Energy 160 (2018) 1008–1020

Contents lists available at ScienceDirect

Energy

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

Potential of three variant machine-learning models for forecasting district level medium-term and long-term energy demand in smart grid environment

Tanveer Ahmad, Huanxin Chen*

School of Energy and Power Engineering, Huazhong University of Science and Technology, Wuhan, Hubei, China

A R T I C L E I N F O

Article history:
Received 3 April 2018
Received in revised form
19 June 2018
Accepted 13 July 2018
Available online 17 July 2018

Keywords:
ANN-NAEMI
MLRM
AdaBoost
Energy prediction
Machine learning models

A B S T R A C T

Medium-term and long-term energy prediction is essential for the planning and operations of the smart grid eco-system. The prediction of next year and next month energy demand of grid station, independent power producers, commercial, domestic and industrial consumers are allowed administrators to optimize and plan their resources. To address the forecasting problems, the basic intention of this study is to propose an accurate and precise medium and long-term district level energy prediction models employing the machine learning based models which are: 1) artificial neural network with nonlinear autoregressive exogenous multivariable inputs model; 2) multivariate linear regression model; and 3) adaptive boosting model. Based on environmental and aggregated energy consumption data as the model's input and output, the load prediction interval is further classified into three main parts, 1-month ahead forecasting, seasonally ahead forecasting and 1-year ahead forecasting. Feature extraction, data transformation and outlier detection are performed through different data tests. The prediction results intimate that the intended models cannot only increase the forecasting accuracy contrasted with previous forecasting models but also produce adequate forecasting intervals in the smart grid environment. Additionally, these techniques describe an essential step-forward, consolidating the spatiotemporal use of energy inconstancies and variations of district level and strong forecasting capabilities of energy usage requirement in future perceptive.

1. Introduction

The modern cities with environment evolution has become continually obvious with 60.0% of the total world's settlement/population that is forecasted to be residing in municipalities from 2028 to 2030, spending approximately 60.0–80.0% of the total world energy generation, and almost an equitable portion of greenhouse gas emanations [1]. An effective energy usage expresses is taking frequently complicated because energy production doesn't extend follow the energy usage pattern [2]. To solve such kinds of difficulties, modern smart grids allow explications: They provide for venerable optimization and management contribute the means for versatility containing the individual buildings as well as the grid station. These economic reasons and environmental circumstances boost the energy demand, forece to build and design dependable commercial and residential buildings in the future perspective. To this point, further studies are required to concede the behavior of buildings environment, and especially to identify the intercommunication between energy consumption and indoor air quality. Numerous techniques are applied to decrease the energy usage in the building environment [3,4].

The role of the smart grid is observed as the enabler for electric companies engaging with new regulatory and economic substances. The new infrastructure of smart grid with its advanced smart metering features for energy consumer, energy measurement has the following aspects such as consumers load directly interact with the utility companies in terms of their peak load requirement, sanctioned load monitoring while new connection installation and load demand estimation in peak hours. Electric companies and independent power producers are seeking the significant feedback and income of modern technology investments that should practice new market transformation and its higher advantages.

* Corresponding author.
E-mail addresses: ahmad_tanveer@hust.edu.cn (T. Ahmad), chenhuanxin@tsinghua.org.cn (H. Chen).

https://doi.org/10.1016/j.energy.2018.07.084
0360-5442/

Developing the techniques and tools of prediction and applying the modern technology is an essential task in reaching that object and goal. Several electric companies already practice prediction to solve the present difficulties, but prediction will rise in significance due to increasing complexity of data availability and challenges by a data-rich smart grid context. Differences in rainfall, temperature, icefall, the severity and frequency of climate effects are increasing various obstacles. To solve such kinds of issues, the smart grid is a novel concept which allows two-way interactions between the consumers and electric utility companies. The role of the smart grid is developed to construct the conventional electrical power grid more efficient, stable, reliable and secure. The modern infrastructure smart grid provides the possibility for the energy consumers to perform a significant performance in their energy requirement, usage and provides further details to manage energy usage efficiently and sensibly. To optimize the demanding schedule, the precise electricity requirements and their usage load curve pattern of the customers is necessary. This is the important task where the energy prediction requirement arises into play. This study examines how efficiently the ML-based approaches can determine the energy usage requirement to improve the district level medium-term and long-term energy demand.

A better exercise is to develop on the large city's requirement estimation and designate the dormant energy sources to the diverse demanded. Consequently, it is essential to propose the prediction algorithms for city-wide building load requirement. Recent past, some models have been examined in prediction usage of the energy of building environment [5]. The scheduling process of loads applying the management of load strategy is to enhance the requirement and to decrease the energy usage [6]. Nevertheless, management of energy usage is quietly confronted in various perspectives described by Shi et al. [7]. From past decades, numerous models to determine a comprehensive explication of energy control and operation in energy networks have been developed. For illustrate, Maziar et al. [8] presents the multiple optimization algorithms (MOA) practiced in accomplishing the numerous electrical generators in a micro-grid system. Some models have been utilized to control total costs from contriving ESSs to acquire energy for the micro-grids including numerous accessible sources [9,10].

1.1. Importance of forecasting

Traditionally, system operators and electricity utilities have attempted to concede the load balancing in supply side including the energy source and reserves. The dispatchability of energy requirement as well as the relevant expense of generating energy by non-renewable sources. Besides increasing the energy requirement as a source of renewable energy in numerous areas or countries, the requirement of the energy industry will require altering its reflection to determine an adequate approach to combine this alternate energy resources into the power grid [28]. Energy usage in the building sector and forecasting is also vital for energy management, planning and energy saving [29].

Energy requirement prediction operates an essential part in the finance/economics of advanced electrical systems. Settlements and decisions are performed which consists of the suspected amount of energy usage at various time orders. Long-term energy prediction, those comprising the time span of 1 years–10 years. It is helpful for intending and planning the sources that require being possible at any period in different time and extending them if required. The power distribution, transmission and generation grids are constructed and designed based on long-term financial investigations. The medium-term energy prediction, those consists the forecasting for energy requirement from one month to 1-year period are applied for designating planning operations, resources and ascertaining the financial circumstances like as rate schedules, energy tariffs and governmental organizations. The short-term energy prediction, normally those comprising time periods of 1 h to one week, are essential in the optimization of transmission flow rate of the networks, programming of power generation, generator shutdowns, etc.

1.2. Existing forecasting methodologies

Energy forecasting modeling that can point to decreases in domestic, commercial and industrial building energy usage has been the fundamental importance of government utilities, agencies and building owners [11]. Boubacar et al., [12] developed a model which consist of the multiple resolution investigations (MRI) in the different time series from means of artificial based neural networks (ABNN) and wavelet disintegration (WD). In this circumstance, applying the erudition of forecast ability, it's conceivable to omit some elements or components, possessing low forecast ability strength, externally a negating impact the accuracy, performance of the forecasting and decreasing the mathematical complexity of the model. Aowabin et al. [13], presents a recurrent based neural networks algorithm to perform medium and long-term energy forecasting, e.g. horizon of time $\geq$ one-week, of energy usage outlines in residential and commercial buildings at 1-h intervals.

Different energy forecasting algorithms such as the artificial neural network (ANN) have further demonstrated to be accurate and efficient building energy usage forecasters. It also discusses the effect of non-linearity (NL) in different kinds of building data as shown in Ref. [14]. In his research, examined a simple algorithm consists of the artificial neural network based on a software (EnergyPlus) and arranged that two techniques were appropriate for energy forecasting [15]. Furthermore, Catalina determined that practicing the linear regression algorithms (LRA) to forecast energy usage in building sector, the more reliable and precise forecasting can be achieved. In their investigation, (270.1) various situations were examined that comprised of two simulations and regression models (RM). The modeling results explicates the highest variation can be observed of 5.10% and mean squared error of 2.0% within the simulated and predicted results. However, providing the application of linear regression algorithms (LRA) as a solid approach for forecasting energy usage in building sector, the better performance can be achieved [16]. In Refs. [37,38], for short, medium and long-term energy demand forecasting, the supervised based model is proposed.

Gunay et al. [17] proposed artificial neural network algorithm applying cooling, heating electrical load in five distinct kinds of buildings. In his study, eighteen algorithms classed are developed. For every one of five-building subsequent that one algorithm is decided to demonstrate how black-box algorithms (BBA) can be applied in decision-making methods (DMM) and process in the building sector. In the recent past, various well-recognized models in the area of machine-learning (ML) like as AdaBoost and support vector machine (SVM) are analyzed in Ref. [18]. Though, such kind of works has only concentrated on binary decision learning (BDL) [19]. In evidence [20,21], the authors practiced the AdaBoost algorithm to the reliable and accurate prediction of wind for prominent attention to the integration and wind power generation.

Lin et al. [40] developed an algorithm, a new sparse AdaBoost is proposed to overcome the ensemble cost and increase the generalization capability. Furthermore, an echo-state system is adapted to increase the nonlinear correlations between multiple factors and electricity requirement. AdaBoost model is the most general method to decrease the deficiencies, that intends to connect various models and improve the generalization capability.

Ensemble-based approaches, such as AdaBoost model have been used to determine various energy usage prediction tasks [41,42]. AdaBoost model frequently trains basic variables with weighted validation samples set, therefore, it can handle the contingency of algorithm by enhancing data difference [43,44].

Results collected by this research designate the developed hybrid algorithm (HA) can overwhelm the non-linear (NN) features of wind speed (WD) in time-series. Furthermore, it can render more precise prediction outcomes contrasted with the different models. Mingjie et al. [22] developed an effective Bayesian structure. It can be implemented with a multivariate linear regression approach (MLRA) which is formed and adopted as the substitute of large accuracy physical based methods in Markov Monte Chain Carlo (MMCC) data sampling. In Refs. [23,24] multivariate analysis (MA) and non-linear multivariate regression (NLMR) are presented for electrical illumination or lightning energy requirement for a general room including distinct building energy variables and the architectural features investigations. In this research [25–27,39], the authors developed the new data mining models (DMM) that comprise of six techniques for forecasting precise and accurate future cooling and heating energy requirement of the water source heat pump (WHSP) and non-technical losses amidst the intention of improving the forecasting accuracy, energy theft detection, decreasing error and the energy management in future preceptive. However, the model's output results further explained that the six algorithms were effective and efficient in energy forecasting and foreseeing the irregular performance and future heating and cooling energy requirement in the building sector.

In this study, the three-machine learning based models applied that are: 1) artificial neural network with nonlinear autoregressive exogenous multivariable Inputs (ANN-NAEMI); 2) multivariate linear regression model (MLRM); and 3) adaptive boosting model (AdaBoost). The ANN-NAEMI model is applied in previous research but the results of this research are better and implemented with the different task of energy forecasting with limited data information for medium and long-term energy forecasting. The MLRM and AdaBoost model are proposed to predict the energy requirement in future preceptive. The MLRM and AdaBoost models have existed in previous studies with different tasks such as above detailed.

ANN-NAEMI algorithm consistently well performed for providing different kinds of tasks such as geographical inundation projections with higher correlation and lower prediction errors. The other major benefit of ANN-NAEMI algorithm is their strength to model non-linear data associations. Moreover, it provides the fast accuracy and efficiency of the system. ANN-NAEMI algorithm takes benefits of exogenous knowledge to obtain greater results with limited past data. The disadvantage is that the ANNAEMI algorithm is a non-minimal description method. The basic benefits of MLRM model is that it provides various predictor parameter which is the component of the regression. Including this versatility, it includes as various parameters which combining several independent parameters which doesn't enhance the condition of the algorithm but reduce it. The AdaBoost model is constructed to reduce the limited generalization capability and instability of individual model and decrease the AdaBoost computational complexity into its inadequate ensemble approach. AdaBoost model is quite sensitive to outliers and noisy data. In some circumstances, it is lower sensitive to overfitting problems then the different other learning models.

1.3. Contributions

The energy requirement for a considerable large system is a non-consistent with multiple intricacies. The precise variety of energy usage compares to an innumerable amount of outlines that analytically cannot be determined and indicate specific determinations on energy consumption of large to small loads which construct its exercise and research of the stochastic environment. It is essential to practice the first move by resembling toward the nature of data and forming some determinations concerning its try to adjust existing methods and behavior to algorithm it. To mitigate the issues of instability and to increase forecasting efficiency and accuracy this study intends the algorithms of getting precise and accurate medium and long-term load forecasts for district-level by using the ANN-NAEMI, MLRM and AdaBoost algorithm in the intermittent form including climate-related parameters applied as the exogenous inputs of the models.

The primary intention of energy consumption forecasting is for utility companies or electricity management companies, industrial and commercial customers by deciding related dates comprises on historical climate and energy consumption data. This study also contributes a correlation between recommended energy prediction algorithms as well as the former or previous research. Amongst other uses, the developed algorithms can also be practiced for real-time purposes like as heating, ventilation and air conditioning systems (HVAC) fault diagnosis and detection. This study also intends to assist and fulfil the gap on medium term and long-term prediction by concentrating on estimate accuracy or performance. It also describes the modeling results for monthly, seasonally and yearly load prediction. Furthermore, the forecasting results are compared with the existing studies for model validation, forecasting performance as well as accuracy.

The rest of the manuscript has the following parts. Section-2 represents the detailed methodology of ML algorithms. Section-3 display the developed energy prediction algorithms as well as the performance evaluation statistics. Section-4 & 5 explicates the selection of variables for training and testing, modeling results, forecasting error and model validation with previous studies. Section-6 conclude this study.

2. Methodology of energy demand forecasting models

2.1. Schematic outline of present study

Fig. 1 demonstrates the schematic overview of the current research. Two-year actual energy consumption data obtained including 30-min intervals are selected for energy forecasting analysis. The preparation of the data comprises into three principal responsibilities/tasks, e.g., feature extraction, data transformation and formulation input candidate pool. The recognition of irregular building usage of energy patterns is obtained by applying outlier detection and clustering analysis. Three ANN-NAEMI, MLRM and AdaBoost algorithm is practiced for energy forecasting.

2.1.1. Testbed and performance metrics

The data of energy usage in this research is obtained from aggerated of various residential, commercial and industrial sector from ISO NEW-England. ISO New England (ISO-NE) is an independent (Regional Transmission Organization (RTO)), serving Maine, Massachusetts, Connecticut, Rhode Island, Vermont and New Hampshire [30]. Relative humidity, outdoor air temperature and global solar radiation are obtained by an environmental location near the station. The ambient temperature, wind speed and relative humidity are used to construct the climate components or factors for the initial prediction are applied due to the uncertainty in degree-days rises as time flies.

2.1.2. Data transformation and outlier detection

Outliers detection are considerations that arise to be incongruous including the rest of a particular set of data. Outliers

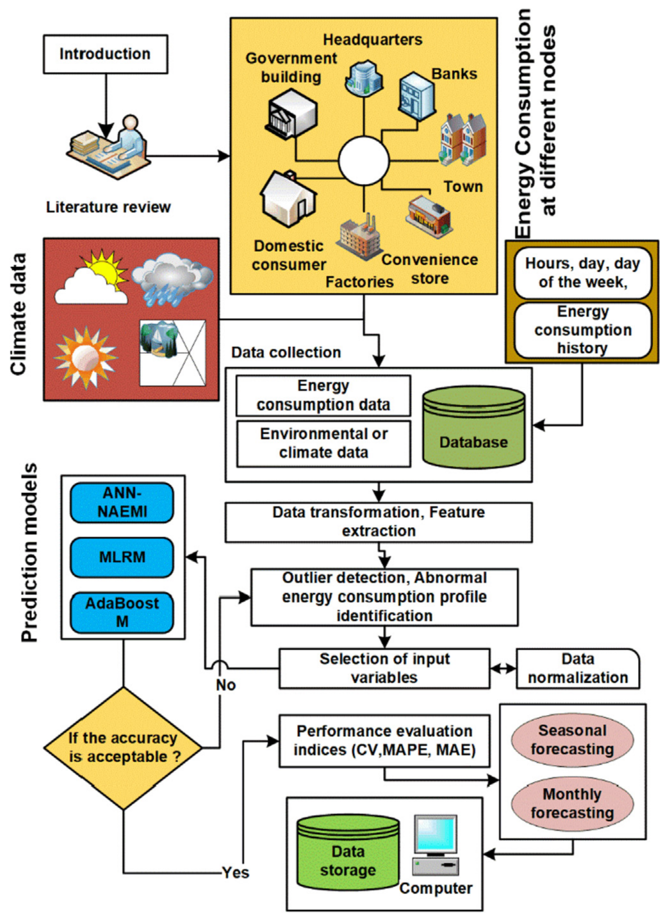

Fig. 1. Proposed methodology of the present work.

detection may occur because for different purposes, like as error of human, modification of system behavior and error in the instrument. Amongst the present models, the generalized studentized extreme deviate model (SEDM) was extremely justified due to of its adaptability under multiple circumstances. It is also executed in identifying the anomaly in energy usage in the building data and demonstrated to be computationally effective in managing extensive data of the building sector [31].

2.1.3. Feature elimination

The input variables selection to a forecasting algorithm is extremely significant when a large amount of inputs are higher, and the forecasting models are complicated as well. It accommodates to decrease the uncertainty of over-fitting, modestly enhance the algorithm, decrease the computation costs and recognize the inherent dimensionality of an assigned predicament. In customary, two usually practiced methods are selected to the algorithm inputs. The first one is comprised of principal component analysis (i.e., feature reconstruction). From protruding toward the initial some principal objectives, a distinct data set including lower dimensions is accomplished within the combination of linear of actual data. One limitation of the before-mentioned model is that the unavailability of actual set of data can be discarded and it might be challenging to render the various inputs. The second method relies on the theory or concept of sub-set determination. Generally, practiced algorithms can be moreover incorporated into the wrapper filter and embedded techniques. The filter model assigns the parameters in the term to specific metrics, like as the correlation of coefficient Pearson analysis. The deprivation or demerits prevails in the superfluity of the data sub-set elected. The rapper technique (RT) assesses the application of data sub-sets from examining particular learning models. Since massive researches of data sub-sets are presented and the wrapper approach may have an exciting development in estimate of the computation costs. This research uses machine learning based models, to choose the data inputs to forecasting algorithms.

2.1.4. Predictive models and performance evaluation indices

In this part, a concise overview of the forecasting models is provided. In total, three models, e.g., ANN-NAEMI, MLRM and AdaBoost algorithms are applied to outline the correlation between outputs and inputs. These models are chosen comprised of two principal deference, e.g., diversity and accuracy. All the decided models have been extensively practiced in interpreting complicated forecasting, modeling and predicaments, and their appearances or performances are recorded to be promising. Additionally, they are chosen to enhance the diversity ensemble, that is the advantage to the accuracy and performance. The detailed overview of the selected models of energy forecasting is presented in Section 3.1 and the performance evaluation indices are briefly discussed in Section 3.2.

2.2. Data pattern and analysis

Fig. 2 shows the scatter plots for dry bulb temperature (DBT) vs. Load and wet bulb temperature (WBT) vs. Load. It can be observed that when the comfort zone temperature is low the usage of energy demand is lower. With increasing the temperature from a particular boundary, the energy usage is also increased. It is because when the outdoor ambient temperature is quite higher, the usage of energy is further high because of the usage of air-conditioning load. The 'zone comfort' is encompassing 22 (°C), when occupants doesn't use the heating or air-conditioning load. Fig. 2 also illustrate the usage of energy at various hours in a day. The load data statistical properties and characteristics are presented in the box plots of Fig. 3. Regard the monthly load characteristic, e.g. variation month to month; the specific decrease in the energy requirement from the weekends, and the abnormality or difference in the energy usage accord to the days in a month and hours in a day.

Additional parameters where similarities could be observed including the aggregated load are periodically parameters: the month in the year (January, February, etc.), that could be specified from the mutable month. Day in a week, due to the energy usage diversifies during the week comprising on the type of day, that could be expressed from the numbers 1 (Thursday) through 7 (Wednesday). The hour in a day, due to the usage of energy also alters relaying on actually, what time it is now: late-night is

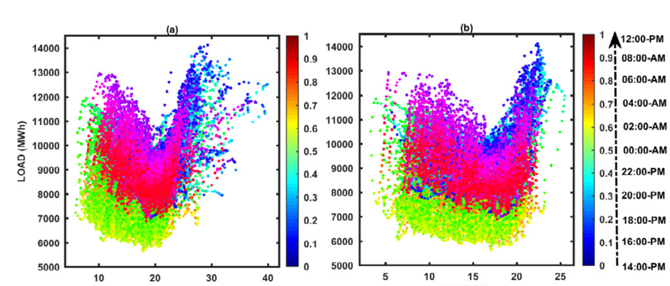

Fig. 2. Impact of DBT (°C) and WBT (°C) on energy consumption (MWh) at different hours in a day.

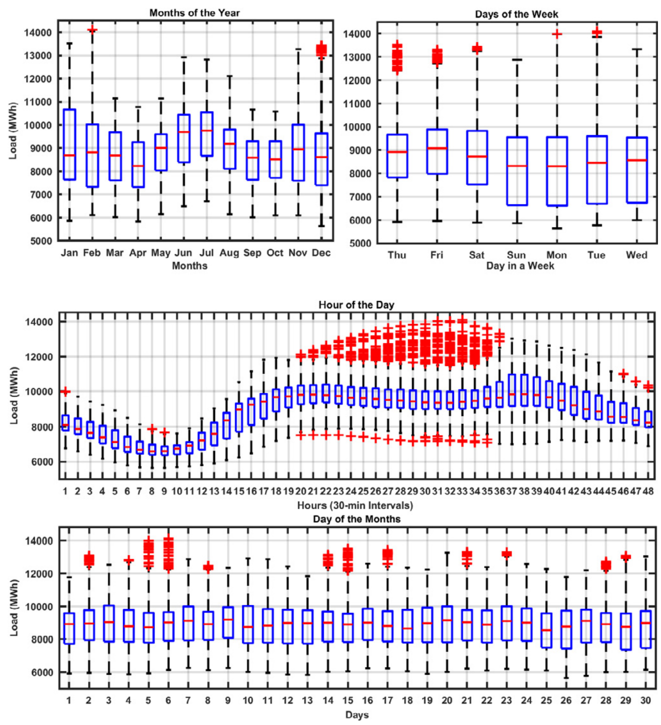

Fig. 3. Electrical load data statistical characteristics at different levels.

normally a least when most of the peoples sleep and there is limited usage of lightning and appliances. In this study, represent the hour in a day from 1 to 48 h. The climate and energy usage data show the 30-min intervals set, and 1 to 48 data samples indicates to 24-h of the day. Also, either day is a work day leave an influence on the requirement of energy: vacations conduce to have a remarkable low energy usage rather than working hours of the day. Fig. 4 describes the effects of the correlation investigation of environmental parameters. The linear correlation among the DBT with dew point temperature (DPT), WBT and ambient relative humidity (ARH) is obtained applying linear regression analysis. The correlation investigation designates that DPT vs WBT has "very strong" correlation. An in-depth analysis of correlation of particular problem demonstrates the robustness of developed ability and its methodology to apprehend the strength of the relationship within input parameters, when the similar recommendation or reference input set of the data is examined in diverse environmental circumstances.

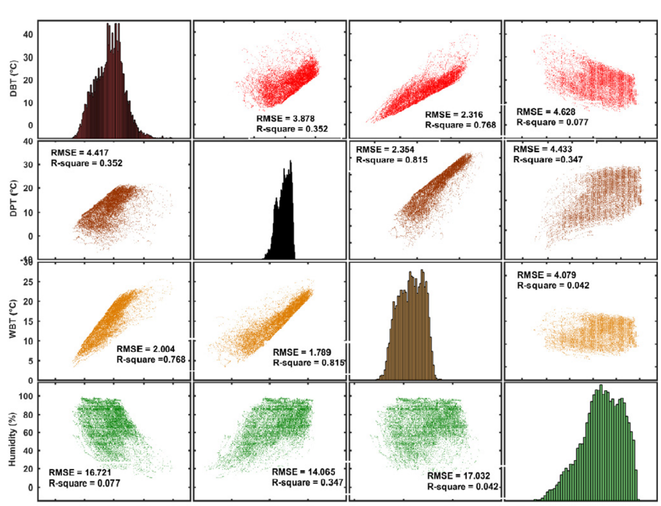

Fig. 4. Linear correlation between climate variables.

3. Proposed energy forecasting models

3.1. Energy forecasting models

Several energy prediction models employ a diversity of high-level approaches to transform the data usage into prevailing long-term, medium-term and short-term prediction algorithms. The algorithms applied in this research are developed an accessible construction stage which efficiently combines with the prediction outcomes and current systems in the future scenario. The particular composition and structure of energy forecasting models are presented below.

3.1.1. The ANN-NAEMI model

The ANN-NAEMI algorithm is an intermittent dynamic model, including the feed-back relationships embedding different kinds of the layer of the model. The ANN-NAEMI algorithm is comprised of the type of linear NAEMI algorithm, that is usually practiced in time series sequence and the modeling. The ANN-NAEMI algorithms can determine to forecast single time sequence provided prior (past) amounts with the similar series of time and the different feedback inputs designated the exogenous or external time range. The ANN-NAEMI based neural networks are practiced determining a simplistic series time problems and energy requirement in different future perceptive.

The ANN-NAEMI algorithm is also applied to forecast the next hour output, an ahead time step of during it will truly resemble. Fig. 5 determines the terminology of a hidden three-layer ANN-NAEMI model, wherever the symbol of hat (^) is utilized to indicate predicted amounts (or functions). The ANN-NAEMI model is an essential aspect of nonlinear discrete-time methods which can be mathematically depicted as

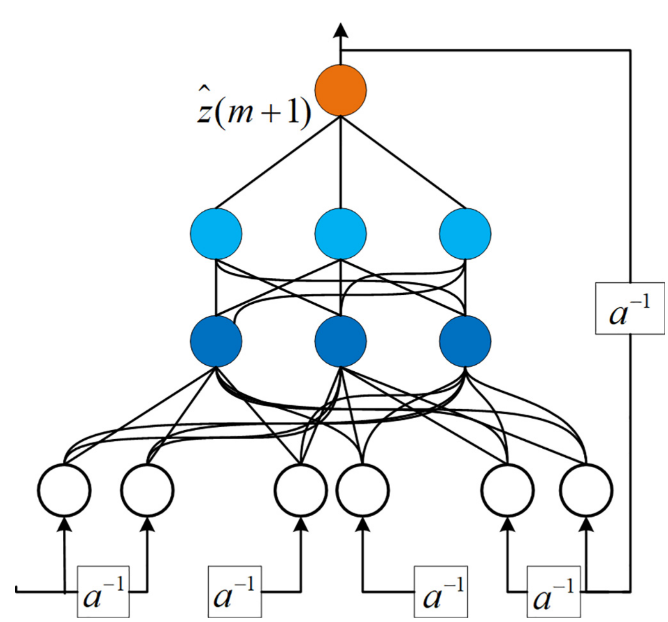

Fig. 5. ANN-NAEMI model overview with $\hat{a}(-1)$ unit time delay.

$$z(o+1) = g[z(o-e_z+1); v(o-l), v(m-l-1), \ldots, v(m-l-e_v=1)], \tag{1}$$

Where $v(o) \in S$ and $z(o) \in S$ denote, sequentially, the model input as well as output at discrete span time stem o, during $e_{v\geq 1}, e_{z\geq 1}$ and $e_v \leq e_z$ are the memory input and memory output methods, sequentially. The variables $l(l \geq 0)$ is an obstruction cycle, identified as the dead-time ahead process. Externally absence of generality or observation, perpetually pretend $l = 0$ in this study, thus getting the following ANN-NAEMI algorithm:

$$z(o+1) = g[z(o), \ldots, z(o-e_z+1); v(o), v(o-1), \ldots, v(o-e_v+1)] \tag{2}$$

which may be composed in form of vector-like as

$$z(o+1) = g[z(o); v(o)] \tag{3}$$

wherever the different vectors $z(o)$ and $v(o)$ indicate the input and output regressors, sequentially. There are several statements and applications for the ANN-NAEMI algorithm. It can also be practiced as a forecaster to predict the next condition for the diverse input signals. It is also employed for different non-linear data filtering in where the targeted output is free from the noise of input signal perceptive. The effectiveness of ANN-NAEMI algorithm is exhibited in the modeling of the nonlinear dynamic system and different significant application.

3.1.2. The MLRM model

The MLRM modeling is practiced for forecasting the amount of one or more than one replies by a set of forecasters. It is also being applied to determine the linear correlation among the forecasters, forecasting and response could be categorical consecutive or a composite of two different construction. The MLR algorithm to the circumstance where the calculated n responses $Z_1, Z_2, \ldots, Z_q$ and the similar set of the data of s forecasters $a_1, a_2, \ldots, a_s$ as on every sampling unit. Each variable response reflects its individual regression algorithm.

$$\begin{bmatrix} Z_1 = \gamma_{01} + \gamma_{11}a_1 + \ldots + \gamma_{s1}a_s + \varepsilon_1 \\ Z_1 = \gamma_{01} + \gamma_{11}a_1 + \ldots + \gamma_{s1}a_s + \varepsilon_1 \\ \ldots \\ \ldots \\ Z_1 = \gamma_{01} + \gamma_{11}a_1 + \ldots + \gamma_{s1}a_s + \varepsilon_1 \end{bmatrix} \tag{4}$$

$\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_q)$ has expectation 0 and matrix variance $\sum_{q\times q}$ The errors ε correlated with various kind of responses on the similar unit may have unconventional variations and may be associated or connected each other. Where $\gamma_{(j)}$ are the $(s+1)$ coefficients of regression in the algorithm for the ith parameters. Assume that practicing a size of sample o, as earlier, the configuration matrix A has different dimension $o*(s+1)$. But now:

$$Z_{o*n} = \begin{bmatrix} Z_{11}\ Z_{12}\ldots Z_{q1} \\ Z_{21}\ Z_{22}\ldots Z_{q2} \\ \ldots \\ \ldots \\ Z_{o1}\ Z_{o2}\ldots Z_{oq} \end{bmatrix} = \left[Z_1\ Z_2 \ldots Z_{(q)}\right] \tag{5}$$

where $Z_{(j)}$ is the supprotive vector of o measures of the ith parameter. The multivariate linear regression algorithm is different by the multiple linear regression algorithms that continuous univariate response as a combination of linear with exogenous courses, an identically distributed error and independent term. The MLR algorithm allows investigating the comparative impacts of these autonomous parameters. These parameters are subservient/dependent or predictor or parameter, those normally consolidated sets of data can influence to false determinations if they aren't investigated accurately.

3.1.3. The AdaBoost model

The Adaboost algorithm is common successful recognizing models in the area of machine learning. It consolidates many vulnerable classifiers to establish a powerful classifier. It is basically displayed for feature issues of regression and classification. Because of its superior classification efficiency and performance, it is commonly utilized in the reproduction of processing and various responsibilities or tasks as well [32]. The Adaboost algorithm's computational steps can be defined to apply the following equation:

The selection of the datasets for training $\{Y_j\}(j = 1, 2, 3, \ldots, o-1, o)$. The weights samplings $\{E_u(j)\}$ of the usage of energy in series $\{Y_j\}$ can be estimated as follows:

$$E_u(j) = \frac{1}{n}, (j = 1, 2, 3, \ldots, o-1, o; u = 1, 2, 3, \ldots U) \tag{6}$$

wherever o is the measurement of the series of energy usage $\{Y_j\}$ and U is the amounts of multi-layer perceptron comprised on energy forecasting. The forecasting I_u is used to predict the energy usage $\{Y_j\}$. Determine the prediction error ε_j of each sampling rate in series $\{Y_j\}$ as well as corresponding to the overall forecasting error ε_u can be described as follows:

$$\begin{cases} \varepsilon_j = \dfrac{\left|Y_j - \widehat{Y}_j\right|}{Y_j}, \ j = 1, 2, 3, \ldots, o-1, o \\ \varepsilon_u = \dfrac{1}{o}\displaystyle\sum_{j=1}^{o} \varepsilon_u \end{cases} \tag{7}$$

To measure the weights of series for the constituted forecaster I_u practicing the following equation as detailed below:

$$V_u = \frac{1}{2} ln\left(\frac{1-\varepsilon_u}{\varepsilon_u}\right) \tag{8}$$

AdaBoost is also an adaptive model in the knowledge domain that consequent type classifiers combined at every step of boosting which are twitched in favour of these occurrences or instances misclassified from the earlier/previous classifier.

3.2. Performance evaluation indices

In this study, three indices are used to estimate the efficiency and performance of the models. These performance evaluation indices are mean absolute percentage error (MAPE), the coefficient of variation (CV) and mean absolute error (MAE). The MAPE refer to a scale independent scale, proposing a straight-forward technique to outline accuracy of the models [25].

$$MAPE = \frac{1}{t}\sum_{i=1}^{t} \frac{\left|Q_i - \widehat{Q}_i\right|}{Q_i} \tag{9}$$

where Q_i is the net measurement; $\widehat{Q}_i$ is the forecasted value; and t is the total number of measurements. The CV signifies the variation amid actual and the predicted energy demand of the system [25].

$$CV = \frac{\frac{1}{M}\sqrt{\sum_{i=1}^{M}(R_i - P_i)^2}}{\overline{P}} \tag{10}$$

where M is the total number of data samples or measurements, R_i is the target value, P_i depicts the predicted energy demand and $\overline{P}$ is the mean of the energy predicted value.

The MAE is a scale-independent metric that proficiently reflects the forecast mistake by averting the balance amid positive and negative inaccuracies [25].

$$MAE = \frac{\sum_{i=1}^{t}\left|Q_i - \widehat{Q}_i\right|}{t} \tag{11}$$

where Q_i is measure energy demand, $\widehat{Q}$ is the forecasted value and t is the total number of measurements.

4. Parameters selection for load forecasting

The primary intention of medium and long-term energy prediction is to forecast the load in future perceptive, for illustrate the year ahead energy forecasting or next season load forecasting etc. The system total load is observed at the generation end of the electrical power system operation, that comprises the total of every kind of loads combined to the operation of the system as well as the system losses. To design the efficient and accurate forecasting model must become the immeasurable knowledge of the attributes of the electrical system. There are numerous circumstances which affect the behavior and performance of the customer energy usage. These parameters can be characterized as the weather as well as time factor. In this research study, certain circumstances as well as their influence on the usage of energy and their importance in medium and long-term prediction is estimated.

4.1. Main factors affecting energy demand

The most significant and important determinant is the peak load hours in medium and long-term energy prediction due to its direct influence on independent power producers and utility companies is higher. By using the load curve of various customer profiles, the consumer load pattern, time of the day, days of the week, peak load demand, weeks in a month and month in a season are determined. Climate features are the common and important parameter for energy prediction. The impact of climate is most notable for agricultural, domestic, and commercial customers, although it could also modify the profile of a load of industrial customers. Energy prediction models are applied at different circumstances to forecast the load in future perceptive, therefore to depreciate the running operational expense or cost, a study conducted in reference in Ref. [33]. Climate conditions are frequently referred as the point of tipping which can create unreliability in the electrical system from reducing the effective and efficient supply of power system. Unforecasted breeze sea, subsequent moon downpours, backdoor disguises/fronts are fascinating the climate parameters which can reduce the outdoor ambient temperature and therefore cause exaggerated load prediction [34]. Furthermore, under prediction of winter and summer temperature can affect load fluctuation (swing) in the building environment [35].

Ambient temperature can further change the transmission line's conductivity. Therefore, the temperature can influence the current carrying capacity of the power system transmission and distribution lines as well [45–47]. Huge outdoor temperature can enhance not only the impedance and resistance of distribution and transmission system, although it can change the line reactance, because to the temperature-influencing increase of the expansion of transmission lines. The climate impact that leaves direct impact on energy forecasting is the temperature (dry bulb, dew point and wet bulb temperature), precipitation, humidity, wind speed, cloud cover, wind chill index, solar radiation and light intensity. The ambient effect throughout the summer accretion in temperature will rise in increase in energy usage. The reduction in ambient temperature will affect the reduction is not in mean everyday energy usage but further will decrease the peak load demand.

Therefore, ambient humidity ratio can enhance the perception of the austerity in ambient temperature and conceive the occupants to utilize further cooling devices. However, because of this experience daily curve of the load will determine immense amount throughout the humid days. The significance of snow, rain or hail at a particular position within a particular interval of time designated precipitation. The precipitation might change the energy usage indirectly and directly. Wind speed and wind direction is influenced by three basic circumstances: i) the gradient of temperature also seldom involved gradient pressure; ii) fractional effectiveness i.e. mountains, forest, and buildings etc; iii) rotation of earth is perceived as Coriolis impact. Therefore, throughout a windy day in the summer season, the energy usage will be low due to subordinate cooling apparatuses will be practiced. The cloud cover effect on energy usage confides simultaneous the thickness and height of time of the day and clouds. The thin and high cloud has the remarkably limited influence on ambient temperature particularly at night-time, so it could be neglected for short-term energy prediction.

The further significant parameter is the thick and low cloud that has the significant influence on dry air temperature. The horizontal and vertical solar radiation drops direct influence on energy usage. In the summer period, the horizontal solar radiation has the higher influence in energy consumption and the winter period it decreased because the change in temperature. However, in this study, four environment variables DBT, DPT, WBT and ambient humidity ratio are applied for future load forecasting and their descriptive statistics are shown in Table 1. The substantial use of environmental variables for load forecasting have the ability to increase the forecasting accuracy but the intended algorithms in this study have the ability the to predict the accurate energy performance applying the limited data information as the model input.

4.2. Training and validation data sets

The 70% of climate and energy usage data is applied for training and rest 30% for model validation. The datasets are applied for training and testing of ANN-NAEMI, MLRM and AdaBoost algorithms. Fig. 6 demonstrates the training and testing datasets of the models. Climate and energy consumption data are acquired at 30 min interval for modeling analysis of energy forecasting. Climate and energy consumption data are further divided into three parts that are: 1) 1-year ahead forecasting; 2) seasonally forecasting; and 3) 1-month ahead forecasting. A total of 1440 data samples for 1-month ahead forecasting, 4368 data samples for seasonally

Table 1
Descriptive statistics of climate parameters.

Descriptive statistics	DBT (°C)	DPT (°C)	WBT (°C)	Humidity Ratio (%)
Mean	18.573	11.795	14.966	67.279
Standard Deviation	4.817	5.486	4.169	17.406
Minimum	5.910	−7.010	3.351	10.631
Maximum	39.920	21.839	25.531	99.014

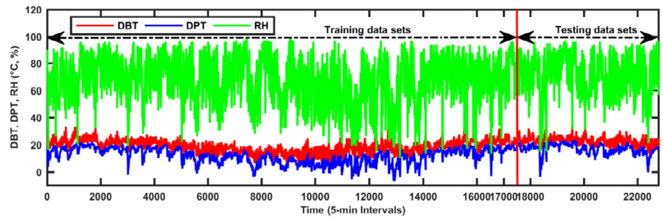

Fig. 6. Training and testing data sets of forecasting models.

forecasting and 17520 for 1-year ahead forecasting were obtained. The testing data set starts from January 1, 2010, to April 30, 2010, and training sets are selected from January 1, 2009, to December 31, 2009, for 1-year ahead energy forecasting. According to different kinds of literature, the dataset for testing should be independent than the training dataset. For an algorithm fit the training data sets can also be applied the data set for testing but a reliable fitting of the test dataset, as opposed to the training dataset, normally shows to overfitting.

Several methods are reported in different studies where the researchers are applied the training for testing the forecasting models. In this regard, it's the better way to validate the models with yearly ahead or seasonally ahead data sets to check the forecasting performance and accuracy. It is the better way to check the performance index of the models instead if we apply the same training data for validation. By applying similar data sets for testing and training, we can decrease the impacts of data errors, but the actual forecasting results and accuracy might be affected. Table 2 shows the testing and training datasets for the model testing and training.

5. Modeling results and discussion

5.1. Forecasting results

To evaluate the efficiency and performance of the energy prediction models, the environmental and energy consumption data is further divided into three classes: i) month ahead load forecasting; ii) seasonally forecasting; and iii) one year ahead load forecasting. The forecasting performance is assessed applying three performance evaluation indices MAE, MSE and MAPE. The MAE is scale-dependent metrics however, the CV and MAPE is scale independent. Scale-dependent indices render a straight-forward approach to estimate the forecasting error though scale-independent indices are helpful for evaluation of model's performance.

5.1.1. Monthly forecasting results

Fig. 7 presents the response of time series of ANN-NAEMI model. This demonstrates the targets, inputs and errors versus time plot. It further shows that time limits were chosen for testing, training and validation. In the validation, training and testing data points the errors are adequate and acceptable. The settled residuals from January 2009 to December 2009 were predicted amidst the envisioned ANN-NAEMI algorithm and it is demonstrated that the algorithm is appropriate for energy prediction.

To define quantitatively of algorithm, the identical measures of error of the three algorithms are enumerated in Table 3. It could also be perceived that contrasted among the other two algorithms, the LSBoost algorithm can produce the lower error in the statistical form of CV, MAPE and MAE. The outcomes concentrate on the forecasting accuracy of the single-valued period forecasting. Normally, the ANN-NAEMI, MLRM and the AdaBoost model can only present the point evaluations of future energy prediction. The forecasting results demonstrate that the models are impressive in forecasting large-scale energy usage and requirement in future preceptive. The model's results designate that AdaBoost can attain the immeasurable performance, including MAPE 0.009% and 0.007% of July and November respectively for forecasting the one-month ahead energy forecasting.

The minimum and the maximum MAPE of ANN-NAEMI is 1.651%

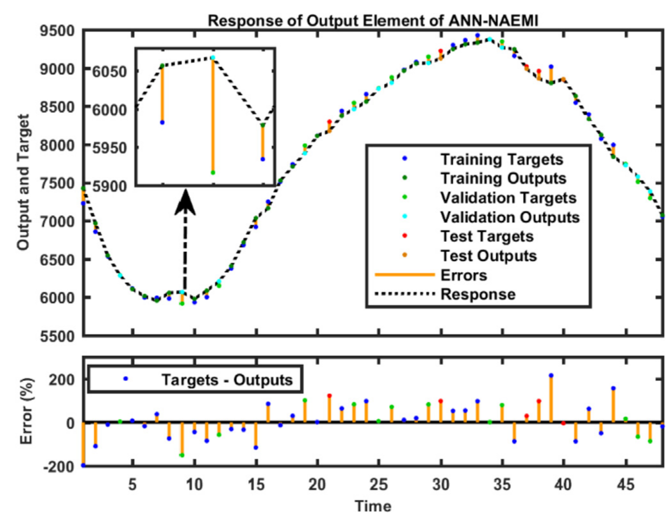

Fig. 7. Time series response with percent at error at training, testing and validation mode of ANN-NAEMI.

Table 2
The training and testing state of forecasting algorithms.

Algorithms	The testing and training of the models for one-month ahead forecasting							
	Training state				Testing state			
ANN-NAEMI	From	To	total data samples		From	To	total data samples	
MLRM	1, Jan. 2009	30, Jan. 2009	daily:60	1-h:1440	1, Jan. 2010	10, Jan. 2010	daily:28.00	1-h:480
AdaBoost	The ratio of training data samples: number of test data samples							3:1
	The testing and training of the models for 1-year ahead forecasting							
	From	To	total data samples		From	To	total data samples	
	1, Jan. 2009	31, Dec.2009	Daily:730	1-h:17520	1, Jan. 2010	30, Apr. 2010	Daily:240	1-h:5760
	The ratio of training data samples: number of test data samples							3:1
	The testing and training of the models for seasonally ahead forecasting							
	From	To	total data samples		From	To	total data samples	
	3, Mar. 2009	5, May 2009	Daily:182	1-h:4368	1, Mar. 2010	31, Mar. 2010	Daily:62	1-h:1488
	The ratio of training data samples: number of test data samples							2.9:1

(1-Hour = 2 × 30 × 24 = 1440.000, 1- hour comprises of 2 data samples (data collection with 30-min intervals).
(1-Dat = 30 × 2 = 60.00, 1- hour comprises of 2 data samples (data collection with 30-min intervals)).

Table 3
Performance evaluation indices for 1-onth ahead load forecasting with different models.

Sr. No.	Model	Performance Indices	Jan.	Feb.	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
1	ANN-NAEMI	CV (%)	4.997	2.161	4.677	4.7817	5.502	6.249	3.303	2.4773	2.241	2.300	4.826	4.639
		MAPE (%)	4.226	1.663	3.758	3.875	4.429	5.104	2.405	1.822	1.780	**1.651**	4.034	3.744
		MAE	368.833	150.231	299.366	302.448	371.056	466.936	229.830	164.859	149.819	137.078	339.272	307.940
2	MLRM	CV (%)	6.4550	5.405	3.399	5.013	2.334	4.667	4.188	3.6914	3.069	4.142	5.651	7.505
		MAPE (%)	4.723	4.135	2.451	3.482	1.788	3.407	3.095	2.626	2.346	3.031	4.043	5.169
		MAE	441.262	376.451	216.902	284.469	157.418	327.81	299.857	235.959	198.637	256.714	373.934	156.896
3	AdaBoost	CV (%)	0.8171	0.9755	0.916	0.541	0.556	1.150	0.214	0.203	0.6962	0.308	0.177	0.958
		MAPE (%)	0.0197	0.0953	0.050	0.031	0.0284	0.0305	0.009	0.085	0.025	0.010	0.007	0.044
		MAE	1.929	7.316	4.744	2.771	2.518	3.3827	2.613	2.723	2.0833	0.6720	0.712	3.179

in October and 4.429% in May respectively. The best CV of ANN-NAEMI model can be witnessed 2.161%, 2.241% and 2.300% in February, September and October respectively. The ANN-NAEMI and MLRM forecasting error are substantially comparable or similar. The ANN-NAEMI forecasting performance is lower than AdaBoost and considerably higher than the MLRM. In the performance evaluation statistics, MAE is a standard of variation/difference within two continuous parameters. The MAE is provided from: It is conceivable to recognize the kinds of variation from resembling at a plot. The primary intention to practice the MAE has measured the mean significance and magnitude of the model's error in a set of forecasting, outwardly recognizing their place or direction. It is the mean over the sample test of the absolute variations among actual and forecasting view and observations wherever every individual variation has similar kind of weight. The AdaBoost MAE is lower contrasted with the other models. The best MAE found 0.712 in November of AdaBoost model.

5.1.2. Seasonally forecasting

In seasonality forecasting, the data is divided into four parts: i) summer; ii) automn; iii) winter; and iv) spring season. Fig. 8 examines the forecasted amounts of the linear regression analysis amidst the true amounts (actual values) from the datasets of ANN-NAEMI algorithm and determines that equation results perform adequately. It is also being described as the visualisation of the amount R-square. The R-square in summer, automn, winter and spring period is 0.960, 0.885, 0.830 and 0.930 respectively and it is highly correlated between actual and forecasted energy requirement.

Fig. 9 explicates the seasonally ahead forecasting results of ANN-NAEMI, MLRM and AdaBoost model respectively that are arbitrarily chosen from the experiment or testing data for visualisation persistence. The algorithms impersonate the single models of energy usage designs witnessed: sinusoid moulded. It is observed that

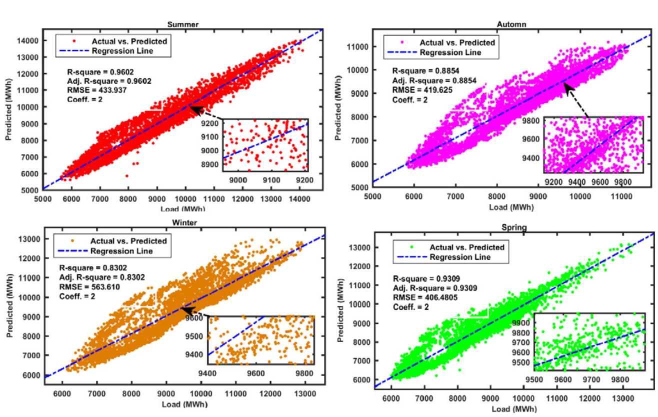

Fig. 8. Actual and predicted energy requirement by ANN-NAEMI.

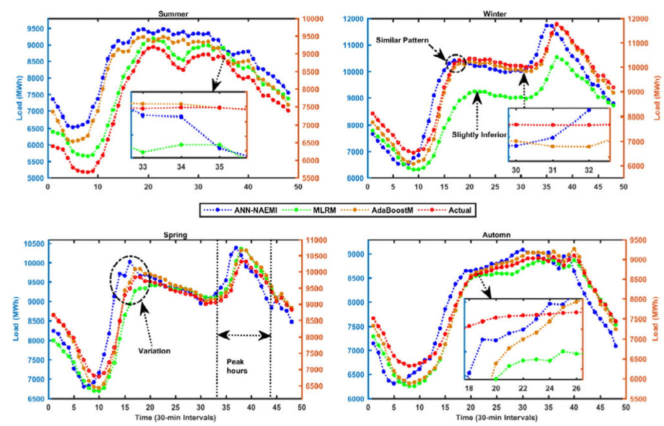

Fig. 9. Seasonally (Summer, Winter, Spring, Automn) actual and forecasted energy demand with ANN-NAEMI, MLRM and AdaBoost model.

the sinusoid designs of all algorithms have stable efficiency and performances. The AdaBoost model's performance is more precise and similar to the actual energy consumption.

The correlation of the forecasted and actual energy usage demonstrated an innumerable auspicious and is compiled in Table 4. The MAPE of ANN-NAEMI and MLRM is 2.441% and 2.806% respectively in winter and automn season is lower than the summer and spring season. The AdaBoost seasonally forecasting performance is almost similar to the 1-month ahead energy forecasting. The detailed overview of performance statistics is presented in Table 4.

5.1.3. Yearly forecasting results

The basic aim to classify the data into three parts monthly, seasonally and the 1-year period is to evaluate the accuracy and performance of the models in distinct duration. Figs. 10 and 11 shows the prediction performance and residuals (error) of MLRM and AdaBoost model between actual and forecasted energy requirement. The application of curve fitting tool generates and contrives a X input (or actual or predictor energy usage data) and Y output (response or forecasted by algorithms). The fit of the curve is a sine sum seizure, and witness the settings fit the ostentation of 5-degree in this study. The fitting curve tool reveals results of adjusting and fitting the specification of the data including a polynomial quadratic etc., for the results window, wherever the perfect description of the model can be observed, the goodness of fit statistics and adapted coefficients. The model's residuals designate that an immeasurable fit, therefore, continues investigating the different kinds of fits to evaluation the sets of data.

Fig. 10 shows the regression line with actual and predicted energy demand with 95% confidence bound. The R-square of regressing fitting is observed 0.872. Fig. 11 explicates the

Table 4
Performance evaluation indices for seasonal ahead load forecasting with different models.

Sr. No.	Model	Performance Indices	Summer	Automn	Winter	Spring
1	ANN-NAEMI	CV (%)	4.865	4.974	3.3671	4.732
		MAPE (%)	4.024	3.996	2.441	3.806
		MAE	344.169	322.491	227.874	308.797
2	MLRM	CV (%)	6.798	4.111	4.447	5.422
		MAPE (%)	4.840	2.806	3.258	3.630
		MAE	439.038	241.249	307.444	320.545
3	AdaBoost	CV (%)	0.041	0.407	0.251	0.284
		MAPE (%)	0.061	0.010	0.005	0.049
		MAE	0.056	0.888	0.5296	0.485

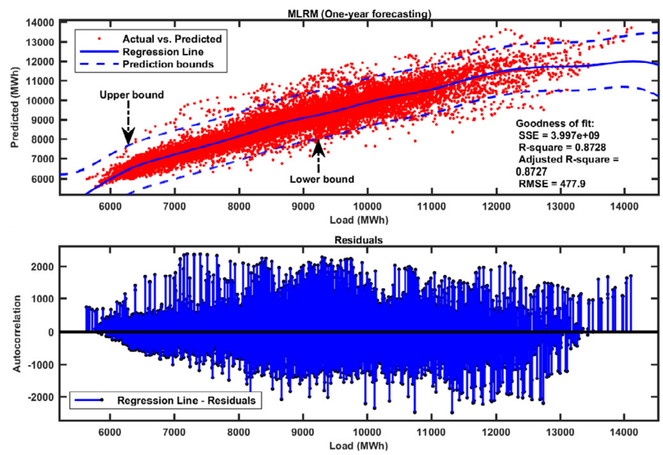

Fig. 10. The MLRM 1-year ahead forecasting with residuals.

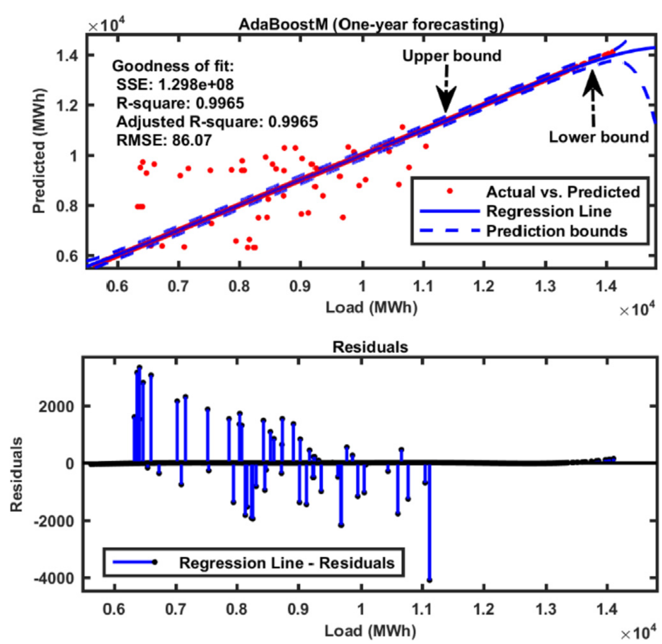

Fig. 11. The AdaBoost model 1-year ahead forecasting with residuals.

forecasting performance of AdaBoost model with 0.996 coefficient of determination. Probably, if an algorithm could justify 100% of the deviation or variance, the adapted conditions would continuously compare the perceived amounts and, consequently, all points of data would decline on the linear fitted regression queue. The AdaBoost model performance is almost similar to actual energy consumption data. The results of the comparison of prediction amounts of particular algorithms and the consolidated predictions are expressed in the Fig. 12. The AdaBoost model performance almost has the similar design with actual data. There is slight variation in ANN-NAEMI model.

The prediction performance practicing various combinations and models of the outcomes are displayed in Table 5. A low error amount designates higher model's performance. The AdaBoost model includes the higher precision data statistics. Comparable to the ANN-NAEMI model, the MLRM model showed better than the ANN-NAEMI model with CV and MAPE statistics extending from 5.944% to 4.073% sequentially.

The visual graphs and performance indices demonstrates that forecasting performance in all session is almost similar. The AdaBoost performance is higher than that the ANN-NAEMI and MLRM model in different sessions. The ANN-NAEMI and MLRM prediction performance in one-month ahead and seasonally ahead is similar. The ANN-NAEMI error is higher in yearly ahead energy prediction. The AdaBoost energy forecasting performance is precise and similar in monthly, seasonally and yearly ahead prediction.

5.2. Forecasting error

Fig. 13 shows the autocorrelation and cross-correlation of ANN-NAEMI model. It represents how the forecasting errors are associated with time pane. It also exposes the forecasting error between

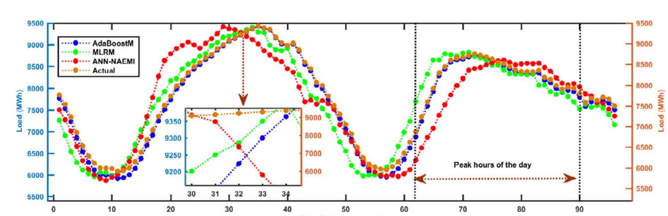

Fig. 12. Yearly actual and forecasted energy demand with AdaBoost, MLRM, ANN-NAEMI model.

Table 5
Performance evaluation indices for 1-year ahead load forecasting with different models.

Sr. No.	Model	Performance Indices	Error
1	ANN-NAEMI	CV (%)	13.092
		MAPE (%)	9.178
		MAE	823.620
2	MLRM	CV (%)	5.944
		MAPE (%)	4.073
		MAE	367.805
3	AdaBoost	CV (%)	0.971
		MAPE (%)	0.082
		MAE	3.931

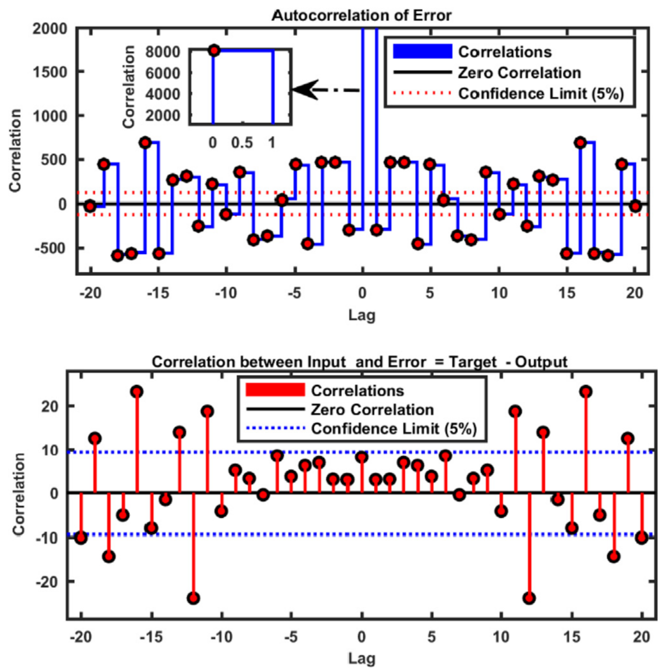

Fig. 13. The error autocorrelation and cross-correlation of ANN-NAEMI model.

predicted and actual energy requirement is quite lower and within the admissible boundary. For an accurate and comprehensive forecasting algorithm, there will only be the single non-zero amount of the auto-correlation function, and it could happen at zero-lag. This would average which the forecasting errors were totally non-correlated including each-other. If there is the important similarity in the forecasting errors, later it could be conceivable to enhance the forecasting - conceivably by expanding the total amount of obstructions in the delay tapped lines. In this scenario, the associations or correlation, without for the single at zero (0) lag, decline generally inwards the 95% confidence bounds approximately zero (0), so algorithm appears to be satisfactory. The cross-correlation between actual and forecasted error demonstrates that how the deviations are associated including the sequence input $x(t)$. For a comprehensive forecasting algorithm, the whole of the similarities should be zero or utmost near to zero.

Fig. 14 show the performance training history to examine for possible overfitting. This also explains that validation and training errors whole declined till iteration (*104*). It doesn't resemble that

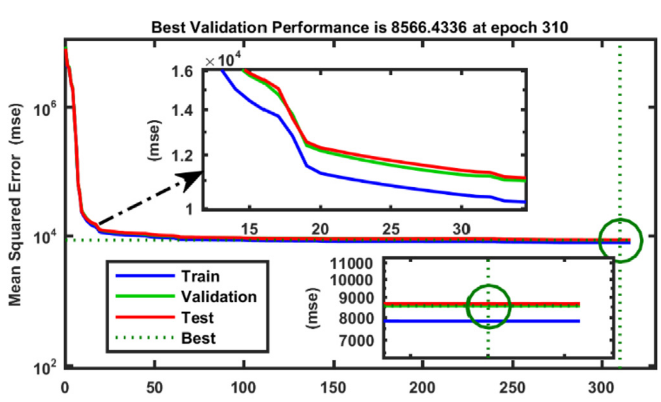

Fig. 14. Best validation performance state at different epochs of ANN-NAEMI model.

each overfitting has happened due to neither validation nor testing error raised prior iteration (*104*).

Fig. 15 demonstrates the error between actual and forecasted demand for one day ahead of ANN-NAEMI model. Fig. 16 shows the seasonally forecasting error comparison of ANN-NAEMI, MLRM and AdaBoost model. It is evident that the error range division of the AdaBoost algorithm is very narrow, and the concentration of degree for the frequency error configuration comparative near to zero (0) is significantly higher than that of another prediction algorithm. Furthermore, the error limit of ANN-NAEMI algorithm is imperceptibly higher than that is far diminutive than the MLRM model. The foregoing dissection indicates that the intended algorithms allow important increases in energy prediction.

5.3. Comparison with the previous study

A comparable algorithm directed were proposed practicing both autoregressive moving average integrated with exogenous parameters (ARIMAX) and artificial neural network (ANN) approach for short-term energy prediction for domestic lower voltage distribution systems [36]. The energy usage data samples applied in the development to forecast the load of algorithms has been obtained and presented from Energex (e.g., the distribution power company for a specific region). The climate and energy consumption data

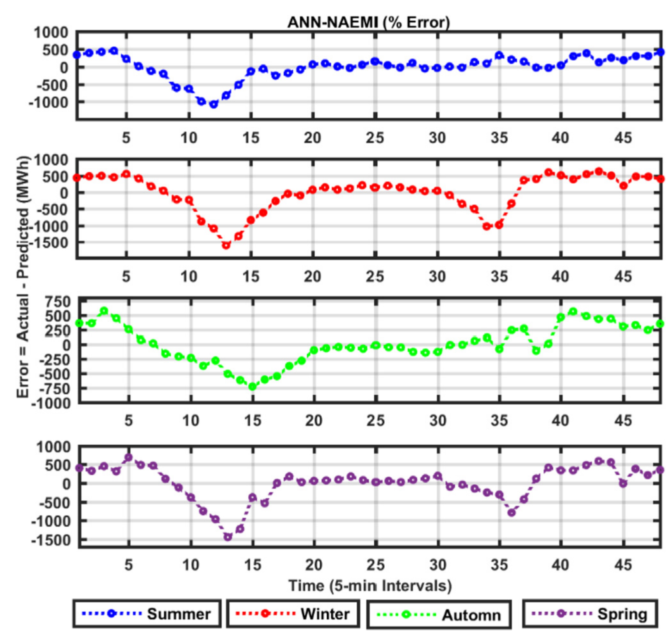

Fig. 15. Seasonally forecasting error of ANN-NAEMI model.

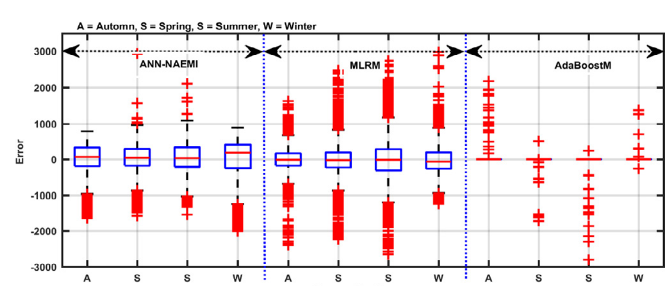

Fig. 16. Seasonally forecasting error comparison of ANN-NAEMI, MLRM and AdaBoost model.

Table 6
Performance evaluation indices comparison with the previous study.

Performance evaluation Indices	Forecasting models					
	ANN-NAEMI	MLRM	AdaBoost	BoostedT [33]	BaggedT [33]	ARIMAX [37]
MAPE (%)	4.226	4.723	0.0197	6.241	3.069	6.970
CV (%)	4.997	6.4550	0.8171	24.742	23.936	–

obtained from a transformer registered at 10 min time interval and it disseminates electricity to almost 128 domestic consumers. The dataset includes the duration from 15 January 2012 to 15 February 2013. A different study [25], intended the six energy prediction models that are - Gaussian process regression, tree bagger, bagged tree, multiple linear regression, neural network and boosted tree. The model's input variables included the designated duration, weather data and energy consumption circumstances of the water source heat pump (WSHP). The model's output was the energy usage of water source heat pump. In this research, modeling period were managed in three main parts - 7-day, 14-day and 1-month ahead forecasting from 8th July 2016 to 7th August 2016. The prediction accuracy of data-mining algorithms was covered by different performance evaluation indices. The performance evaluation metrics that were practiced in evaluating the forecasting efficiently were - R, MAE, CV, MSE, RMSE and MAPE. A straightforward observation or comparison might not be very suitable as the characteristics of data and usages of energy are considerably complicated and the forecasting duration is also inconsistent. Although, a correlation of the forecasting performance and accuracy is presented in Table 6. The ANN-NAEMI, MLRM and AdaBoost 1-month ahead prediction error compared with the previous studies which shows in Table 6. The intended models in this study performed adequately in terms of the best result and the variations (error) of forecasting with the previous study.

6. Conclusion and future direction

Medium-term and long-term load forecasting is the indispensable infrastructure for numerous industrial, utilities, and energy management responsibilities in the building sector, like as demand-side management, fault diagnosis and detection and optimization control. Traditional approaches, which massively compromise on physical based postulates/principles, have restricted or limited energy usage as their efficiency and performance is subjected to various physical postulates. The machine learning based models can be practiced unless in a supervised deportment to improve the forecasting algorithms including addressed output and inputs (e.g., cooling and heating load and grid load forecasting), or an unsupervised based to extricate significantly characteristics from models input as a raw data. Aggregated energy consumption data is acquired from ISO-New England and weather variables are applied as the model's input. Total usage of energy from various kinds of the load is the algorithm's output.

The model's results indicate that machine learning models can improve the efficiency and performance of energy load forecasting, particularly when applied in an unsupervised manner for envisioning the higher-level characteristics or features as algorithm inputs. This study examines the massive potential of machine learning models in load prediction from in medium and long-term perceptive for district-level applying ANN-NAEMI, MLRM and AdaBoost model. The recognition of irregular building usage of energy patterns is obtained by applying outlier detection and clustering analysis. The correlation analysis is conducted to observe the correlation between climate variables and net energy consumption. The autocorrelation and cross-correlation relation is observed of ANN-NAEMI model. The performance evaluation indices MAE, MAPE, CV are applied to assess the forecasting performance of the models. In the duration of forecasting algorithms, the model's outcomes determine that the forecasting methods performed enormously immeasurable.

The AdaBoost algorithm proves its preponderance in forecasting when contrasted with others. The excellent forecasting performance is achieved by AdaBoost models. The ANN-NAEMI and MLRM performance are almost similar with AdaBoost model. One impressive conclusion of this study is that meaningful increase in forecasting performance. By contradiction or comparison, the model results with the previous studies, the prediction effects are immeasurable in term of MAPE. The ML techniques applied in this study can obtain and accomplish reliable and accurate forecasting. The variation between the forecasted and observed load can be accepted as signs or indicators for identifying irregularities in prediction operations.

In future perceptive, the performance indicators selection that has been considered, although the succession of certain pointers and indicators with system training, topologies algorithms and function of activation requires being further investigated. Furthermore, a study will be conducted to overcome the forecasting error with a large number of data samples and different input feature parameter sets.

Acknowledgement

The authors appreciatively acknowledge the assistance of National Natural Science Foundation of China (Grant 51576074 and 51328602).

References

[1] Hall S. Development and initial trial of a tool to enable improved energy &human performance in existing commercial buildings. Renew Energy 2014;67:109–18.
[2] Mauser I, Müller J, Allerding F, Schmeck H. Adaptive building energy management with multiple commodities and flexible evolutionary optimization. Renew Energy 2016;87:911–21.
[3] Farhanieh B, Sattari S. Simulation of energy saving in Iranian buildings using integrative modelling for insulation. Renew Energy 2006;31:417–25.
[4] Petersen S, Svendsen S. Method for component-based economical optimisation for use in design of new low-energy buildings. Renew Energy 2012;38:173–80.
[5] Yu D. A two-step approach to forecasting city-wide building energy demand. Energy Build 2018;160:1–9.
[6] Eissa MM. First time real time incentive demand response program in smart grid with"i-Energy" management system with different resources. Appl Energy 2018;212:607–21.
[7] Shi H, Blaauwbroek N, Phuong H. Nguyen, energy management in multi-commodity smart energy systems with a greedy approach. Appl Energy 2016;167:385–96.
[8] Maziar I, Majid G, Alireza R, Abdollah A. Short-term resource scheduling of a renewable energy based micro grid. Renew Energy March 2015;75:598–606.
[9] Muhammad K, Abdollah A, Andrey S, Vassilios A. Minimizing the energy cost for microgrids integrated with renewable energy resources and conventional generation using controlled battery energy storage. Renew Energy 2016;97:646–55.
[10] Roldan-Blay C, Escriva-Escriv G, Porta CR, Alvarez-Bel C. An optimisation algorithm for distributed energy resources management in micro-scale energy hubs. Energy 2017;132:126–35.
[11] Cao X, Dai X, Liu J. Building energy-consumption status worldwide and the

state-of-the-art technologies for zero-energy buildings during the past decade. Energy Build 2016;128:198–213.
[12] Doucoure B, Agbossou K, Cardenas A. Time series prediction using artificial wavelet neural network and multi-resolution analysis: application to wind speed data. Renew Energy 2016;92:202–11.
[13] Rahmana A, Srikumarb V, Amanda D. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Appl Energy 2018;212:372–85.
[14] Rafe Biswas MA, Robinson MD, Fumo N. Prediction of residential building energy consumption: a neural network approach. Energy 2016;117:84–92.
[15] Neto AH, Fiorelli FAS. Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. Energy Build 2008;12:2169–76.
[16] Tiberiu C, Virgone J, Blanco E. Development and validation of regression models to predict monthly heating demand for residential buildings. Energy Build 2008;10:1825–32.
[17] Gunay B, Shen W, Newsham G. Inverse blackbox modeling of the heating and cooling load in office buildings. Energy Build 2017;142:200–10.
[18] Domahidi A, Ullmann F, Morari M, Jones CN. Learning decision rules for energy efficient building control. J Process Contr 2014;24(6):763–72.
[19] Le K, Bourdais R, Guéguen H. From hybrid model predictive control to logical control for shading system: a support vector machine approach. Energy Build 2014;84:352–9.
[20] Peng T, Zhou J, Zhang C, Zheng Y. Multi-step ahead wind speed forecasting using a hybrid model based on twostage decomposition technique and AdaBoost-extreme learning machine. Energy Convers Manag 2017;153: 589–602.
[21] Liu H, Tian H, Li Y, Zhang L. Comparison of four Adaboost algorithm based artificial neural networks in wind speed predictions. Energy Convers Manag 2015;92:67–81.
[22] Chen M, Andrew FB, Robert J, Mellors, Ramirez AL, Dyer KM, Yang X, Wagoner JL. An efficient Bayesian inversion of a geothermal prospect using a multivariate adaptive regression spline method. Appl Energy 2014;136: 619–27.
[23] Valerio RM, Versoa L, Pellegrinoa Anna, Pellerey Franco. A multivariate nonlinear regression model to predict the energy demand for lighting in rooms with different architectural features and lighting control systems. Energy Build 2014;76:151–63.
[24] Olofsson T, Andersson S, gren JU. Building energy parameter investigations based on multivariate analysis. Energy Build 2009;41:71–80.
[25] Ahmad T, Chen H. Short and medium-term forecasting of cooling and heating load demand in building. Environ Data Min Base Approaches 2018;166: 460–76.
[26] Ahmad T, Chen H, Wang J, Guo Y. A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: a review 2018;165:301–20.
[27] Ahmad T, Chen H, Wang J, Guo Y. Review of various modeling techniques for the detection of electricity theft in smart grid environment. Renew Sustain Energy Rev 2018;82:2916–33.
[28] Lerner J, Grundmeyer M, Garvert M. The importance of wind forecasting. Renew Energy Focus 2009;10:64–6.
[29] Amasyali K, El-Gohary NM. A review of data-driven building energy consumption prediction studies. Renew Sustain Energy Rev 2018;81:1192–205.
[30] ISO new England forecasts adequate resources to meet summer electricity demand. AllBusiness.com. Retrieved 2009-10-01.
[31] Seem SE. Using intelligent data analysis to detect abnormal energy consumption in buildings. Energy Build 2007;39:52–8.
[32] Guo L, Ge PS, Zhang MH, Li LH, Zhao YB. Pedestrian detection for intelligent transportation systems combining AdaBoost algorithm and support vector machine. Expert Syst Appl 2012;39(4):4274–86.
[33] Amaraa F, Agbossoua K, Dubé Y, Kelouwani S, Cardenas A, Bouchard J. Household electricity demand forecasting using adaptive conditional density estimation. Energy Build 2017;156:271–80.
[34] Rahman S. Formulation and analysis of a rule-based short-term load forecasting algorithm. Proc IEEE 2002;78:805–16.
[35] Papakostasa K, Mavromatisb T, Kyriakisa N. Impact of the ambient temperature rise on the energy consumption for heating and cooling in residential buildings of Greece. Renew Energy 2010;35(7):1376–9.
[36] Bennett C, Stewart RA, Lu J. Autoregressive with exogenous variables and neural network short-term load forecast models for residential low voltage distribution networks. Energies 2014;7:2938–60.
[37] Ahmad T, Chen H. Utility companies strategy for short-term energy demand forecasting using machine learning based models. Sustain Cities Soc 2018;39: 401–17.
[38] Ahmad T, Chen H, Huang Ronggeng, Yabin Guo, Wang Jiangyu, Shair Jan, Azeem Akram Hafiz Muhammad, Mohsan Syed Agha Hassnain, Kazim Muhammad. Supervised based machine learning models for short, medium and long-term energy prediction in distinct building environment. Energy 2018;158:17–32.
[39] Ahmad T. Non-technical loss analysis and prevention using smart meters. Renew Sustain Energy Rev 2017;72:573–89.
[40] Wang L, Lv Sheng-Xiang, Zeng Yu-Rong. Effective sparse adaboost method with ESN and FOA for industrial electricity consumption forecasting in China. Energy 2018;155:1013–31.
[41] Taieb SB, Hyndman RJ. A gradient boosting approach to the kaggle load forecasting competition. Int J Forecast 2014;30:382–94.
[42] Gaillard P, Goude Y. Forecasting electricity consumption by aggregating experts; how to design a good set of experts. In: Modeling and stochastic learning for forecasting in high dimensions. Springer; 2015.
[43] Ren Y, Suganthan PN, Srikanth N. Ensemble methods for wind and solar power forecasting a state-of-the-art review. Renew Sustain Energy Rev 2015;50: 82–91.
[44] Ren Y, Qiu X, Suganthan PN. Empirical mode decomposition based adaboost backpropagation neural network method for wind speed forecasting. In: Proceedings of IEEE symposium on computational intelligence in ensemble learning; 2014. p. 1–6.
[45] Rui Y, Sun K, Liu F, Mei S. Efficient simulation of temperature evolution of overhead transmission lines based on analytical solution and NWP. IEEE Trans Power Deliv 2018;33(4):1576–88.
[46] Bartos M, Chester Mikhail, Johnson Nathan, Gorman Brandon, Eisenberg Daniel, Linkov, Bates Matthew. Impacts of rising air temperatures on electric transmission ampacity and peak electricity load in the United States. Environ Res Lett 2016;11:114008.
[47] Adedayo G, Samson Adeleke B. Effect of temperature change on the resistance of transmission line losses in electrical power network. Int J Renew Energy Technol Res 2017;6:1–8.