy prediction in a Smart Grid context using reinforcement cross-building transfer learning. *Energy and Buildings, 116*, 646–655. https://doi.org/10.1016/J.ENBUILD.2016.01.030.
mySVM – TU Dortmund. (n.d.). Retrieved from https://www-ai.cs.uni-dortmund.de/SOFTWARE/MYSVM/index.html.
Neto, A. H., & Fiorelli, F. A. S. (2008). Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. *Energy and Buildings, 40*(12), 2169–2176. https://doi.org/10.1016/J.ENBUILD.2008.06.013.
Neural Network Toolbox – MATLAB. (n.d.). Retrieved from https://www.mathworks.com/products/neural-network.html.
NeuroLab 0.3.5, Neural Network Library for Python. (n.d.). Retrieved from https://pythonhosted.org/neurolab/.
Newsham, G. R., & Birt, B. J. (2010). Building-level occupancy data to improve ARIMA-based electricity use forecasts. *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building – BuildSys'10,* 13. https://doi.org/10.1145/1878431.1878435 ACM Press New York, New York, USA.
Nilashi, M., Dalvi-Esfahani, M., Ibrahim, O., Bagherifard, K., Mardani, A., & Zakuan, N. (2017). A soft computing method for the prediction of energy performance of residential buildings. *Measurement, 109*, 268–280. https://doi.org/10.1016/J.MEASUREMENT.2017.05.048.
Obey, D. R. (2009). *Text – H.R.1 – 111th Congress (2009–2010): American Recovery and Reinvestment Act of 2009.* Retrieved from https://www.congress.gov/bill/111th-congress/house-bill/1/text.
Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering, 22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191.
Papadopoulos, S., Azar, E., Woon, W.-L., & Kontokosta, C. E. (2017). Evaluation of tree-based ensemble learning algorithms for building energy performance estimation. *Journal of Building Performance Simulation,* 1–11. https://doi.org/10.1080/19401493.2017.1354919.
Parti, M., & Parti, C. (1980). The total and appliance-specific conditional demand for electricity in the household sector. *The Bell Journal of Economics, 11*(1), 309. https://doi.org/10.2307/3003415.
Paudel, S., Elmitri, M., Couturier, S., Nguyen, P. H., Kamphuis, R., Lacarrière, B., & Le Corre, O. (2017). A relevant data selection method for energy consumption prediction of low energy building based on support vector machine. *Energy and Buildings, 138*, 240–256. https://doi.org/10.1016/J.ENBUILD.2016.11.009.
Pedersen, L. (2007). Use of different methodologies for thermal load and energy estimations in buildings including meteorological and sociological input parameters. *Renewable and Sustainable Energy Reviews, 11*(5), 998–1007. https://doi.org/10.

1016/J.RSER.2005.08.005.
Peschiera, G., & Taylor, J. E. (2012). The impact of peer network position on electricity consumption in building occupant networks utilizing energy feedback systems. *Energy and Buildings, 49*, 584–590. https://doi.org/10.1016/J.ENBUILD.2012.03.011.
Pisello, A. L., & Asdrubali, F. (2014). Human-based energy retrofits in residential buildings: A cost-effective alternative to traditional physical strategies. *Applied Energy, 133*, 224–235. https://doi.org/10.1016/J.APENERGY.2014.07.049.
Pulido-Arcas, J. A., Pérez-Fargallo, A., & Rubio-Bellido, C. (2016). Multivariable regression analysis to assess energy consumption and CO_2 emissions in the early stages of offices design in Chile. *Energy and Buildings, 133*, 738–753. https://doi.org/10.1016/J.ENBUILD.2016.10.031.
Python. (n.d.). Retrieved from https://www.python.org/.
Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*, 81–106.
Quinlan, J. R. (1993). *C4.5: Programs for machine learning.* Morgan Kaufmann Publishers.
R: The R Project for Statistical Computing. (n.d.). Retrieved from https://www.r-project.org/.
Rahman, A., Srikumar, V., & Smith, A. D. (2018). Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. *Applied Energy, 212*(October 2017), 372–385. https://doi.org/10.1016/j.apenergy.2017.12.051.
Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning.* MIT Press Retrieved from http://f3.tiera.ru/2/Cs_Computer science/CsLn_Lecture notes/Advanced Lectures on Machine Learning 2003(LNCS3176, Springer, 2004)(ISBN 3540231226)(248s).pdf#page=70.
Rathore, M. M., Paul, A., Hong, W.-H., Seo, H., Awan, I., & Saeed, S. (2018). Exploiting IoT and big data analytics: Defining Smart Digital City using real-time urban data. *Sustainable Cities and Society, 40*, 600–610. https://doi.org/10.1016/J.SCS.2017.12.022.
Ribeiro, M., Grolinger, K., ElYamany, H. F., Higashino, W. A., & Capretz, M. A. M. (2018). Transfer learning with seasonal and trend adjustment for cross-building energy forecasting. *Energy and Buildings, 165*, 352–363. https://doi.org/10.1016/J.ENBUILD.2018.01.034.
Rokach, L., & Maimon, O. (2005). Clustering methods. *Data mining and knowledge discovery handbook,* 321–352. https://doi.org/10.1007/0-387-25465-X_15 Springer-Verlag New York.
Ruparathna, R., Hewage, K., & Sadiq, R. (2017). Economic evaluation of building energy retrofits: A fuzzy based approach. *Energy and Buildings, 139*, 395–406. https://doi.org/10.1016/J.ENBUILD.2017.01.031.
Schapire, R. E., & Singer, Y. (1999). Improved boosting algorithms using confidence-rated predictions. *Machine Learning, 37*(3), 297–336. https://doi.org/10.1023/A:1007614523901.
Scikit-Learn: Machine learning in Python. (n.d.). Retrieved from https://scikit-learn.org/stable/.
Sekhar Roy, S., Roy, R., & Balas, V. E. (2018). Estimating heating load in buildings using multivariate adaptive regression splines, extreme learning machine, a hybrid model of MARS and ELM. *Renewable and Sustainable Energy Reviews, 82*, 4256–4268. https://doi.org/10.1016/J.RSER.2017.05.249.
Shi, G., Liu, D., & Wei, Q. (2016). Energy consumption prediction of office buildings based on echo state networks. *Neurocomputing, 216*, 478–488.
Siddharth, V., Ramakrishna, P. V., Geetha, T., & Sivasubramaniam, A. (2011). Automatic generation of energy conservation measures in buildings using genetic algorithms. *Energy and Buildings, 43*(10), 2718–2726. https://doi.org/10.1016/J.ENBUILD.2011.06.028.
Silva, B. N., Khan, M., & Han, K. (2018). Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities. *Sustainable Cities and Society, 38*(August 2017), 697–713. https://doi.org/10.1016/j.scs.2018.01.053.
Smart Metering deployment in the European Union|JRC Smart Electricity Systems and Interoperability. (n.d.). Retrieved from https://ses.jrc.ec.europa.eu/smart-metering-deployment-european-union.
Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing, 14*, 199–222.
Song, Q., & Chissom, B. S. (1993). Fuzzy time series and its models. *Fuzzy Sets and Systems, 54*(3), 269–277. https://doi.org/10.1016/0165-0114(93)90372-O.
Spandagos, C., & Ng, T. L. (2018). Fuzzy model of residential energy decision-making considering behavioral economic concepts. *Applied Energy, 213*, 611–625. https://doi.org/10.1016/J.APENERGY.2017.10.112.
StatsModels: Statistics in Python — statsmodels 0.9.0 documentation. (n.d.). Retrieved from https://www.statsmodels.org/stable/index.html.
Storn, R., & Price, K. (1997). Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization, 11*(4), 341–359. https://doi.org/10.1023/A:1008202821328.
Sütterlin, B., Brunner, T. A., & Siegrist, M. (2011). Who puts the most energy into energy conservation?. A segmentation of energy consumers based on energy-related behavioral characteristics. *Energy Policy, 39*, 8137–8152.
Swan, L. G., & Ugursal, V. I. (2009). Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews, 13*(8), 1819–1835. https://doi.org/10.1016/J.RSER.2008.09.033.
Tahmassebi, A., & Gandomi, A. H. (2018). Building energy consumption forecast using multi-objective genetic programming. *Measurement, 118*, 164–171. https://doi.org/10.1016/J.MEASUREMENT.2018.01.032.
Tang, F., Kusiak, A., & Wei, X. (2014). Modeling and short-term prediction of HVAC system with a clustering algorithm. *Energy and Buildings, 82*, 310–321. https://doi.org/10.1016/j.enbuild.2014.07.037.
Tardioli, G., Kerrigan, R., Oates, M., O'Donnell, J., & Finn, D. (2015). Data driven approaches for prediction of building energy consumption at urban level. *Energy Procedia, 78*, 3378–3383. https://doi.org/10.1016/J.EGYPRO.2015.11.754.
TensorFlow. (n.d.). Retrieved from https://www.tensorflow.org/.
Toffanin, D. (2016). *Generation of customer load profiles based on smart-metering time series, building-level data and aggregated measurements.* Swiss Federal Institute of Technology (ETH).
TRNSYS: Transient System Simulation Tool. (n.d.). Retrieved from http://www.trnsys.com/.
Tsanas, A., & Xifara, A. (2012). Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and Buildings, 49*, 560–567. https://doi.org/10.1016/J.ENBUILD.2012.03.003.
Tsinghua University Building Energy Research Center (2016). *China building energy use 2016*.
Tso, G. K. F., & Yau, K. K. W. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy, 32*(9), 1761–1768. https://doi.org/10.1016/J.ENERGY.2006.11.010.
U.S. Energy Information Administration (EIA) (2018). *How many smart meters are installed in the United States, and who has them?* Retrieved from https://www.eia.gov/tools/faqs/faq.php?id=108&t=3.
Valgaev, O., & Kupzog, F. (2016). Building power demand forecasting using K-nearest neighbors model – initial approach. *2016 IEEE PES Asia-Pacific power and energy engineering conference (APPEEC),* 1055–1060. https://doi.org/10.1109/APPEEC.2016.7779700 IEEE.
Valgaev, O., Kupzog, F., & Schmeck, H. (2017). Building power demand forecasting using K-nearest neighbours model – practical application in Smart City Demo Aspern project. *CIRED – Open Access Proceedings Journal, 2017*(1), 1601–1604. https://doi.org/10.1049/oap-cired.2017.0419.
Wahid, F., & Kim, D. (2016). A prediction approach for demand analysis of energy consumption using *k*-nearest neighbor in residential buildings. *International Journal of Smart Home, 10*(2), 97–108. https://doi.org/10.14257/ijsh.2016.10.2.10.
Wang, J., Wang, J., Li, Y., Zhu, S., & Zhao, J. (2014). Techniques of applying wavelet denoising into a combined model for short-term load forecasting. *International Journal of Electrical Power and Energy Systems, 62*, 816–824. https://doi.org/10.1016/j.ijepes.2014.05.038.
Wang, Z., & Srinivasan, R. S. (2017). 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*, 796–808. https://doi.org/10.1016/J.RSER.2016.10.079.
Wang, Z., Wang, Y., & Srinivasan, R. S. (2018a). A novel ensemble learning approach to support building energy use prediction. *Energy and Buildings, 159*, 109–122. https://doi.org/10.1016/J.ENBUILD.2017.10.085.
Wang, Z., Wang, Y., Zeng, R., Srinivasan, R. S., & Ahrentzen, S. (2018b). Random Forest based hourly building energy prediction. *Energy and Buildings, 171*, 11–25. https://doi.org/10.1016/j.enbuild.2018.04.008.
Wei, Y., Zhang, X., Shi, Y., Xia, L., Pan, S., Wu, J., & … Zhao, X. (2018). A review of data-driven approaches for prediction and classification of building energy consumption. *Renewable and Sustainable Energy Reviews, 82*, 1027–1047. https://doi.org/10.1016/J.RSER.2017.09.108.
Weka 3 – Data Mining with Open Source Machine Learning Software in Java. (n.d.). Retrieved from https://www.cs.waikato.ac.nz/ml/weka/index.html.
Wolpert, D. H. (1992). Stacked generalization. *Neural Networks, 5*(2), 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1.
Woo, Y. E., & Cho, G. H. (2018). Impact of the surrounding built environment on energy consumption in mixed-use building. *Sustainability (Switzerland), 10*(3), https://doi.org/10.3390/su10030832.
Xifara, A., & Tsanas, A. (n.d.). UCI machine learning repository: Energy efficiency Data Set. Retrieved from https://archive.ics.uci.edu/ml/datasets/energy+efficiency.
Xu, X., Taylor, J. E., Pisello, A. L., & Culligan, P. J. (2012). The impact of place-based affiliation networks on energy conservation: An holistic model that integrates the influence of buildings, residents and the neighborhood context. *Energy and Buildings, 55*, 637–646.
Yalcinoz, T., & Eminoglu, U. (2005). Short term and medium term power distribution load forecasting by neural networks. *Energy Conversion and Management, 46*(9–10), 1393–1405. https://doi.org/10.1016/J.ENCONMAN.2004.07.005.
Yang, J., Santamouris, M., & Lee, S. E. (2016, January). Review of occupancy sensing systems and occupancy modeling methodologies for the application in institutional buildings. *Energy and Buildings, 121*, 344–349. https://doi.org/10.1016/j.enbuild.2015.12.019.
Yildiz, B., Bilbao, J. I., & Sproul, A. B. (2017). A review and analysis of regression and machine learning models on commercial building electricity load forecasting. *Renewable and Sustainable Energy Reviews, 73*, 1104–1122. https://doi.org/10.1016/J.RSER.2017.02.023.
Yu, Z., Haghighat, F., Fung, B. C. M., & Yoshino, H. (2010). A decision tree method for building energy demand modeling. *Energy and Buildings, 42*(10), 1637–1646. https://doi.org/10.1016/J.ENBUILD.2010.04.006.
Yun, K., Luck, R., Mago, P. J., & Cho, H. (2012). Building hourly thermal load prediction using an indexed ARX model. *Energy and Buildings, 54*, 225–233. https://doi.org/10.1016/J.ENBUILD.2012.08.007.
Zhang, C., Cao, L., & Romagnoli, A. (2018). On the feature engineering of building energy data mining. *Sustainable Cities and Society*. https://doi.org/10.1016/J.SCS.2018.02.016.
Zhang, C., Zhao, Y., Zhang, X., Fan, C., & Li, T. (2017). An improved cooling load prediction method for buildings with the estimation of prediction intervals. *Procedia Engineering, 205*, 2422–2428. https://doi.org/10.1016/J.PROENG.2017.09.967.
Zhang, F., Deb, C., Lee, S. E., Yang, J., & Shah, K. W. (2016). Time series forecasting for building energy consumption using weighted Support Vector Regression with

differential evolution optimization technique. *Energy and Buildings, 126*, 94–103. https://doi.org/10.1016/J.ENBUILD.2016.05.028.
Zhao, D., Zhong, M., Zhang, X., & Su, X. (2016). Energy consumption predicting model of VRV (Variable refrigerant volume) system in office buildings based on data mining. *Energy, 102*, 660–668.
Zhao, H., & Magoulès, F. (2012). A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews, 16*(6), 3586–3592. https://doi.org/10.1016/J.RSER.2012.02.049.
Zheng, Z., Zhuang, Z., Lian, Z., & Yu, Y. (2017). Study on building energy load prediction based on monitoring data. *Procedia Engineering, 205*, 716–723. https://doi.org/10.1016/j.proeng.2017.09.894.