Renewable and Sustainable Energy Reviews 81 (2018) 1192–1205

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

A review of data-driven building energy consumption prediction studies

Kadir Amasyali, Nora M. El-Gohary*

Dept. of Civil and Environmental Engineering, Univ. of Illinois at Urbana-Champaign, 205 N. Mathews Ave., Urbana, IL 61801, USA

ARTICLE INFO

Keywords:
Building energy
Energy consumption prediction
Data-driven methods
Machine learning

ABSTRACT

Energy is the lifeblood of modern societies. In the past decades, the world's energy consumption and associated CO_2 emissions increased rapidly due to the increases in population and comfort demands of people. Building energy consumption prediction is essential for energy planning, management, and conservation. Data-driven models provide a practical approach to energy consumption prediction. This paper offers a review of the studies that developed data-driven building energy consumption prediction models, with a particular focus on reviewing the scopes of prediction, the data properties and the data preprocessing methods used, the machine learning algorithms utilized for prediction, and the performance measures used for evaluation. Based on this review, existing research gaps are identified and future research directions in the area of data-driven building energy consumption prediction are highlighted.

1. Introduction

Buildings represent a large portion of the world's energy consumption and associated CO_2 emissions. For example, the building sector represents 39% and 40% of the energy consumption and 38% and 36% of the CO_2 emissions in the U.S. [1] and Europe [2], respectively. The use of energy that is generated from fossil fuels contributes CO_2 emissions and causes air pollution and global warming. Prediction of building energy consumption is crucial for improved decision making towards reducing energy consumption and CO_2 emissions, because it can assist in evaluating different building design alternatives and building operation strategies (in terms of their energy efficiency) and improving demand and supply management. However, building energy consumption prediction remains to be a challenging task due to the variety of factors that affect the consumption such as the physical properties of the building, the installed equipment, the outdoor weather conditions, and the energy-use behavior of the building occupants [3].

Two main approaches have been taken for building energy consumption prediction: physical modelling approach and data-driven approach. Physical models (also known as engineering methods or white-box models) rely on thermodynamic rules for detailed energy modelling and analysis. Examples of building energy simulation software that utilize physical models include EnergyPlus, eQuest, and Ecotect. These types of software calculate building energy consumption based on detailed building and environmental parameters such as building construction details; operation schedules; HVAC design information; and climate, sky, and solar/shading information [4]. However, some of such detailed data may not be available to the users at the time of simulation. Failure to provide accurate input can result in poor prediction performance.

Data-driven building energy consumption prediction modelling, on the other hand, does not perform such energy analysis or require such detailed data about the simulated building, and instead learns from historical/available data for prediction. Data-driven energy consumption prediction has gained a lot of research attention in recent years [5], despite its possible limitations (as discussed in Section 8). In response, a number of review studies on the analysis of existing data-driven approaches has been published. The reviews mostly focused on the machine learning methods/algorithms used in previous research efforts. Despite the importance of these efforts, there is still a lack of review studies that analyze existing data-driven approaches from a more multivariate perspective, including data aspects such as what data types and sizes were used and what features were selected for learning. Such a review would help reveal existing research gaps in the field of data-driven building energy consumption prediction and point towards future research directions.

To address this gap, this paper offers a review of data-driven building energy consumption prediction studies that utilized machine learning algorithms, including support vector machines (SVM), artificial neural networks (ANN), decision trees, and other statistical algorithms. The paper focuses on reviewing the types of buildings, temporal granularities, types of energy consumption predicted, types of data, types of features, and data sizes in the existing studies; and

* Corresponding author.
E-mail addresses: amasyal2@illinois.edu (K. Amasyali), gohary@illinois.edu (N.M. El-Gohary).

http://dx.doi.org/10.1016/j.rser.2017.04.095
Received 4 January 2016; Received in revised form 23 March 2017; Accepted 27 April 2017
Available online 11 September 2017
1364-0321/

provides a discussion of the review results and future research directions. The paper is organized as follows. Section 2 provides a concise overview of existing review studies on data-driven building energy consumption prediction and identifies the gaps in this area. Section 3 gives a brief introduction on the background of data-driven approaches. Section 4 defines the methodology used in this review study. Section 5 reviews previous studies in terms of the scopes of prediction, the data properties and the data preprocessing methods used, the machine learning algorithms utilized for prediction, and the performance measures used for evaluation. Section 6 discusses the previous studies in terms of the temporal granularities of prediction, the types of buildings, and the types of energy consumption predicted. Finally, Section 7 discusses future research directions, Section 8 discusses the limitations of data-driven energy consumption prediction, and Section 9 summarizes the conclusions.

2. Existing review studies on data-driven building energy consumption prediction

Data-driven building energy consumption prediction gained a lot of attention in recent years. In response, a number of review studies has focused on the analysis of existing data-driven efforts. For example, Zhao and Magoulès [4] classified building energy consumption prediction methods as elaborate engineering methods, simplified engineering methods, statistical methods, ANN-based methods, SVM-based methods, and grey models; and conducted some comparative analysis in terms of model complexity, ease of use, running speed, inputs needed, and accuracy. Ahmad et al. [2] focused on the review of ANN-based, SVM-based, and hybrid methods and discussed the principles, advantages, and disadvantages of these methods. Fumo [6] summarized the classification of building energy consumption prediction methods proposed by various studies and placed a special emphasis on the review of model calibration and verification and weather data used for modelling. Li and Wen [7] conducted an inclusive review; they reviewed state-of-the-art studies not only on building energy modelling and prediction but also on building critical component modelling (e.g., photovoltaic power generation modelling), building energy modelling for demand response (e.g., weather condition forecasting), agent-based building energy modelling, and system identification for building energy modelling. Li et al. [8] reviewed the methods for building energy benchmarking and proposed a flowchart that intends to assist users in choosing the proper prediction method. Chalal et al. [9] focused on both building scale and urban scale energy consumption prediction and further classified and discussed the available methods within each scale. Wang and Srinivasan [10] reviewed and compared the principles, applications, advantages, and disadvantages of single AI-based methods (e.g., ANN and SVM) and ensemble methods.

The majority of these studies provided a comprehensive review on energy consumption prediction research efforts with a particular focus on the machine learning methods/algorithms used in these research studies. Despite the importance of these review efforts, there is still a lack of review studies that cover building energy consumption prediction research in terms of the scopes of prediction (e.g., heating energy consumption), the types of data used (e.g., real data, simulated data), the types of features used for prediction (e.g., outdoor weather conditions, indoor environmental conditions), the sizes of the data (e.g., duration of data collection, number of data instances), and the data preprocessing methods utilized (e.g., data reduction). Such a review is essential for identifying the research gaps and highlighting the future research directions in the field of data-driven building energy consumption prediction.

3. Background

Developing a data-driven model, typically, consists of four primary steps: data collection, data preprocessing, model training, and model testing. In the field of building energy consumption prediction, data collection involves collecting historical/available data for model training such as outdoor weather condition and electricity consumption data. Data preprocessing may include data cleaning, data integration, data transformation, and/or data reduction. Model training is the training of the model using a training dataset. Model testing aims to evaluate the model using standard evaluation measures.

SVM, ANN, decision trees, and other statistical algorithms are the most commonly-used supervised machine learning algorithms for model training. SVM is a kernel-based machine learning algorithm, which can be used for both regression and classification [11]. The algorithm is good at solving non-linear problems even with a relatively small amount of training data [4]. SVM solves a non-linear problem through transforming the non-linearity between features x_i (e.g., dry-bulb temperature and global solar radiation) and target y_i (e.g., cooling energy consumption) using linear mapping in two steps. First, it projects the non-linear problem into a high-dimensional space and determines the function f(x) that fits best in the high-dimensional space. Second, it applies a kernel function to make the complex non-linear map a linear problem. For further details on the prediction principle using SVM, the readers are referred to [9]. SVM is one of the most robust and accurate algorithms and has been listed in the top-ten most influential data mining algorithms in the research community by the IEEE International Conference on Data Mining [11]. It was found to outperform other machine learning algorithms in numerous applications. In order to increase the computational efficiency of SVM, least squares SVM (LS-SVM) (e.g., [12]) and parallel SVM (e.g., [13]) were also implemented in the field of building energy consumption prediction.

ANN is a non-linear computational model, inspired by the human brain. A typical ANN includes three sequential layers: the input layer, the hidden layer, and the output layer. Each layer has a number of interconnected neurons, and each neuron has an activation function. Three types of parameters are typically used to define ANNs: the interconnection pattern between the neurons of the different layers, the learning process of updating the weights of the interconnections, and the activation function that converts a neuron's weighted input to its output activation [14]. In ANN, each feature (e.g., dry-bulb temperature) is multiplied by its corresponding neuron weight and summed up with the bias. The activation function is then applied to determine the output (e.g., cooling energy consumption). For further details on the prediction principle using ANN, the readers are referred to [9]. ANN is one of the most popular algorithms used in building energy consumption prediction [2]. Examples of ANNs include the back propagation neural networks (BPNN), radial basis function neural networks (RBFNN), general regression neural networks (GRNN), feed forward neural network (FFNN), and adaptive network-based fuzzy inference system (ANFIS). Other methods that can be used in conjunction with ANN include the hierarchical mixture of experts (HME), fuzzy c-means (FCC), and multilayer perceptron (MLP).

Decision tree algorithms use a tree to map instances into predictions. In a decision tree model, each non-leaf node represents one feature, each branch of the tree represents a different value for a feature, and each leaf node represents a class of prediction. Decision trees is a flexible algorithm that could grow with an increased amount of training data [15]. The classification and regression trees (CART), chi-squared automatic interaction detector (CHAID), random forest (RF), and boosting trees (BT) are the most widely-used decision tree methods in the area of building energy consumption prediction.

Other statistical algorithms include multiple linear regression (MLR), general linear regression (GLR), ordinary least squares regression (OLS), autoregressive (AR), autoregressive integrated moving average (ARIMA), Bayesian regression, polynomial regression (poly), exponential regression, multivariate adaptive regression splines (MARS), case-based reasoning (CBR), and k-nearest neighbors (kNN).

Algorithms used for developing energy consumption prediction models have advantages and disadvantages. For example, ANN and SVM require many parameters and might become computationally expensive, but their prediction accuracy is, in many cases, better than decision trees and statistical algorithms. Decision trees and other statistical algorithms, on the other hand, are generally easy to use and computationally inexpensive, but their performance is usually fair [4].

4. Methodology

The research methodology was composed of five primary steps:

- Conducting a keyword-based search: A keyword-based search of research articles and abstracts was conducted using Google Scholar. Examples of the keywords that were used are: building energy estimation, building energy use prediction, building energy consumption forecasting, building energy modelling. Google Scholar was selected, because it can rank articles based on some factors such as number of citations, authors, and publisher.
- Screening the retrieved articles: The articles were screened for relevance using the following criteria: (1) the approach must be data-driven; and (2) the purpose must be to predict building energy consumption.
- Identifying and screening additional articles: The articles that cited or were cited by an article that passed the screening test were further identified as additional candidate articles. These articles were further screened using the same two relevance criteria defined above.
- Reviewing all relevant articles: All articles identified in steps 2 and 3 were analytically reviewed to define their purpose of prediction, scope of prediction, data properties and data preprocessing methods, machine learning algorithm(s), and performance.
- Analyzing the review results to identify gaps and future directions: The review results were analyzed to identify the research gaps in the field of data-driven building energy consumption and highlight future research directions.

5. Review of existing data-driven energy consumption prediction models

5.1. Scope of prediction

The scope of the studies was classified in terms of type of building, temporal granularity, and type of energy consumption predicted. Two types of buildings (residential and non-residential), five types of temporal granularities (sub-hourly, hourly, daily, monthly, and yearly), and four types of energy consumption (heating, cooling, lighting, and overall energy consumption) were defined.

Existing models covered residential and/or non-residential buildings, with different temporal granularities and for different types of energy consumption. Fig. 1. shows the distribution of the reviewed models according to type of building, temporal granularity, and type of energy consumption. Only 19% of these models focused on residential buildings, with the remaining models focusing on non-residential buildings including commercial and educational buildings. The majority of these models, 57%, were developed for predicting hourly energy consumption, while 12%, 15%, 4%, and 12% of the models focused on sub-hourly, daily, monthly, and yearly consumption, respectively. Overall, 47% of the models focused on predicting overall energy consumption, with 31% and 20% focusing on cooling and heating energy consumption, respectively, and only 2% focusing on lighting energy consumption prediction. The scope of each reviewed model is summarized in Table 1, in terms of building type, temporal granularity, type of energy consumption, and purpose of prediction.

5.2. Data properties and data preprocessing

5.2.1. Types of data: real, simulated, or benchmark

Data were classified into three types: (1) real data, (2) simulated data, and (3) public benchmark data (e.g., datasets provided for energy consumption prediction competitions). Fig. 1. shows the distribution of the reviewed studies by type of data used for training and testing. The majority (67%) of these studies used real data to train and test their models, while 19% and 14% of the studies used simulated and public benchmark data, respectively. Table 1 shows the types of data used in the reviewed studies.

Real data cover data collected through smart energy meters, sensors, building management systems, and weather stations; in addition to utility bills, energy consumption surveys, and energy consumption statistics and reports [16]. Sensor-based approaches have several advantages and disadvantages. On one hand, sensor-based approaches provide actual indoor environmental condition data and energy consumption levels. On the other hand, installing sensors brings an additional cost and effort not only to install the required sensors, but also to test and ensure the quality of the data collected [12]. Otherwise, sensor data may include noise, missing values, and/or outliers, which would affect the performance of the prediction models adversely.

Simulation-based studies, on the other hand, model an existing or unexisting building in a building energy simulation software tool – such as EnergyPlus, DeST, DOE2, or Ecotect – and obtain the needed data through running the simulations. By nature of modelling, a model cannot fully represent its prototype or exactly behave same as it does. For example, Li et al. [17] showed that current building energy software tools are, in some cases, limited in evaluating the performance of energy conservation measures. Simulation data are, however, useful in cases where real data are limited (e.g., when instrumenting a building is difficult due to technical difficulties and/or economic reasons).

Other studies (e.g., [12,18,19]) utilized publicly-available benchmark datasets such as the ASHRAE's Great Building Energy Predictor Shootout and EUNITE dataset. This type of datasets provides benchmark data that can be used to compare the performance of different models.

5.2.2. Types of features

A machine learning model predicts energy consumption based on a set of features. These features can be related to outdoor weather conditions, indoor environmental conditions, building characteristics, time, occupancy and occupant energy use behavior, and/or historical energy consumption. Outdoor weather condition features include dry-bulb temperature, dew point temperature, relative humidity, global solar radiation, wind speed, wind direction, degree of cloudiness, pressure, rainfall amount, and evaporation. Indoor environmental condition features include room temperature, room relative humidity, and indoor lighting level. Building characteristic features include relative compactness, surface area, wall area, roof area, overall height, orientation, glazing area, glazing area distribution, mean heat transfer coefficient of building walls, mean thermal inert index of building walls, roof heat transfer coefficient, building size coefficient, absorption coefficient for solar radiation of exterior walls, eastern window-wall ratio (WWR), western WWR, southern WWR, northern WWR, mean WWR, shading coefficient (SC) of eastern window, SC of western window, SC of southern window, SC of northern window, and integrated SC. Time features include the type of day (e.g., weekday, weekend, holiday) and the type of hour (e.g., daytime, nighttime). Occupant energy use behavior and occupancy features include building use schedule, heat gain through lights and people, water temperature, and number of occupants.

For all these types of features, some studies used data considering various past time steps (e.g., past hour) in history. For example, Li et al.

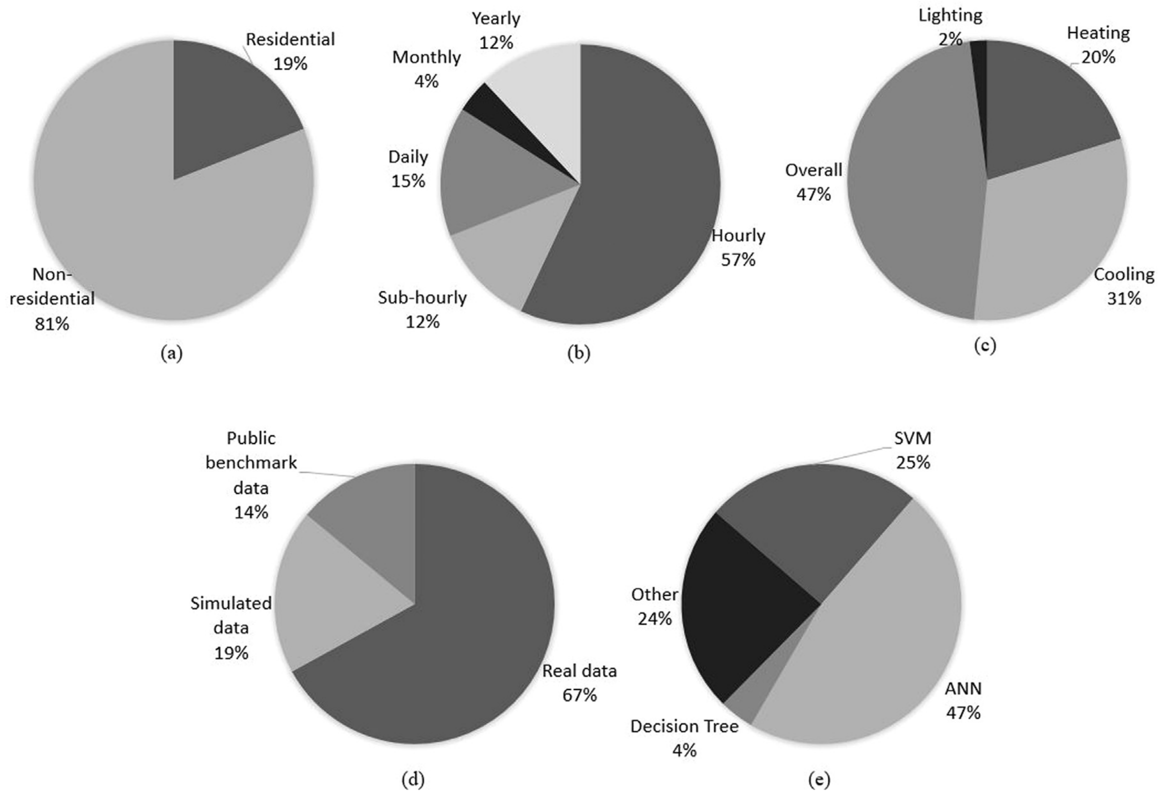

Fig. 1. Distribution of the reviewed models according to (a) type of building, (b) temporal granularity, (c) type of energy consumption, (d) type of data, (e) machine learning algorithm.

[20] used current outdoor dry-bulb temperature, outdoor dry-bulb temperature of an hour ago, outdoor dry-bulb temperature of two hours ago, current relative humidity, current solar radiation, and solar radiation of an hour ago to predict building cooling load. Jain et al. [16] used electricity consumption of the previous two time steps, current temperature, current solar flux, a denote for weekend/holiday or weekday, sine of current hour, and cosine of current hour to predict the electricity consumption of a multi-family residential building. Table 1 summarizes the features used in the reviewed models.

5.2.3. Data sizes

The sizes of datasets varied from 2-week (e.g., [21]) to 4-year energy consumption data (e.g. [22,23]). A small dataset may not be able to capture a representative sample of data, whereas a large dataset requires a lot of computational effort to process. The majority (56%) of the reviewed studies utilized one-month to one-year long datasets; 9% utilized datasets shorter than one-month; and 31% utilized datasets longer than one-year. Table 1 shows the dataset sizes used in the reviewed studies.

5.2.4. Data preprocessing

Data preprocessing is essential for any data-driven approach, because any incorrect or inconsistent data can cause errors in the analysis [24]. Data preprocessing may include data cleaning, data integration, data transformation, and/or data reduction. Data cleaning is the process of detecting and correcting (completing, modifying, replacing, and/or removing) the incomplete, incorrect, inaccurate, irrelevant, and/or noisy parts of the data. For example, data collected through sensors are usually noisy and often incomplete [25]. Data integration is the process of combining multiple data from different sources. For example, outdoor weather condition data and hourly electricity consumption data come from different sources, but are combined in a single dataset for training and testing. Data transformation is the process of transforming the data into the format that is required by the learning algorithm. Data transformation may include normalization, smoothing, aggregation/disaggregation, and/or generalization of the data. Data reduction is the process of reducing the dimensionality of the dataset, which is not only computationally more efficient but may also enhance the performance of the machine learning algorithm by removing non-discriminative features. There are different techniques for data reduction including principal component analysis (PCA) and kernel PCA (KPCA). For example, Xuemei et al. [26] applied PCA and KPCA for reducing the dimensionality of the data and compared the performances of SVM with PCA, SVM with KPCA, and SVM without any data reduction techniques. They also applied C-mean clustering to ensure that the training samples were chosen based on the similarity degree of the input samples and compared the performances of fuzzy C-means (FCM) fuzzy SVM, FCM-SVM, and SVM without any clustering [27].

5.3. Machine learning algorithms

A machine learning algorithm is needed to train an energy consumption prediction model. Previous studies in data-driven building energy consumption prediction have utilized SVM, ANN, decision trees, and/or other statistical algorithms. Fig. 1. shows the distribution of the studies by type of machine learning algorithm. Overall, 47% and 25% of the studies utilized ANN and SVM, respectively, to train their models. Only 4% of the studies utilized decision trees. On the other hand, 24% of the studies utilized other statistical algorithms such as MLR, OLS, and ARIMA.

Some studies also compared the effectiveness of different algorithms in energy consumption prediction. For example, Li et al. [20] compared SVM and BPNN; Borges et al. [28] compared SVM and AR; Xuemei et al. [29] compared LS-SVM and BPNN; Liu and Chen [21] compared SVM and ANN; Penya et al. [30] compared poly, exponential, mixed, AR, ANN, SVM, and Bayesian Network; Platon et al. [31] compared ANN and CBR; Jain et al. [32] compared SVM and MLR; Hou et al. [33] compared ARIMA and ANN; Penya et al. [34] compared AR, ARIMA, ANN, and Bayesian Network; Fan et al. [35] compared

Table 1
Scope, data properties, algorithms, and performance of the energy consumption prediction models [3,12,13,16,18-23,26-78].

Reference	Learning algorithm (type)	Building type	Temporal granularity	Type of energy consumption predicted	Purpose of prediction	Type of dataset (simulation tool)	Types of feature	Data size	Performance (metric)
[26]	SVM (RBF)	Non-residential	Hourly	Cooling	HVAC system operation improvement	Real (N/A)	Date, daily average temperature, daily lowest temperature, daily highest temperature	620 instances	0.17 (RMSE)
	PCA-SVM (RBF)								0.04 (RMSE)
	KPCA-SVM (RBF)								0.02 (RMSE)
[41]	SVM (RBF)	Non-residential	Hourly	Cooling	N/S	Real (N/A)	Date, daily average temperature, daily lowest temperature, daily highest temperature	620 instances	0.17 (RMSE)
	PCA-SVM (RBF)								0.04 (RMSE)
	PCA-WSVM (RBF)								0.03 (RMSE)
[20]	SVM (RBF)	Non-residential	Hourly	Cooling	HVAC system design	Simulated (DeST)	Dry-bulb temperature, relative humidity, solar radiation	5 months	1.15% - 1.18% (CV)
	ANN(BPNN)								2.22% - 2.36% (CV)
[37]	SVM (RBF)	Non-residential	Hourly	Cooling	HVAC system design	Simulated (DeST)	Dry-bulb temperature, relative humidity, solar radiation	5 months	1.15% - 1.18% (CV)
	ANN(BPNN)								2.22% - 2.36% (CV)
	ANN(RBFNN)								1.43% - 1.51% (CV)
	ANN(GRNN)								1.19% - 1.20% (CV)
[29]	LS-SVM (RBF)	Non-residential	Hourly	Cooling	HVAC system optimization	Simulated (DeST)	Dry-bulb temperature, relative humidity, solar radiation	4 months	5.56% (CV)
	ANN(BPNN)								11.84% (CV)
[27]	SVM (RBF)	Non-residential	Hourly	Cooling	HVAC system optimization	Real (N/A)	N/S	6 months	3.85% (CV)
	FCM-SVM (RBF)								2.68% (CV)
	FCM-FSVM (RBF)								1.24% (CV)
[42]	SVM (RBF)	Non-residential	Hourly	Overall	HVAC system efficiency improvement	Real (N/A)	Temperature, dew point temperature, pressure, wind direction, wind speed, humidity, precipitation	~27.5 months	0.71 - 0.95 (R^2)
[36]	ANN(BPNN)	Residential	Hourly	Cooling	Energy-efficient building design	Simulated (Ecotect)	Relative compactness, surface area, wall area, roof area, overall height, orientation, glazing area, glazing area distribution	N/S	1.68 kW (RMSE)
	SVM (RBF)								1.65 kW (RMSE)
	DT (CART)								1.84 kW (RMSE)
	DT (CHAID)								1.86 kW (RMSE)
	GLR								1.74 kW (RMSE)
	ANN(BPNN)			Heating					0.61 kW (RMSE)
	SVM (RBF)								0.35 kW (RMSE)
	DT (CART)								0.80 kW (RMSE)
	DT (CHAID)								0.91 kW (RMSE)
	GLR								1.04 kW (RMSE)
[12]	MLR	Residential	Hourly	Overall	N/S	Real (N/A)	140 different sensor data	A year	26.27% - 38.53% (CV)
	ANN(FFNN)								24.32% - 37.15% (CV)
	SVM								21.32% - 31.88% (CV)
	LS-SVM								20.05% - 30.66% (CV)
	ANN(HME-REG)								26.14% - 38.22% (CV)
	ANN(HME-FFNN)								20.15% - 32.98% (CV)
	ANN(FCM-FFNN)								20.53% - 32.92% (CV)
	MLR	N/S				ASHRAE dataset (N/A)	Temperature, solar flux, date, sin of current hour, cosine of current hour	6 months	4.07% (CV)
	ANN(FFNN)								2.93% (CV)
	SVM								3.97% (CV)
	LS-SVM								6.35% (CV)
	ANN(HME-REG)								4.05% (CV)
	ANN(HME-FFNN)								2.75% (CV)
	ANN(FCM-FFNN)								2.71% (CV)
[22]	OLS	Non-residential	Hourly	Overall	Green electricity production management	Real (N/A)	Occupancy, recency, temperature, irradiance, time	4 years	2.05 (MAE)
	SVM (RBF)								1.94 (MAE)
[39]	MLR	Non-residential	Hourly	Overall	Daily power system operation and control	Real (N/A)	Temperature, relative humidity, solar radiation, indoor temperature, indoor relative humidity, indoor light level, occupancy, date	~A year	4.68% (MAPE)
	ANN(MLP)								0.45% (MAPE)
	SVM (PUK)								0.06% (MAPE)
[34]	Poly	Non-residential	Hourly	Overall	N/S	Real (N/A)	Day of week, type of day, season, wind direction, humidity, precipitation, sigma direction, sigma speed, air temperature, average speed	N/S	7.43% - 13.86% (MAPE)
	Exponential								18% (MAPE)
	Mixed								7.59% - 23.00% (MAPE)
	AR								5.30% - 8.72% (MAPE)
	ANN (N/S)								7.89% - 12.55% (MAPE)
	SVM (RBF)								5.79% - 9.28% (MAPE)

(continued on next page)

Table 1 *(continued)*

	Bayesian Network								5.92% - 11.31% (MAPE)
[13]	SVM (RBF)	Non-residential	Hourly	Overall	Energy conservation	Simulated (EnergyPlus)	Day is holiday or not, weather conditions, zone mean air temperatures, infiltration volume, heat gain through each window, heat gain through lights and people, zone internal total heat gain	100 instances	~0.00 (MSE)
	Parallel SVM (RBF)								~0.00 (MSE)
[43]	SVM (RBF)	Non-residential	Hourly	Overall	Demand and supply management	Real (N/A)	Outside air temperature	7 months	50kW - 51kW(RMSE)
[21]	SVM (RBF)	Non-residential	Hourly	Lighting	Abnormal energy usage identification	Real (N/A)	Number of people in building, solar radiation	168 instances	0.66 (MSE)
	ANN (RBF)								3.14 (MSE)
[16]	SVM (RBF)	N/S	Hourly	Overall	N/S	ASHRAE dataset (N/A)	Temperature, solar flux, date, sine of current hour, cosine of current hour	6 months	3.30% (CV)
		Residential	Sub-hourly			Real (N/A)	Temperature, date, sine of current hour, cosine of current hour	~3.5 months	10.47% - 133.24% (CV)
			Hourly						2.16% - 11.30% (CV)
			Daily						5.52% - 11.39% (CV)
[32]	Lasso	Residential	Sub-hourly	Overall	N/S	Real (N/A)	Outdoor temperature, data, sine of current hour, cosine of current hour	84 days	14.88% - 86.18% (CV)
			Hourly						12.03% - 97.39% (CV)
[44]	ANN (NARX)	Non-residential	Hourly	Overall	Energy demand management	Real (N/A)	Date, outdoor temperature, outdoor humidity, solar radiation, outdoor wind speed, outdoor wind direction, state of pumps, state of boilers, state of absorption machine, state of cooling tower, state of heat pump	18 months	0.81% - 1.73% (MAPE)
[3]	ANN(PENN)	Non-residential	Hourly	Cooling	N/S	Real (N/A)	Outdoor temperature, relative humidity, rainfall, wind speed, bright sunshine duration, solar radiation, occupancy area, occupancy rate	1053 instances	11.41% - 17.17% (CV)
[45]	ANN(FFNN)	Non-residential	Monthly, yearly	Cooling, heating	N/S	Real (N/A)	Outdoor temperature, relative humidity, setpoint temperature, occupancy schedule	159 instances	N/S
[46]	ANN(GRNN)	Non-residential	Hourly	Cooling	HVAC thermal energy storage optimization	Simulated (ESP-r)	Temperature	4 years	0.91 - 0.96 (R^2)
[33]	ANN(MRAN)	N/S	Hourly	Cooling	HVAC system operation improvement	Real (N/A)	Parameters of 11 AHUs	288 hours	3.65% (MRE)
	ARIMA								9.17% (MRE)
[30]	AR	Non-residential	Hourly	Overall	N/S	Real (N/A)	Day of the week, type of day, season, wind direction, humidity, precipitation, sigma direction, sigma speed, air temperature, average speed, temperature humidity index, wind chill index	N/S	4.26% - 8.14% (MAPE)
	ARIMA								13.54% - 19.13% (MAPE)
	ANN (N/S)								3.46% - 4.11% (MAPE)
	Bayesian Network								6.87% - 22.75% (MAPE)
[47]	ANN (BPNN)	Non-residential	Hourly	Cooling	Operational planning	Real (N/A)	Air temperature, relative humidity	45 days	7.89% (Relative error)
[31]	ANN(FFNN)	Non-residential	Hourly	Overall	N/S	Real (N/A)	Outside air temperature, outside air relative temperature, boiler outlet water temperature, boiler outlet water flowrate, chiller outlet water temperature, chiller outlet water flowrate, supply air temperatures - hot duct for ahus, supply air temperatures - cold duct for ahus, supply air fan VFD control settings for ahus, return air fan VFD control settings for ahus, indoor air temperatures of different zones	A year	7.30% - 8.48% (CV)
	CBR								13.15% - 14.32% (CV)
[18]	ANN(FFNN)	Non-residential	Hourly	Overall	N/S	ASHRAE dataset (N/A)	Temperature, solar flux, humidity, wind speed, date, sine and cosine of hour of day, sin and cosine of day of week, sin and cosine of day of year	6 months	2.44% (CV)
						Real (N/A)	Temperature, humidity, date, sine and cosine of hour of day, sin and cosine of day of week, sin and cosine of day of year	A year	2.95% (CV)
[19]	ANN (feedback)	Non-residential	Hourly	Overall	Building management system operation	ASHRAE dataset (N/A)	Current and forecasted temperature, date	N/S	1.44% (CV)
						Proben dataset (N/A)		N/S	2.55% (CV)
[48]	ANN (Levenberg–Marquart)	Non-residential	Hourly	Cooling	Building daily operation	Simulated (N/S)	On/off status of compressors, temperature of water entering	9 months	4.00% - 40.00% (CV)

(continued on next page)

Table 1 *(continued)*

					optimization and control strategy selection	Real (N/A)	ice tank, temperature of water entering evaporator, temperature of water leaving evaporator, outdoor relative humidity, outdoor temperature, chilled water prepared in ice tanks or not, percentage of chilled water prepared in ice tanks, holiday indicator, date, electric current used by chiller		23.00% - 253.00% (CV)
[49]	ANN (MLP)	Non-residential	Hourly	Cooling	Energy auditing	Real (N/A)	Outdoor temperature, relative humidity, rainfall, wind speed, bright sunshine duration, solar radiation, occupancy area, occupancy rate	1053 instances	12.12% - 16.36% (CV)
[40]	AR	Non-residential	Hourly	Overall	Demand side management	Real (N/A)	N/A	18 months	7.34% - 13.78% (MAPE)
						ASHRAE dataset (N/A)		6 months	5.74% (MAPE)
						EUNITE dataset (N/A)		24 months	6.69% (MAPE)
	SVM (RBF)					Real (N/A)		18 months	7.92 - 14.25% (MAPE)
						ASHRAE dataset (N/A)		6 months	5.88% (MAPE)
						EUNITE dataset (N/A)		24 months	7.34% (MAPE)
	Poly					Real (N/A)		18 months	11.91 - 19.78% (MAPE)
						ASHRAE dataset (N/A)		6 months	6.94% (MAPE)
						EUNITE dataset (N/A)		24 months	7.36% (MAPE)
	ANN (N/S)					Real (N/A)		18 months	13.46 - 17.64% (MAPE)
						ASHRAE dataset (N/A)		6 months	6.63% (MAPE)
						EUNITE dataset (N/A)		24 months	7.78% (MAPE)
[50]	SVM (N/S)	Mixed	Sub-hourly	Overall	N/S	Real (N/A)	Outdoor air temperature, relative humidity, solar radiation, wind speed, wind direction	A year	N/S
[51]	ANN (MLP)	Non-residential	Sub-hourly	Heating	Heating load management	Real (N/A)	Outside temperature, solar radiation, work/off day, occupancy profiles, operational power level characteristics, transitional characteristics	27 days	0.15 (MSE)
[52]	ANN (MLP)	Non-residential	Sub-hourly	Overall	Demand side management	Real (N/A)	External temperature	N/S	3.16 (MAPE)
[53]	ANN (N/S)	Non-residential	Sub-hourly	Overall	Demand and supply management	Real (N/A)	Current temperature of external environment, status, building usage profile,	N/S	11.06% (MAPE)
[54]	ANN (BPNN)	Non-residential	Sub-hourly	Overall	N/S	Real (N/A)	Date, 24-hour-ahead average load, day-ahead load, 7 days-ahead load, day-ahead temperature	A year	6.97% - 11.15% (CV)
[28]	SVM (RBF)	Non-residential	Hourly	Overall	Zero energy building operation	Real (N/A)	N/A	~18 months	6.38% - 13.29% (MAPE)
						ASHRAE dataset (N/A)		6 months	4.62% (MAPE)
	AR (N/A)					Real (N/A)		~18 months	6.03% - 12.86% (MAPE)
						ASHRAE dataset (N/A)		6 months	4.63% (MAPE)
[35]	MLR	Non-residential	Daily	Overall	N/S	Real (N/A)	Maximum dry-bulb temperature, average dry-bulb temperature, minimum dry-bulb temperature, average dew point temperature, average relative humidity, average pressure, average amount of cloud, total rainfall, number of hours of reduced visibility, solar radiation, total evaporation, average wind speed	A year	4.23% (MAPE)
	ARIMA								5.45% (MAPE)
	SVM (RBF)								3.11% (MAPE)
	DT (RF)								3.17% (MAPE)
	MLP (ANN)								4.75% (MAPE)
	DT (BT)								4.07% (MAPE)
	MARS								3.97% (MAPE)
	Knn								4.01% (MAPE)
[55]	SVM (N/S)	Residential	Daily	Overall	Energy conservation	Real (N/A)	Date, outdoor temperature, bedroom temperature, living temperature, living room humidity, bedroom humidity, outdoor humidity, water temperature	15 months	0.88 (Pearson coefficient)
[56]	MLR	Residential	Daily	Overall	Demand side management	Real (N/A)	Outside temperature, date	3 years	12.36% (MAPE)
[57]	ANN (MLP)	Non-residential	Daily	Cooling	Energy-efficient building design	Simulated (EnergyPlus)	Daily average dry-bulb temperature, daily average wet-bulb temperature, daily global solar radiation, daily average	8760 instances	725 - 1410 kWh (RMSE)
				Heating					607 - 785 kWh (RMSE)

(continued on next page)

Table 1 *(continued)*

				Lighting			clearness index, solar aperture, daylight aperture, overhang, side-fins projections, date		224 - 396 kWh (RMSE)
				Overall					2118 - 2904 kWh (RMSE)
[58]	ANN(FFNN)	Non-residential	Daily	Overall	Demand side management	Simulated (EnergyPlus)	Daily minimum and maximum external dry-bulb temperature, date	54 days	10.5% - 21.0% (average error)
							Daily maximum external dry-bulb temperature, relative humidity, solar radiation, diffuse solar radiation, date		9.5% - 16.5% (average variation)
[59]	ANN (MLR)	Non-residential	Daily	Cooling	Response to climate change analysis	Simulated (VisualDOE4.1)	Temperature, humidity, solar radiation	29 years	1.4 - 1.5 MWh (RMSE)
				Overall					1.2 - 1.3 MWh (RMSE)
[23]	SVM (RBF)	Non-residential	Monthly	Overall	Energy performance contracting	Real (N/A)	Dry-bulb temperature, relative humidity, solar radiation	4 years	0.99% - 2.69% (CV)
[60]	ANN(FFNN)	Residential	Monthly	Heating	Demand side management	Real (N/A)	Monthly average external temperature, heat transfer rate through envelope, heat transfer rate through wall next to staircase, heat flow rate due to infiltration/natural ventilation, solar gain through transparent elements, internal gains, income level, occupant per room	5 years	0.83 (R)
[61]	ANN	Non-residential	Monthly	Overall	Demand side management	Real (N/A)	N/S	3 years	15.70% - 17.97% (RMSPE)
[38]	SVM (N/S)	Residential	Yearly	Overall	N/S	Real (N/A)	Average heat transfer coefficient of building walls, mean thermal inert index of building walls, roof heat transfer coefficient, building size coefficient, absorption coefficient for solar radiation of exterior walls, eastern window-wall ratio, western window-wall ratio, southern window-wall ratio, northern window-wall ratio, mean window-wall ratio, shading coefficient of eastern window, shading coefficient of western window, shading coefficient of southern window, shading coefficient of northern window and integrated shading coefficient	59 instances	2.40% (RMSE)
	ANN(BPNN)								14.46% (RMSE)
	ANN(RBFNN)								12.44% (RMSE)
	ANN(GRNN)								5.24% (RMSE)
[62]	ANN(BPNN)	Residential	Yearly	Overall	N/S	Simulated (KEP-IYTE-ESS)	Width/length, wall overall heat transfer coefficient, area/volume, total external surface area, total window area/total external surface area	148 instances	5.06 (MAPE)
[63]	ANN (MLP)	Mixed	Yearly	Overall except heating	Energy-efficient building design	Real (N/A)	Building activity, building environment, heating fuel, age, primary material, geometry data, adjacency shading data, adjacency sheltering factor, orientation, glazing, weather data	1872 instances	32.70% (CV)
				Heating					25.80% (CV)
[64]	ANN (BPNN)	N/S	Yearly	Heating	Energy-efficient building design	Simulated (N/S)	Transparency ratio, orientation, insulation thickness	135 instances	0 - 0.4 (δ)
[65]	ANN(BPNN)	Residential buildings of a city	Yearly	Overall	Supply side management	Real (N/A)	Locale (i.e., urban and rural), total population in urban areas, average number of people per household, electrification rate, penetration of device or appliance, types of lighting bulb, number of lighting bulb of type per household, power of bulb of type, hours of use of bulb of type, fuel type, lighting energy use of fuel, cooking and water heating energy use of fuel per household per year, space heating and cooling energy use of fuel, other end use devices	12 years	0.09% (MRPE)
[66]	ANN(BPNN)	N/S	Hourly	Overall	N/S	ASHRAE dataset (N/A)	Temperature, solar flux, date, sin of current hour, cosine of current hour	6 months	10.52% (CV)
	ANN(ANFIS)								9.83% (CV)
	ANN(BPNN)						Short past values of energy consumption, temperature, date, sine of current hour		3.41% (CV)
	ANN(ANFIS)								2.78% (CV)
	ANN(BPNN)	Non-residential				Real (N/A)	Daily temperature, cosine of hour of day, hourly occupancy	900 hours	5.2% (CV)
	ANN(ANFIS)								4.47% (CV)
	ANN(BPNN)						Short past values of energy consumption, daily temperature, cosine of hour of day, hourly occupancy		3.01% (CV)
	ANN(ANFIS)								2.66% (CV)

(continued on next page)

Table 1 *(continued)*

[67]	MLR	Non-residential	Hourly	Heating	Demand and supply management	Simulated (EnergyPlus)	Dry-bulb temperature, relative humidity, wind speed, direct irradiation, occupancy	6 months	173.2% - 249.8% (CV)
				Cooling					47.5% - 48.5% (CV)
	AR			Heating			Previous loads		124.6% - 185.9% (CV)
				Cooling					32.4% - 37.6% (CV)
	ARX			Heating			Dry-bulb temperature, relative humidity, wind speed, direct irradiation, occupancy, previous loads		49.6% - 178.8% (CV)
				Cooling					8.8% - 34.0% (CV)
	ANN(BPNN)			Heating					121.2% -174.9% (CV)
				Cooling					25.3% - 30.6% (CV)
	MLR	Residential		Heating			Dry-bulb temperature, relative humidity, wind speed, direct irradiation, occupancy		58.5% - 70.7% (CV)
				Cooling					17.9% - 25.9% (CV)
	AR			Heating			Previous loads		14.7% - 23.7% (CV)
				Cooling					8.3% - 8.6% (CV)
	ARX			Heating			Dry-bulb temperature, relative humidity, wind speed, direct irradiation, occupancy, previous loads		10.8% - 22.0% (CV)
				Cooling					5.0% - 6.9% (CV)
	ANN(BPNN)			Heating					23.0% - 23.3% (CV)
				Cooling					6.4% - 7.1% (CV)
[68]	ANN(BPNN)	Non-residential	Hourly	Cooling	Demand side management	Real (N/A)	Dry bulb temperature, wet bulb temperature, global radiation, wind speed, rainfall, visibility and cloud condition, operation schedule of pretreated air units, hour type, occupancy space power demand	2 years	11.01% - 11.13% (CV)
			Daily				Dry bulb temperature, wet bulb temperature, global radiation, wind speed, rainfall, visibility and cloud condition, operation schedule of pretreated air units, day type, occupancy space power demand		5.27% - 5.51% (CV)
[69]	Bivariate regression	Non-residential	Yearly	Cooling	N/S	Simulated (EnergyPlus)	Building cooling demand	23,040 instances	0.89 - 3.29 kWh/m2/yr (RMSD)
				Heating			Building heating demand		2.12 - 3.74 kWh/m2/yr (RMSD)
	Multivariate regression			Cooling			Building cooling and heating demand		0.53 - 2.41 kWh/m2/yr (RMSD)
				Heating					1.27 - 3.13 kWh/m2/yr (RMSD)
[70]	Poly	Residential	Yearly	Heating	Energy-efficient building design	Real (N/A)	Building global heat loss coefficient, south equivalent surface, difference between indoor heating set point and sol-air temperature	17 instances	0.36% (MAPE)
[71]	ANN(BPNN)	Non-residential	Hourly	Overall	Demand side management	Real (N/A)	Average temperature of three hours before time of prediction, cooling degree days, heating degree days, expected consumption in the previous hour	A year	6.31% (MAPE)
[72]	ANN(MLP)	Non-residential	Daily	Overall	Demand side management	Real (N/A)	Load history, temperature	10 months	~3.5% - 9.00% (MAPE)
[73]	ANN(FFNN)	Non-residential	Daily	Heating	Above-normal energy consumption detection	Real (N/A)	Heating consumption of previous day, mean daily outside temperature and day of week	3 years	5.25% (MAPE)
	ANN(RBFN)								5.43% (MAPE)
	ANN(ANFIS)								5.43% (MAPE)
[74]	ANN (N/S)	Non-residential	Hourly	Overall	Demand and supply management	Real (N/A)	Outdoor temperature of current and previous time, recorded energy consumption of previous time, day type, time type	3 months	10.47% (MAPE)
	SVM (N/S)								18.03% (MAPE)
	ARIMA								32.76% (MAPE)
[75]	ANN(ESN)	Non-residential	Hourly	Overall	N/S	Real (N/A)	Air temperature and building occupancy	4 years	3.72% (CV)
[76]	ANN(ELM)	Residential	Yearly	Heating and cooling	Energy-efficient building design	Simulated (EnergyPlus)	Insulation K value, insulation thickness	180 instances	74.02 kWh (RMSE)
[77]	ANN(FFNN)	Non-residential	Sub-hourly	Overall	Demand side management	Real (N/A)	Day indicator, interval stamp, HVAC operation schedule, outdoor dry-bulb temperature, outdoor relative humidity	A month	~10% (CV)
[78]	SVM (RBF)	Non-residential	Hourly	Overall	Supply side management	Real (N/A)	Electric load, temperature, calendar, school schedule, working schedule, classroom size, classroom devices, expert knowledge, occupancy	24 months - 38 months	18.05% - 18.11% (MAPE)

N/A: Not Applicable
N/S: Not Specified

MLR, ARIMA, SVM, RF, MLP, BT, MARS, and kNN; Chou and Bui [36] compared ANN, SVM, CART, CHAID, and GLR; Edwards et al. [12] compared MLR, FFNN, SVM, LS-SVM, HME-FFNN, and FCM-FFNN; Li et al. [37] and Li et al. [38] compared SVM, BPNN, RBFNN, and GRNN; Dagnely et al. [22] compared OLS and SVM; Massana et al. [39] compared MLR, MLP, and SVM; and Fernandez et al. [40] compared AR, poly, ANN, and SVM.

5.4. Performance evaluation

Model testing is the evaluation of the prediction model using some standard evaluation measures. The most commonly-used evaluation measures of energy consumption prediction models are the coefficient of variation (CV), mean absolute percentage error (MAPE), and root mean square error (RMSE). These measures can be calculated using Eqs. (1 to 3). Overall, 41%, 29%, and 16% of the reviewed studies utilized CV, MAPE, and RMSE, respectively, to evaluate their models. Other measures used for evaluating energy consumption prediction include the mean absolute error (MAE), mean bias error (MBE), mean squared error (MSE), R-squared (R^2), and error rate (δ). These measures can be calculated using Eqs. (4 to 8). CV is the most commonly-used evaluation measure probably for two reasons. First, it is one of the performance evaluation measures recommended by ASHRAE for evaluating energy consumption prediction models. Second, it normalizes the prediction error by the average energy consumption and provides a unitless measure that is more convenient for comparison purposes.

$$\text{Coefficient of Variation (CV) (\%)} = \frac{\sqrt{\frac{\sum_{i=1}^{n}(y_{predict,i} - y_{data,i})^2}{n}}}{\bar{y}_{data}} \times 100 \tag{1}$$

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

$$\text{Root Mean Square Error (RMSE)} = \sqrt{\frac{\sum_{i=1}^{n}(y_{predict,i} - y_{data,i})^2}{n}} \tag{3}$$

$$\text{Mean Absolute Error (MAE)} = \frac{1}{n}\sum_{i=1}^{n}\left|y_{predict,i} - y_{data,i}\right| \tag{4}$$

$$\text{Mean Bias Error (MBE) (\%)} = \frac{\frac{\sum_{i=1}^{n}(y_{predict,i} - y_{data,i})}{n}}{\bar{y}_{data}} \times 100 \tag{5}$$

$$\text{Mean Squared Error (MSE)} = \frac{1}{n}\sum_{i=1}^{n}(y_{predict,i} - y_{data,i})^2 \tag{6}$$

$$\text{R} - \text{Squared } (R^2) = 1 - \frac{\sum_{i=1}^{n}(y_{predict,i} - y_{data,i})^2}{\sum_{i=1}^{n}(y_{data,i} - \bar{y}_{data})^2} \tag{7}$$

$$\text{Error Rate } (\delta_i) = y_{data,i}(1 - y_{data,i})(y_{predict,i} - y_{data,i}) \tag{8}$$

where $y_{predict,i}$ is the predicted energy consumption at time point i, $y_{data,i}$ is the actual energy consumption at time point i, $\bar{y}_{data}$ is the average energy consumption, and n is the total number of data points in the dataset.

6. Discussion

6.1. Temporal granularities

Both short-term (e.g., sub-hourly, hourly, or daily) and long-term (e.g., yearly) energy consumption prediction are essential for building and grid design and operation. For example, “HVAC operations including adjusting the starting time of cooling to meet start-up loads, minimizing or limiting the electric on-peak demand, optimizing the costs and energy utilization in cool storage systems, and related energy and cost needs in other HVAC systems” all benefit from short-term energy consumption prediction [26]. Short-term energy consumption prediction models are also utilized for maintaining economic and secure operation of power grids and for providing energy consumption data to building occupants to better negotiate energy prices with energy retailers [40]. Among the reviewed literature, 84% of the studies focused on short-term energy consumption prediction because of its direct relation to the day-to-day operations of buildings [35].

Only 12% of the studies focused on long-term (yearly) energy consumption prediction. This might be caused by several reasons. First, to achieve good performance, long-term energy consumption prediction requires a relatively higher amount of data that covers a long time span [79]. For example, prediction errors of annual energy consumption prediction models, which were developed based on 1-day, 1-week, and 3-month measurements, were 100%, 30%, and 6%, respectively [80]. Second, nonlinearity in long-term data is usually more prominent compared to short-term data [81]. Third, uncertainties in long-term energy consumption prediction are usually higher because of the many changes that may occur in the supply and demand over a long time span. Long-term energy consumption prediction, thus, requires specific long-term prediction models due to the non-homogeneity and significant changes that may occur on the long-run [82]. Despite their challenges, long-term energy consumption prediction models are essential; they are required when studying decisions of long-term implications such as capacity expansion, energy supply strategy, and capital investment [83].

6.2. Building types

About 81% of the reviewed research efforts focused on developing energy consumption prediction models for commercial and/or educational buildings, with only 19% focusing on residential buildings. The relative lack of studies on residential buildings could be due to a number of reasons. First, the lack of data – specifically sensor-based data – could be a main reason. The majority, 73%, of non-residential building energy consumption prediction models rely on sensor data for algorithm training. Such data are much harder to obtain for residential buildings because the majority of buildings are not sufficiently metered in a way that allows for sensing at high granularity [10]. Another reason could be the complexity of predicting energy consumption in residential contexts because of the relatively higher variability of occupant behavior compared to the commercial context [16]. Occupant behavior is the greatest uncertainty in building energy consumption prediction [84]; ignoring, misunderstanding, and/or underestimating the role of occupant behavior in affecting energy consumption is one of the main causes for the deviations between the predicted and the actual consumption levels [85].

Despite their challenges, residential building energy consumption predictions are needed because of the high energy consumption share of this sector and the potential high gain that can be achieved if successful energy reducing strategies are implemented. Residential buildings represent 21% of the total energy consumption in the US, which is greater than the share of commercial buildings [86]. Further studies are, thus, needed on the residential sector. For example, experimental studies could be conducted to see if/how existing data-driven commercial building energy consumption prediction models could be extended to the residential context.

6.3. Energy consumption types

As discussed in Section 5, 46%, 31%, 20%, and 2% of the reviewed research efforts focused on predicting overall, cooling, heating, and lighting energy consumption, respectively. This shows a relative lack of studies on predicting lighting loads. This might be caused by the predominant impact of occupant behavior on lighting energy consumption. Lighting use is directly impacted by building occupancy and occupant behavior patterns [87]. For example, 500 lx is the recommended illuminance level for office buildings [88]. Theoretically, people who have access to natural lighting, when the outdoor illumination is sufficient, are expected to use artificial lightings less [89]. However, Yun et al. [87] showed that there are no statistically significant relationships between outdoor illuminance and artificial lighting use patterns.

Despite these reasons, lighting energy consumption prediction is essential for building energy efficiency and for efficient supply-side management. Lighting represents almost 20% of the global electricity consumption [90]. Since it is a major heat source, lighting is not only a significant piece of building energy consumption by itself, but it also impacts the cooling energy demand [77]. In general, one-third of the cooling energy consumption can be saved if a good balance between natural light and solar heat can be achieved [57]. In addition, different building design features – in terms of building envelope, architectural features, and building materials – may have different impacts on lighting energy consumption [91]. Lighting energy consumption prediction models, thus, require more attention to better understand lighting energy consumption trends and conservation opportunities, the interaction between cooling load and lighting, and the impacts of various design features on consumption levels.

7. Future research directions

Many of the research challenges discussed above can be attributed to insufficiency of data (in terms of representativeness, size, etc.) and/or complexity of occupant energy use behavior. Two future research directions are discussed in this regard.

One growing research direction is big energy data analytics. With the advent of smart meters and advanced metering infrastructure (AMI) larger sizes of monitoring data will become available. Making these data accessible to the research community may open unprecedented opportunities for researches to better understand building energy efficiency. Establishing a roadmap – including which buildings to monitor and in which locations to ensure data representativeness – could also help consolidate the many research efforts in the area of building energy efficiency, in order to eliminate duplication of efforts, provide more coverage of research questions and methods, and create a stronger research impact in the area of building energy consumption prediction. Future research directions in the area of big energy data analytics include building energy efficiency retrofitting, occupant behavior analysis, and smart energy management. For example, Mathew et al. [92] presented a vision for the potential use of big data analytics in energy efficiency retrofits. Zhou and Yang [93] proposed a vision for interdisciplinary research to analyze and understand individuals′ energy consumption behavior using big energy data analytics. Zhou et al. [94] presented a comprehensive vision for big-data-driven smart energy management, including smart power generation, power transmission, power distribution and transformation, and demand side management.

Another important research direction is behavioral energy efficiency. More efforts to capture and study occupant energy use behavior are needed to better understand how energy use behavior affects energy consumption, what the energy wasting and saving behaviors are, and how much improved behaviors can save energy. For example, Turner and Hong [95] recently proposed a framework to capture occupant energy use behavior but did not test their framework in a real-world setting. Empirical studies for capturing occupant energy use behavior and studying their impact on energy consumption are thus needed. Three sub-challenges are, however, associated with energy use behavior studies. One is the cost and time associated with real data collection, as noted above. Another is the difficulty in conducting such studies on a representative sample of occupants; behavior is highly personal and variable across different types of people and more difficult to generalize than other types of energy data. The last is the potential privacy concerns associated with tracking the behavior of occupants.

In addition to these two primary directions, future research efforts could also explore the use of other types of machine learning algorithms in energy consumption prediction. For example, deep learning algorithms have been proven to outperform other machine learning algorithms in many other fields (e.g., image classification and multi-modal data analysis [96]) but have not been sufficiently studied in the field of building energy consumption prediction yet.

As new data-driven models are developed, sharing more information about the development process and purpose, validation, and reusability of these models will be essential to avoid unnecessary duplication of research efforts. Some important model information (e.g., purpose of prediction) are sometimes not reported or not sufficiently described. Insufficient information offers limited guidance on whether certain models are applicable in a new context or not, which could inhibit the reusability of the models.

8. Limitations of data-driven energy consumption prediction and applicability considerations

Despite the importance of data-driven approaches, data-driven energy consumption prediction has two main limitations. First, data-driven prediction models may not perform well outside of their training range. Assumptions made by the learning algorithm have implications on the model's ability to cope with new data outside of the training data and whether it would generalize well beyond the training range or not [7]. For example, a model that was trained by learning from a limited dataset (e.g., data collected from a small set of buildings) may not perform well outside of the training data (e.g., different types of buildings in terms of physical properties, operation strategies, weather conditions, occupant behavior, etc.). The dataset used for training must, thus, be representative of the range of application and contain sufficient variety. Collecting such sufficiently representative and wide-ranging data may be difficult, costly, and/or time consuming [9]. It is, therefore, crucial to consider the training range when determining the suitability of using a data-driven model in a specific application. For example, using a data-driven approach for exploratory analysis of what-if-scenarios outside of the training range may be unsuitable or may be used with caution.

Second, data-driven prediction models are black-box models – their internals are not known. A black-box model may provide sufficient prediction accuracy, but may be limited in providing a detailed understanding of the different parameters and its behavior in terms of energy consumption [97].

Hybrid or grey-box modelling approaches, on the other hand, offer a combination of physical and data-driven prediction models, thereby leveraging the advantages and minimizing the disadvantages of both approaches. In grey-box models, some internal parameter and equations are physically interpretable. Grey-box models may also show better performance compared to black-box and white-box models. For example, Dong et al. [98] developed a hybrid model, which couples a data-driven model and a thermal network model, for predicting the total and non-AC energy consumptions of residential buildings and compared its prediction performance to ANN-, SVM-, LSSVM-, Gaussian mixture model (GMM)-, Gaussian process regression (GPR)-based models. Similarly, Li et al. [99] developed a hybrid improved particle swarm optimization (iPSO)-ANN model for predicting building electricity consumption. The results of both studies showed that these hybrid models offered some performance improvement.

9. Conclusions

This paper presented an overview of recent research efforts in the area of data-driven building energy consumption prediction. The scope of a set of models was reviewed in terms of building types (i.e., residential and non-residential), temporal granularities of prediction (i.e., sub-hourly, hourly, daily, monthly, and yearly), and types of energy consumption predicted (i.e., heating, cooling, lighting, and overall). The properties of the data used for training and testing these models were reviewed, including the types of data (i.e., real, simulation, and public benchmark data), the types of features (i.e., features related to outdoor weather conditions, indoor environmental conditions, building characteristics, time, occupant energy use behavior and occupancy, and historical energy consumption data), and the sizes of the data. The machine learning algorithms and the performance levels of these prediction models were also reviewed. The paper concluded with a discussion of the results, research gaps, and future research directions.

As seen from the review, data-driven building energy consumption prediction has been attracting significant research attention. Different models serve different purposes, have different scopes, were trained on different datasets, and use different features for prediction. All of the models have their own strengths and weaknesses and perform differently under different circumstances. There is no one-size-fits-all model that can be utilized under all conditions. Application-specific model development is, therefore, essential and requires case-by-case consideration of all the aspects analyzed in this paper, including data properties and machine learning algorithms.

The results of this review indicate some research areas that may require more attention: long-term building energy consumption prediction, residential building energy consumption prediction, and lighting building energy consumption prediction. The relative lack of research efforts in these areas could be attributed to insufficiency of data and/or complexity of occupant energy use behavior in these contexts. Sufficient data – in terms of types, sizes, temporal coverage, and representativeness – are essential. Capturing occupant behavior, and taking it into account, is also critical for improved energy consumption prediction. Future research directions that may lead to major improvements in these areas and beyond include big energy data analytics and behavioral energy efficiency.

Acknowledgements

This publication was made possible by NPRP Grant #6–1370-2–552 from the Qatar National Research Fund (a member of Qatar Foundation). The findings achieved herein are solely the responsibility of the authors.

References

[1] Becerik-Gerber B, Siddiqui MK, Brilakis I, El-Anwar O, El-Gohary N, Mahfouz T, et al. Civil engineering grand challenges: opportunities for data sensing, information analysis, and knowledge discovery. J Comput Civ Eng 2014;28:4014013. http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000290.

[2] Ahmad AS, Hassan MY, Abdullah MP, Rahman HA, Hussin F, Abdullah H, et al. A review on applications of ANN and SVM for building electrical energy consumption forecasting. Renew Sustain Energy Rev 2014;33:102–9. http://dx.doi.org/10.1016/j.rser.2014.01.069.

[3] Kwok SSK, Lee EWM. A study of the importance of occupancy to building cooling load in prediction by intelligent approach. Energy Convers Manag 2011;52:2555–64. http://dx.doi.org/10.1016/j.enconman.2011.02.002.

[4] Zhao H, Magoulès F. A review on the prediction of building energy consumption. Renew Sustain Energy Rev 2012;16:3586–92. http://dx.doi.org/10.1016/j.rser.2012.02.049.

[5] Esen H, Esen M, Ozsolak O. Modelling and experimental performance analysis of solar-assisted ground source heat pump system. J Exp Theor Artif Intell 2017;29:1–17. http://dx.doi.org/10.1080/0952813X.2015.1056242.

[6] Fumo N. A review on the basics of building energy estimation. Renew Sustain Energy Rev 2014;31:53–60. http://dx.doi.org/10.1016/j.rser.2013.11.040.

[7] Li X, Wen J. Review of building energy modeling for control and operation. Renew Sustain Energy Rev 2014;37:517–37. http://dx.doi.org/10.1016/j.rser.2014.05.056.

[8] Li Z, Han Y, Xu P. Methods for benchmarking building energy consumption against its past or intended performance: an overview. Appl Energy 2014;124:325–34. http://dx.doi.org/10.1016/j.apenergy.2014.03.020.

[9] Chalal ML, Benachir M, White M, Shrahily R. Energy planning and forecasting approaches for supporting physical improvement strategies in the building sector: a review. Renew Sustain Energy Rev 2016;64:761–76. http://dx.doi.org/10.1016/j.rser.2016.06.040.

[10] Wang Z, Srinivasan RS A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. Renew Sustain Energy Rev n.d. http://dx.doi.org/10.1016/j.rser.2016.10.079.

[11] Wu X, Kumar V, Ross Quinlan J, Ghosh J, Yang Q, Motoda H, et al. Top 10 algorithms in data mining. Knowl Inf Syst 2008;14:1–37. http://dx.doi.org/10.1007/s10115-007-0114-2.

[12] Edwards RE, New J, Parker LE. Predicting future hourly residential electrical consumption: a machine learning case study. Energy Build 2012;49:591–603. http://dx.doi.org/10.1016/j.enbuild.2012.03.010.

[13] Zhao H, Magoulès F. Parallel support vector machines applied to the prediction of multiple buildings energy consumption. J Algorithm Comput Technol 2010;4:231–49.

[14] Wang Z, Srinivasan RS. A review of artificial intelligence based building energy prediction with a focus on ensemble prediction models. 2015 Winter Simul Conf 2015:3438–48. http://dx.doi.org/10.1109/WSC.2015.7408504.

[15] Domingos P. A few useful things to know about machine learning. Commun ACM 2012;55:78–87. http://dx.doi.org/10.1145/2347736.2347755.

[16] Jain RK, Smith KM, Culligan PJ, Taylor JE. Forecasting energy consumption of multi-family residential buildings using support vector regression: investigating the impact of temporal and spatial monitoring granularity on performance accuracy. Appl Energy 2014;123:168–78. http://dx.doi.org/10.1016/j.apenergy.2014.02.057.

[17] Li N, Yang Z, Becerik-Gerber B, Tang C, Chen N. Why is the reliability of building simulation limited as a tool for evaluating energy conservation measures?. Appl Energy 2015;159:196–205. http://dx.doi.org/10.1016/j.apenergy.2015.09.001.

[18] Karatasou S, Santamouris M, Geros V. Modeling and predicting building's energy use with artificial neural networks: methods and results. Energy Build 2006;38:949–58. http://dx.doi.org/10.1016/j.enbuild.2005.11.005.

[19] González PA, Zamarreño JM. Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. Energy Build 2005;37:595–601. http://dx.doi.org/10.1016/j.enbuild.2004.09.006.

[20] Li Q, Meng Q, Cai J, Yoshino H, Mochida A. Applying support vector machine to predict hourly cooling load in the building. Appl Energy 2009;86:2249–56. http://dx.doi.org/10.1016/j.apenergy.2008.11.035.

[21] Liu D, Chen Q. Prediction of building lighting energy consumption based on support vector regression. 2013 9th Asian Control Conf; 2013: p. 1–5. doi:10.1109/ASCC.2013.6606376.

[22] Dagnely P, Ruette T, Tourwé T, Tsiporkova E, Verhelst C. Predicting hourly energy consumption. Can you beat an autoregressive model. Proceeding 24th Annu. Mach. Learn. Conference Belgium Netherlands, Benelearn, Delft, Netherlands, vol. 19; 2015.

[23] Dong B, Cao C, Lee SE. Applying support vector machines to predict building energy consumption in tropical region. Energy Build 2005;37:545–53. http://dx.doi.org/10.1016/j.enbuild.2004.09.009.

[24] Hellerstein JM. Quantitative Data Cleaning for Large Databases. White Paper, United Nations Economic Commission for Europe. ⟨http://db.cs.berkeley.edu/jmh/papers/cleaning-unece.pdf⟩; 2008.

[25] Pattipati K, Kodali A, Luo J, Choi K, Singh S, Sankavaram C, et al. An integrated diagnostic process for automotive systems. In: Prokhorov D, editor. Comput. Intell. Automot. Appl.. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. p. 191–218. http://dx.doi.org/10.1007/978-3-540-79257-4_11.

[26] Xuemei L, Lixing D, Jinhu L, Gang X, Jibin L. A Novel Hybrid Approach of KPCA and SVM for Building Cooling Load Prediction. 2010 Third Int Conf Knowl Discov Data Min 2010:1. http://dx.doi.org/10.1109/WKDD.2010.137.

[27] Xuemei L, Yuyan D, Lixing D, Liangzhong J. Building cooling load forecasting using fuzzy support vector machine and fuzzy C-mean clustering. 2010 Int Conf Comput Commun Technol Agric Eng 2010;1:438–41. http://dx.doi.org/10.1109/CCTAE.2010.5543577.

[28] Borges CE, Penya YK, Fernández I, Prieto J, Bretos O. Assessing tolerance-based robust short-term load forecasting in buildings. Energies 2013;6:2110. http://dx.doi.org/10.3390/en6042110.

[29] Xuemei L, Jin-hu L, Lixing D, Gang X, Jibin L. Building cooling load forecasting model based on LS-SVM. 2009 Asia-Pac Conf Inf Process 2009;1:55–8. http://dx.doi.org/10.1109/APCIP.2009.22.

[30] Penya YK, Borges CE, Agote D, Fernández I. Short-term load forecasting in air-conditioned non-residential buildings. 2011 IEEE Int Symp Ind Electron 2011:1359–64. http://dx.doi.org/10.1109/ISIE.2011.5984356.

[31] Platon R, Dehkordi VR, Martel J. Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis. Energy Build 2015;92:10–8. http://dx.doi.org/10.1016/j.enbuild.2015.01.047.

[32] Jain RK, Damoulas T, Kontokosta CE. Towards data-driven energy consumption forecasting of multi-family residential buildings: feature selection via The Lasso. Comput Civ Build Eng 2016. http://dx.doi.org/10.1061/9780784413616.208.

[33] Hou Z, Lian Z, Yao Y, Yuan X. Cooling-load prediction by the combination of rough set theory and an artificial neural-network based on data-fusion technique. Appl Energy 2006;83:1033–46. http://dx.doi.org/10.1016/j.apenergy.2005.08.006.

[34] Penya YK, Borges CE, Fernández I. Short-term load forecasting in non-residential buildings. AFRICON 2011;2011:1–6. http://dx.doi.org/10.1109/AFRCON.2011.6072062.

[35] Fan C, Xiao F, Wang S. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. Appl Energy 2014;127:1–10. http://dx.doi.org/10.1016/j.apenergy.2014.04.016.

[36] Chou J-S, Bui D-K. Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. Energy Build 2014;82:437–46. http://dx.doi.org/

10.1016/j.enbuild.2014.07.036.
[37] Li Q, Meng Q, Cai J, Yoshino H, Mochida A. Predicting hourly cooling load in the building: a comparison of support vector machine and different artificial neural networks. Energy Convers Manag 2009;50:90–6. http://dx.doi.org/10.1016/j.enconman.2008.08.033.
[38] Li Q, Ren P, Meng Q. Prediction model of annual energy consumption of residential buildings. 2010 Int Conf Adv Energy Eng 2010:223–6. http://dx.doi.org/10.1109/ICAEE.2010.5557576.
[39] Massana J, Pous C, Burgas L, Melendez J, Colomer J. Short-term load forecasting in a non-residential building contrasting models and attributes. Energy Build 2015;92:322–30. http://dx.doi.org/10.1016/j.enbuild.2015.02.007.
[40] Fernández I, Borges CE, Penya YK. Efficient building load forecasting. ETFA 2011:1–8. http://dx.doi.org/10.1109/ETFA.2011.6059103.
[41] Jinhu L, Xuemei L, Lixing D, Liangzhong J. Applying principal component analysis and weighted support vector machine in building cooling load forecasting. 2010 Int Conf Comput Commun Technol Agric Eng 2010;1:434–7. http://dx.doi.org/10.1109/CCTAE.2010.5543476.
[42] Solomon DM, Winter RL, Boulanger AG, Anderson RN, Wu LL. Forecasting energy demand in large commercial buildings using support vector machine regression. Department of Computer Science, Columbia University; 2011, [Tech. Rep. CUCS-040-11, September 2011].
[43] Paudel S, Nguyen PH, Kling WL, Elmitri M, Lacarrière B, Le Corre O. Support Vector Machine in Prediction of Building Energy Demand Using Pseudo Dynamic Approach. ECOS 2015-In: Proceedings of the 28th International Conference Effic. Cost, Optim. Simul. Environ. Impact Energy Syst., Pau; 2015.
[44] Mena R, Rodríguez F, Castilla M, Arahal MR. A prediction model based on neural networks for the energy consumption of a bioclimatic building. Energy Build 2014;82:142–55. http://dx.doi.org/10.1016/j.enbuild.2014.06.052.
[45] Yezioro A, Dong B, Leite F. An applied artificial intelligence approach towards assessing building performance simulation tools. Energy Build 2008;40:612–20. http://dx.doi.org/10.1016/j.enbuild.2007.04.014.
[46] Ben-Nakhi AE, Mahmoud MA. Cooling load prediction for buildings using general regression neural networks. Energy Convers Manag 2004;45:2127–41. http://dx.doi.org/10.1016/j.enconman.2003.10.009.
[47] Yokoyama R, Wakui T, Satake R. Prediction of energy demands using neural network with model identification by global optimization. Energy Convers Manag 2009;50:319–27. http://dx.doi.org/10.1016/j.enconman.2008.09.017.
[48] Yang J, Rivard H, Zmeureanu R. On-line building energy prediction using adaptive artificial neural networks. Energy Build 2005;37:1250–9. http://dx.doi.org/10.1016/j.enbuild.2005.02.005.
[49] Kwok SSK, Yuen RKK, Lee EWM. An intelligent approach to assessing the effect of building occupancy on building cooling load prediction. Build Environ 2011;46:1681–90. http://dx.doi.org/10.1016/j.buildenv.2011.02.008.
[50] Georgescu M, Eccles E, Manjunath V, Swindle E, Mezic I. MACHINE LEARNING METHODS FOR SITE-LEVEL BUILDING ENERGY FORECASTING AND DATA RECTIFICATION. Build. Simul. Optim. - Second IBPSA, London; 2014.
[51] Paudel S, Elmtiri M, Kling WL, Corre O Le, Lacarrière B. Pseudo dynamic transitional modeling of building heating energy demand using artificial neural network. Energy Build 2014;70:81–93. http://dx.doi.org/10.1016/j.enbuild.2013.11.051.
[52] Escrivá-Escrivá G, Álvarez-Bel C, Roldán-Blay C, Alcázar-Ortega M. New artificial neural network prediction method for electrical consumption forecasting based on building end-uses. Energy Build 2011;43:3112–9. http://dx.doi.org/10.1016/j.enbuild.2011.08.008.
[53] Kamaev VA, Shcherbakov MV, Panchenko DP, Shcherbakova NL, Brebels A. Using connectionist systems for electric energy consumption forecasting in shopping centers. Autom Remote Control 2012;73:1075–84. http://dx.doi.org/10.1134/S0005117912060124.
[54] Bagnasco A, Fresi F, Saviozzi M, Silvestro F, Vinci A. Electrical consumption forecasting in hospital facilities: an application case. Energy Build 2015;103:261–70. http://dx.doi.org/10.1016/j.enbuild.2015.05.056.
[55] Lai F, Magoulès F, Lherminier F. Vapnik's learning theory applied to energy consumption forecasts in residential buildings. Int J Comput Math 2008;85:1563–88. http://dx.doi.org/10.1080/00207160802033582.
[56] Iwafune Y, Yagita Y, Ikegami T, Ogimoto K. Short-term forecasting of residential building load for distributed energy management. 2014 IEEE Int Energy Conf 2014:1197–204. http://dx.doi.org/10.1109/ENERGYCON.2014.6850575.
[57] Wong SL, Wan KKW, Lam TNT. Artificial neural networks for energy analysis of office buildings with daylighting. Appl Energy 2010;87:551–7. http://dx.doi.org/10.1016/j.apenergy.2009.06.028.
[58] Neto AH, Fiorelli FAS. Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. Energy Build 2008;40:2169–76. http://dx.doi.org/10.1016/j.enbuild.2008.06.013.
[59] Lam JC, Wan KKW, Wong SL, Lam TNT. Principal component analysis and long-term building energy simulation correlation. Energy Convers Manag 2010;51:135–9. http://dx.doi.org/10.1016/j.enconman.2009.09.004.
[60] Popescu D, Ungureanu F. Prediction of space heating consumption in district heated apartments. Energy ASME 2013;6B. http://dx.doi.org/10.1115/IMECE2013-64422.
[61] Ferlito S, Atrigna M, Graditi G, Vito SDe, Salvato M, Buonanno A. et al. Predictive models for building's energy consumption: An Artificial Neural Network (ANN) approach. 2015 XVIII AISEM Annu Conference:1–4. doi:http://dx.doi.org/10.1109/AISEM.2015.7066836; 2015.
[62] Turhan C, Kazanasmaz T, Uygun IE, Ekmen KE, Akkurt GG. Comparative study of a building energy performance software (KEP-IYTE-ESS) and ANN-based building heat load estimation. Energy Build 2014;85:115–25. http://dx.doi.org/10.1016/j.enbuild.2014.09.026.
[63] Hawkins D, Hong SM, Raslan R, Mumovic D, Hanna S. Determinants of energy use in UK higher education buildings using statistical and artificial neural network methods. Int J Sustain Built Environ 2012;1:50–63. http://dx.doi.org/10.1016/j.ijsbe.2012.05.002.
[64] Ekici BB, Aksoy UT. Prediction of building energy consumption by using artificial neural networks. Adv Eng Softw 2009;40:356–62. http://dx.doi.org/10.1016/j.advengsoft.2008.05.003.
[65] Farzana S, Liu M, Baldwin A, Hossain MU. Multi-model prediction and simulation of residential building energy in urban areas of Chongqing, South West China. Energy Build 2014;81:161–9. http://dx.doi.org/10.1016/j.enbuild.2014.06.007.
[66] Li K, Su H, Chu J. Forecasting building energy consumption using neural networks and hybrid neuro-fuzzy system: a comparative study. Energy Build 2011;43:2893–9. http://dx.doi.org/10.1016/j.enbuild.2011.07.010.
[67] Yun K, Luck R, Mago PJ, Cho H. Building hourly thermal load prediction using an indexed ARX model. Energy Build 2012;54:225–33. http://dx.doi.org/10.1016/j.enbuild.2012.08.007.
[68] Leung MC, Tse NCF, Lai LL, Chow TT. The use of occupancy space electrical power demand in building cooling load prediction. Energy Build 2012;55:151–63. http://dx.doi.org/10.1016/j.enbuild.2012.08.032.
[69] Korolija I, Zhang Y, Marjanovic-Halburd L, Hanby VI. Regression models for predicting UK office building energy consumption from heating and cooling demands. Energy Build 2013;59:214–27. http://dx.doi.org/10.1016/j.enbuild.2012.12.005.
[70] Catalina T, Iordache V, Caracaleanu B. Multiple regression model for fast prediction of the heating energy demand. Energy Build 2013;57:302–12. http://dx.doi.org/10.1016/j.enbuild.2012.11.010.
[71] Roldán-Blay C, Escrivá-Escrivá G, Álvarez-Bel C, Roldán-Porta C, Rodríguez-García J. Upgrade of an artificial neural network prediction method for electrical consumption forecasting using an hourly temperature curve model. Energy Build 2013;60:38–46. http://dx.doi.org/10.1016/j.enbuild.2012.12.009.
[72] Jetcheva JG, Majidpour M, Chen W-P. Neural network model ensembles for building-level electricity load forecasts. Energy Build 2014;84:214–23. http://dx.doi.org/10.1016/j.enbuild.2014.08.004.
[73] Jovanović RŽ, Sretenović AA, Živković BD. Ensemble of various neural networks for prediction of heating energy consumption. Energy Build 2015;94:189–99. http://dx.doi.org/10.1016/j.enbuild.2015.02.052.
[74] Zhao D, Zhong M, Zhang X, Su X. Energy consumption predicting model of VRV (variable refrigerant volume) system in office buildings based on data mining. Energy 2016;102:660–8. http://dx.doi.org/10.1016/j.energy.2016.02.134.
[75] Shi G, Liu D, Wei Q. Energy consumption prediction of office buildings based on echo state networks. Neurocomputing 2016;216:478–88. http://dx.doi.org/10.1016/j.neucom.2016.08.004.
[76] Naji S, Keivani A, Shamshirband S, Alengaram UJ, Jumaat MZ, Mansor Z, et al. Estimating building energy consumption using extreme learning machine method. Energy 2016;97:506–16. http://dx.doi.org/10.1016/j.energy.2015.11.037.
[77] Chae YT, Horesh R, Hwang Y, Lee YM. Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. Energy Build 2016;111:184–94. http://dx.doi.org/10.1016/j.enbuild.2015.11.045.
[78] Massana J, Pous C, Burgas L, Melendez J, Colomer J. Short-term load forecasting for non-residential buildings contrasting artificial occupancy attributes. Energy Build 2016;130:519–31. http://dx.doi.org/10.1016/j.enbuild.2016.08.081.
[79] Zhao H. Artificial intelligence models for large scale buildings energy consumption analysis. Ecole Centrale Paris; 2011.
[80] Cho S-H, Kim W-T, Tae C-S, Zaheeruddin M. Effect of length of measurement period on accuracy of predicted annual heating energy consumption of buildings. Energy Convers Manag 2004;45:2867–78. http://dx.doi.org/10.1016/j.enconman.2003.12.017.
[81] Li X, Wen J, Bai EW. Building energy forecasting using system identification based on system characteristics test. 2015 Work Model Simul Cyber-Phys Energy Syst 2015:1–6. http://dx.doi.org/10.1109/MSCPES.2015.7115401.
[82] Azadeh A, Ghaderi SF, Sohrabkhani S. Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors. Energy Convers Manag 2008;49:2272–8. http://dx.doi.org/10.1016/j.enconman.2008.01.035.
[83] Ekonomou L. Greek long-term energy consumption prediction using artificial neural networks. Energy 2010;35:512–7. http://dx.doi.org/10.1016/j.energy.2009.10.018.
[84] O'Brien W, Gunay HB. Mitigating office performance uncertainty of occupant use of window blinds and lighting using robust design. Build Simul 2015;8:621–36. http://dx.doi.org/10.1007/s12273-015-0239-2.
[85] Azar E, Menassa CC. Agent-based modeling of occupants and their impact on energy use in commercial buildings. J Comput Civ Eng 2012;26:506–18.
[86] EIA. Energy Consumption by Sector. Mon Energy Rev December:29. ⟨http://www.eia.gov/totalenergy/data/monthly/pdf/sec.2_3.pdf⟩; 2016.
[87] Yun GY, Kong HJ, Kim H, Kim JT. A field survey of visual comfort and lighting energy consumption in open plan offices. Energy Build 2012;46:146–51. http://dx.doi.org/10.1016/j.enbuild.2011.10.035.
[88] ISO. Lighting of indoor work places. doi:http://dx.doi.org/10.3403/02609865; 2002.
[89] Zhou X, Yan D, Hong T, Ren X. Data analysis and stochastic modeling of lighting energy use in large office buildings in China. Energy Build 2015;86:275–87. http://dx.doi.org/10.1016/j.enbuild.2014.09.071.
[90] IEA. About lighting. ⟨https://www.iea.org/topics/energyefficiency/subtopics/lighting/⟩; 2016.
[91] Cheng S, Lei L, Zhe K Study on Lighting Energy Consumption Prediction Model of Office Buildings in Harbin,China using Diva. BS2015 In: Proceedings of the 14th Conference International Build. Perform. Simul. Assoc., Hyderabad: 2015, p. 1221–1228.
[92] Mathew PA, Dunn LN, Sohn MD, Mercado A, Custudio C, Walter T. Big-data for building energy performance: lessons from assembling a very large national

database of building energy use. Appl Energy 2015;140:85–93. http://dx.doi.org/10.1016/j.apenergy.2014.11.042.
[93] Zhou K, Yang S. Understanding household energy consumption behavior: the contribution of energy big data analytics. Renew Sustain Energy Rev 2016;56:810–9. http://dx.doi.org/10.1016/j.rser.2015.12.001.
[94] Zhou K, Fu C, Yang S. Big data driven smart energy management: from big data to big insights. Renew Sustain Energy Rev 2016;56:215–25. http://dx.doi.org/10.1016/j.rser.2015.11.050.
[95] Turner WJN, Hong T. A technical framework to describe occupant behavior for building energy simulations. Sacramento; 2013.
[96] Ooi BC, Tan K-L, Wang S, Wang W, Cai Q, Chen G. et al. SINGA: A Distributed Deep Learning Platform. In: Proceedings of the 23rd ACM International Conference Multimed., New York, NY, USA: ACM. doi:http://dx.doi.org/10.1145/2733373.2807410; 2015, p. 655–688.
[97] Amara F, Agbossou K, Cardenas A, Dube Y, Kelouwani S. Comparison and simulation of building thermal models for effective energy management. Smart Grid Renew Energy 2015;6:95–112. http://dx.doi.org/10.4236/sgre.2015.64009.
[98] Dong B, Li Z, Rahman SMM, Vega R. A hybrid model approach for forecasting future residential electricity consumption. Energy Build 2016;117:341–51. http://dx.doi.org/10.1016/j.enbuild.2015.09.033.
[99] Li K, Hu C, Liu G, Xue W. Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis. Energy Build 2015;108:106–13. http://dx.doi.org/10.1016/j.enbuild.2015.09.002.