Energy and Buildings 147 (2017) 77–89

Contents lists available at ScienceDirect

Energy and Buildings

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

Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption

Muhammad Waseem Ahmad*, Monjur Mourshed, Yacine Rezgui

BRE Centre for Sustainable Engineering, School of Engineering, Cardiff University, Cardiff CF24 3AA, United Kingdom

ARTICLE INFO

Article history:
Received 31 October 2016
Received in revised form 17 February 2017
Accepted 6 April 2017
Available online 23 April 2017

Keywords:
HVAC systems
Artificial neural networks
Random forest
Decision trees
Ensemble algorithms
Energy efficiency
Data mining

ABSTRACT

Energy prediction models are used in buildings as a performance evaluation engine in advanced control and optimisation, and in making informed decisions by facility managers and utilities for enhanced energy efficiency. Simplified and data-driven models are often the preferred option where pertinent information for detailed simulation are not available and where fast responses are required. We compared the performance of the widely-used feed-forward back-propagation artificial neural network (ANN) with random forest (RF), an ensemble-based method gaining popularity in prediction – for predicting the hourly HVAC energy consumption of a hotel in Madrid, Spain. Incorporating social parameters such as the numbers of guests marginally increased prediction accuracy in both cases. Overall, ANN performed marginally better than RF with root-mean-square error (RMSE) of 4.97 and 6.10 respectively. However, the ease of tuning and modelling with categorical variables offers ensemble-based algorithms an advantage for dealing with multi-dimensional complex data, typical in buildings. RF performs internal cross-validation (i.e. using out-of-bag samples) and only has a few tuning parameters. Both models have comparable predictive power and nearly equally applicable in building energy applications.

1. Introduction

Globally, buildings contribute towards 40% of total energy consumption and account for 30% of the total CO_2 emissions [1]. In the European Union, buildings account for 40% of the total energy consumption and approximately 36% of the greenhouse gas emissions (GHG) come from buildings [1]. Rapidly increasing GHGs from the burning of fossil fuels for energy is the primary cause of global anthropogenic climate change [2], mandating the need for a rapid decarbonisation of the global building stock. Decarbonisation strategies require energy and environmental performance to be embedded in all lifecycle stages of a building – from design through operation to recycle or demolition. On the other hand, enhancing energy efficiency in buildings requires an in-depth understanding of the underlying performance. Gathering data on and the evaluation of energy and environmental performance are thus at the heart of decarbonising building stock.

There is an abundance of readily available historical data from sensors and meters in contemporary buildings, as well as from utility smart meters that are being installed as part of the transition to the smart grid [3]. The premise is that high temporal resolution metered data will enable real-time optimal management of energy use – both in buildings, and in low- and medium-voltage electricity grids, with predictive analytics playing a significant role [4]. However, their full potential is seldom realised in practice due to the lack of effective data analytics, management and forecasting. Apart from their use during operation stage for control and management, data-driven analytics and forecasting algorithms can be used for the energy-efficient design of building envelope and systems. They are especially suitable for use during early design stages and where parameter details are not readily available for numerical simulation. Their use helps in reducing the operating cost of the system, providing thermally comfortable environment to the occupants, and minimising peak demand.

Building energy forecasting models are one of the core components of building energy control and operation strategies [5]. Also, being able to forecast and predict building energy consumption is one of the major concerns of building energy managers and facility managers. Precise energy consumption prediction is a challenging task due to the complexity of the problem that arises from seasonal variation in weather conditions as well as system non-linearities and delays. In recent years, a number of approaches – both detailed and simplified, have been proposed and applied to predict building energy consumption. These approaches can be classified into three main categories: numerical, analytical and predictive (e.g.

* Corresponding author.
E-mail addresses: AhmadM3@Cardiff.ac.uk (M.W. Ahmad), MourshedM@Cardiff.ac.uk (M. Mourshed), RezguiY@Cardiff.ac.uk (Y. Rezgui).

http://dx.doi.org/10.1016/j.enbuild.2017.04.038

Nomenclature

w_{ij}	weight from ith input node to jth hidden layer node
θ_j	threshold value between input and hidden layers
h_j	vector of hidden-layer neurons
$f_h()$	logistic sigmoid activation function
λ	slope control variable of the sigmoid function
w_{kj}	weight from jth hidden layer node to kth output layer node
θ_k	threshold value between hidden and output layers
δ_k	errors' vector for each output neurons
d_k	target activation of output layer
f'_h	local slope of the node activation function for output nodes
δ_j	errors' vector for each hidden layer's neurons
x	inputs
y	outputs

Subscripts

i	input node
j	hidden layer's neuron
k	output node
c	decision tree number c

artificial neural network, decision trees, etc.) methods. Numerical methods (e.g. TRNSYS,[1] EnergyPlus,[2] DOE-2[3]) often enable users to evaluate designs with reduced uncertainties, mainly due to their multi-domain modelling capabilities [6]. However, these simulation programs do not perform well in predicting the energy use of occupied buildings as compared to the design stage energy prediction. This is mainly due to insufficient knowledge about how occupants interact with their buildings, which is a complicated phenomenon to predict. Also, these prediction engines require a considerable amount of computation time, making them unsuitable for online or near real-time applications. Ahmad et al. [7,8] have also stressed on the need of developing and using predictive models instead of whole building simulation program for real-time optimization problems. On the other hand, analytical models rely on the knowledge of the process and the physical laws governing the process. Key advantage of these models over predictive models is that once calibrated, they tend to have better generalization capabilities. These models require detailed knowledge of the process and mostly require significant effort to develop and calibrate.

According to a recent review by Ahmad et al., [9] significant advances have been made in the past decades on the application of computational intelligence (CI) techniques. The authors reviewed several CI techniques in the paper for HVAC systems; most of these techniques use data available from building energy management system (BEMS) for developing the system model, defining expert rules, etc., which are then used for prediction, optimization and control purposes. These models require less time to perform energy predictions and therefore, can be used for real-time optimization purposes. However, most of these models rely on historical data to infer complex relationships between energy consumption and dependent variables. Among them ensemble-based methods (e.g. random forest) are less explored by the building energy research community. Random forest offers different appealing characteristics, which makes it an attracting tool for HVAC energy prediction. Among these characteristics are [10]; (i) it incorporates interaction between predictors, (ii) it is based on ensemble learning theory, which allows it to learn both simple and complex problems; (iii) random forest does not require much fine-tuning of its hyper-parameters as compared to other machine learning techniques (e.g. artificial neural network, support vector machine, etc.) and often default parameters can result in excellent performance. Artificial neural networks have been extensively used to predict building energy consumption. The literature has demonstrated their ability to solve non-linear problems. ANNs can easily model noisy data from building energy systems as they are fault-tolerant, noise-immune and robust in nature.

Of different building types, hotel and restaurants are the third largest consumer of energy in the non-domestic sector, accounting for 14%, 30% and 16% in the USA, Spain, and UK respectively [11]. Therefore, it is important to address energy predictions of this type of buildings. It is also worth mentioning that hotels and restaurants do not have distinct energy consumption pattern as compared to other building types e.g. offices and schools. Fig. 1 shows electricity consumption (for 5 weeks) of a BREEAM excellent rated school in Wales, UK. It can be seen that the energy consumption during night and weekends is lower as compared to weekdays. This clear pattern makes it easy for machine learning or statistical algorithms to predict energy consumption accurately. However, in hotels and restaurants, energy consumption does not exhibit clear patterns, which makes prediction challenging. Grolinger et al. [12] tackled this type of problem for an event-organizing venue and used event type, event day, seating configuration, etc. as inputs to the models to predict the energy consumption. For hotels and restaurants, energy consumption could depend on many factors e.g. whether there are meetings held at the hotel, sports event in the city, holiday season, weather conditions, time of the day, etc. Most of this information can be collected from building energy management systems (BEMS) and hotel reservation system. However, in this work we have tried to use as little information as possible to develop reliable and accurate models to predict the hotel's HVAC energy consumption.

This paper compares the accuracy in predicting heating, air conditioning and ventilation energy consumption of a hotel in Madrid, Spain by using two different machine learning approaches: artificial neural network and random forest (RF). The rest of the article is organised as follows. Section 2 details literature review covering ANN and decision tree based studies used to energy prediction. In Section 3, we describe ANN and RF in detail, including their mathematical formulation. Methodology is described in Section 4, whereas results and discussion are detailed in Section 5. Concluding remarks and future research directions are presented at the end of the paper.

2. Related work

A large number of studies have investigated building energy prediction using different computational intelligence methods. Based on applications, computational intelligence techniques can be classified into different categories: control and diagnosis, prediction and optimization. In the building energy domain, ANNs are the most popular choice for predicting energy consumption in buildings [9]. They were also used as prediction engines for control and diagnosis purposes. They are able to learn complex relationship manifested in a multi-dimensional domain. ANNs are fault-tolerant, noise-immune and robust in nature, which can easily model noisy data from building energy systems. On the other hand, there are limited research that studied decision trees (DTs) for energy predictions.

González and Zamarreño [13] used simple back-propagation NN for short term load prediction in buildings. The models used current and forecasted values of current load, temperature, hour and day

[1] TRNSYS. http://sel.me.wisc.edu/trnsys.
[2] EnergyPlus. http://energyplus.gov.
[3] DOE-2. http://doe2.com.

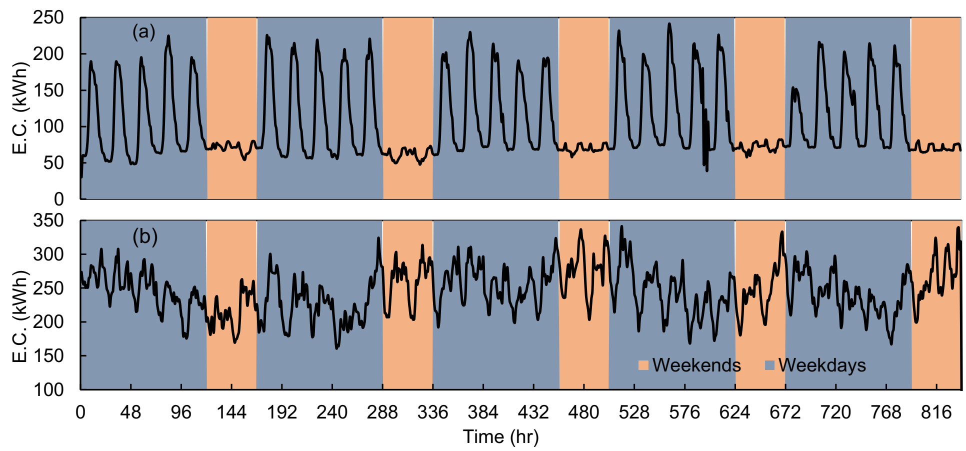

Fig. 1. (a) Building electricity consumption of an BREEAM excellent rated school in Wales, UK. (b) Building electricity consumption of a hotel in Madrid, Spain.

as the inputs to predict hourly energy consumption. It was demonstrated that the proposed model results in accurate results. Nizami and Al-Garni [14] showed that a simple neural network can be used to relate energy consumption to the number of occupants and weather conditions (outdoor air temperature and relative humidity). The authors compared the results with a regression model and it was concluded that ANN performed better. In most of the studies ANN models were developed to be used as a surrogate model instead of using a detailed dynamic simulation programs as they are much faster and can be applied for real-time control applications.

Ben-Nakhi and Mahmoud [15] used general regression neural networks (GRNN) to predict cooling load for three buildings. GRNNs are suitable for prediction purposes due to their quick learning ability, fast convergence and easy tuning as compared to standard back propagation neural networks. The authors used external hourly temperatures readings for a 24 h as inputs to the network to predict next day cooling load. ANN was also used by Kalogirou and Bojic [16] to predict energy consumption of a passive solar building. The authors used four different modules for predicting the electrical heaters' state, outdoor dry-bulb temperatures, indoor dry-bulb temperature for the next step and solar radiation.

Recurrent neural networks are also used in the domain of building energy prediction. Kreider et al. [17] reported the use of these neural networks to predict cooling and heating energy consumption by using only weather and time stamp information. Cheng-wen and Jian [18] used artificial neural network to predict energy consumption for different climate zones by using 20 input parameters; 18 building envelope performance parameters, heating and cooling degree days. The proposed model performed well with a prediction rate of above 96%. Azadeh et al. [19] predicted the annual electricity consumption of manufacturing industries. The authors tackled a challenging task as the energy consumption shows high fluctuations in these industries and the results showed that the ANN are capable of forecasting energy consumption.

Decision trees are one of the most widely used machine learning techniques. They use a tree-like structure to classify a set of data into various predefined classes (for classification) or target values (for regression problems). By this way, they provide a description, categorization and generalisation of the dataset [20]. A decision tree predicts the values of target variable(s) by using inputs variables. One of the main advantages of decision trees is that they produce a trained model which can represent logical statements/rules, which can then be used for prediction purposes through the repetitive process of splitting. Tso and Yau [21] presented a study to compare regression analysis, decision tree and neural networks to predict electricity energy consumption. It was found that decision trees can be a viable alternative in understanding energy patterns and predicting energy consumption. Yu et al. [20] used a decision tree based methods to predict energy demand. The authors modelled building energy use intensity levels to estimate residential building energy performance indices. It was concluded that decision tree based methods could be used to generate fairly accurate models and could be used by users without needing computational knowledge.

Hong et al. [22] developed a decision support model to reduce electric energy consumption in school buildings. Among different computational intelligence techniques, the authors also used decision trees to form a group of educational buildings based on electric energy consumption. From results it was found that decision tree improved the prediction accuracy by 1.83–3.88%. Decision trees were also used by Hong et al. [23] to cluster a type of multi-family housing complex based on gas consumption. The authors used a combination of genetic algorithm, artificial neural network and multiple regression analysis. It was found from the results that decision tree improved the prediction power by 0.06–01.45%. These results clearly demonstrate the importance and usefulness of decision trees for prediction purposes.

3. Machine learning techniques for energy forecasting

3.1. Artificial neural networks

Artificial neural network stores knowledge from experience (e.g. using historical data) and makes it available for use [24]. Fig. 2 shows a schematic diagram of a feed-forward neural network architecture, consisting of two hidden layers. The number of hidden layers depends on the nature and complexity of the problem. ANNs do not require any information about the system as they operate like black box models and learn relationship between inputs and outputs. Different neural network strategies have been developed in the literature e.g. feed-forward, Hopfield, Elman, self-organising maps, and radial basis networks [25]. Among them, feed-forward is the most widely and generic neural network type and has been used for most of the problems. This study also uses a feed-forward neural networks and back-propagation algorithm for modelling the HVAC energy consumption of a hotel. The process of back-propagation algorithm as suggested by many [24,26,27], is summarised as follows [28]:

1. Presenting training samples and propagating through the neural network to obtain desired outputs.
2. Using small random and threshold values to initialise all weights.

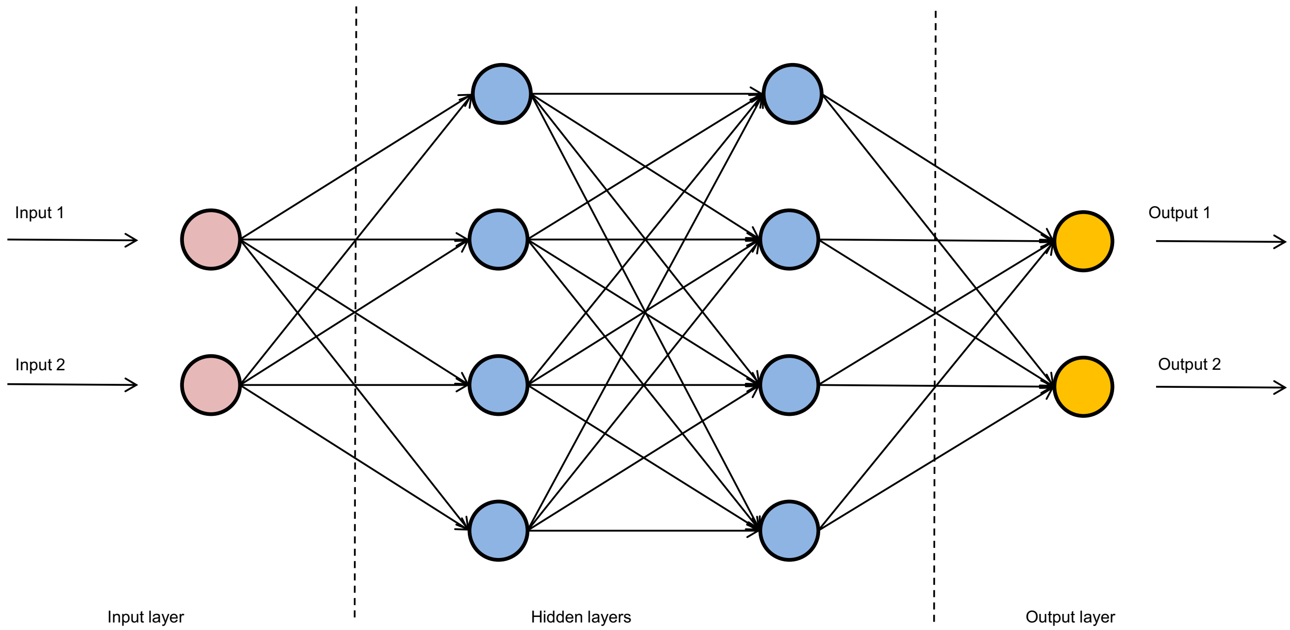

Fig. 2. Schematic diagram of a feed-forward artificial neural network.

Source: Ahmad et al., [9].

3. Calculating input to the j-th node in the hidden layer using Eq. (1).

$$net_j = \sum_{i=1}^{n} w_{ij}x_i - \theta_j \quad (1)$$

4. Calculating output from the j-th node in the hidden layer using Eqs. (2) and (3):

$$h_j = f_h\left(\sum_{i=1}^{n} w_{ij}x_i - \theta_j\right) \quad (2)$$

$$f_h(x) = \frac{1}{1+e^{-\lambda_h x}} \quad (3)$$

5. Calculating input to the k-th node in the hidden layer using Eq. (4).

$$net_k = \sum_{j} w_{kj}x_j - \theta_k \quad (4)$$

6. Calculating output of the k-th node of the output layer by using Eqs. (5) and (6):

$$y_k = f_k\left(\sum_{j} w_{kj}x_j - \theta_k\right) \quad (5)$$

$$f_k(x) = \frac{1}{1+e^{-\lambda_k x}} \quad (6)$$

7. Using Eqs. (7) and (8) to calculate errors from the output layer:

$$\delta_k = -(d_k - y_k)f'_k \quad (7)$$

$$f'_k = y_k(1-y_k) \quad (8)$$

In the above equation, δ_k depends on the error $(d_k - y_k)$. The errors from hidden layers are represented by Eqs. (9) and (10):

$$\delta_k = f'_k \sum_{k=1}^{n} w_{kj}\delta_k \quad (9)$$

$$f'_h = h_j(1-h_j) \quad (10)$$

Eq. (10) is for sigmoid function.

8. Adjusting the weights and thresholds in the output layer.

3.2. Random forest

In recent years, decision trees have become a very popular machine learning technique because of its simplicity, ease of use and interpretability [29]. There have been different studies to overcome the shortcomings of conventional decision trees; e.g. their suboptimal performance and lack of robustness [30]. One of the popular techniques that resulted from these works is the creation of an ensemble of trees followed by a vote of most popular class, labelled forest [31]. Random forest (RF) is an ensemble learning methodology and like other ensemble learning techniques, the performance of a number of weak learners (which could be a single decision tree, single perceptron, etc.) is boosted via a voting scheme. According to Jiang et al. [32], the main hallmarks for RF include (1) bootstrap resampling, (2) random feature selection, (3) out-of-bag error estimation, and (4) full depth decision tree growing. An RF is an ensemble of C trees $T_1(X)$, $T_2(X)$, ..., $T_C(X)$, where $X = x_1, x_2, \ldots, x_m$ is a m-dimension vector of inputs. The resulting ensemble produces C outputs $\hat{Y}_1 = T_1(X)$, $\hat{Y}_2 = T_2(X)$, ..., $\hat{Y}_C = T_C(X)$. $\hat{Y}_c$ is the prediction value by decision tree number c. Output of all these randomly generated trees is aggregated to obtain one final prediction $\hat{Y}$, which is the average values of all trees in the forest. An RF generates C number of decision trees from N number of training samples. For each tree in the forest, bootstrap sampling (i.e. randomly selecting sample number of samples with replacement) is performed to create a new training set, and the samples which are not selected are known as out-of-bag (OOB) samples [32]. This new training set is then used to fully grow a decision tree without pruning by using CART methodology [29]. In each split of node of a decision tree, only a small number of m features (input variables) are randomly selected instead of all of them (this is known as random feature selection). This process is repeated to create M decision trees in order to form a randomly generated forest.

The training procedure of a randomly generated forest can be summarised as [33]:

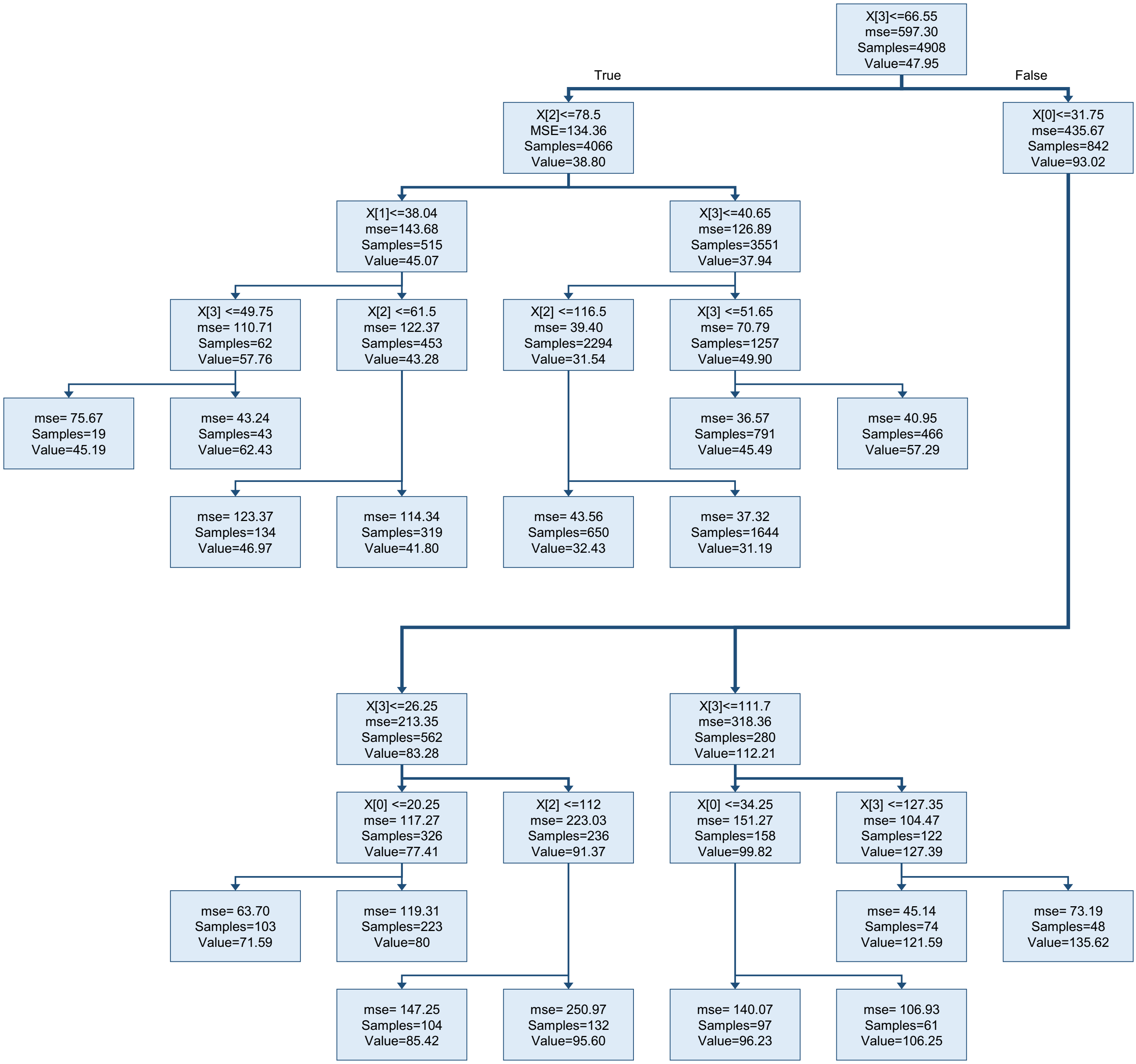

Fig. 3. Decision tree from a random forest for predicting Hotel's HVAC energy consumption. Note: $X[0]$: outdoor air temperature, $X[1]$: relative humidity, $X[2]$: number of rooms booked, $X[3]$: previous value of E.C.

1. Draw a bootstrap sample from the training dataset;
2. Grow a tree for each bootstrap sample drawn in above 1 with following modification: at each node, select the best split among a randomly selected subset of input variables (m_{try}), which is the tuning parameter of random forest algorithm. The tree is fully grown until no further splits are possible and not pruned.
3. Repeat steps 1 and 2 until C such trees are grown.

Fig. 3 shows a decision tree from a random forest for the prediction of HVAC energy consumption of the hotel. It is worth mentioning that this decision tree is only for demonstration purposes. The actual random forest used in the analysis of below results contains complex decision trees i.e. the actual decision trees are more deep and more than 3 features were tried in an individual tree. The decision tree shown in Fig. 3 is taken from a forest which only considers outdoor air temperature, relative humidity, the total number of rooms booked and the previous value of energy consumption as input variables. All 4 features were allowed to be tried in an individual tree, and the maximum depth of a tree is restricted to 4.

4. Methodology

4.1. Data description

The data set of 5 min historical values of HVAC electricity consumption for the studied hotel building was gathered from its building energy management system. Total daily number of guests and rooms booked were also acquired from the reservation system. The 30 min weather conditions including outdoor air temperature,

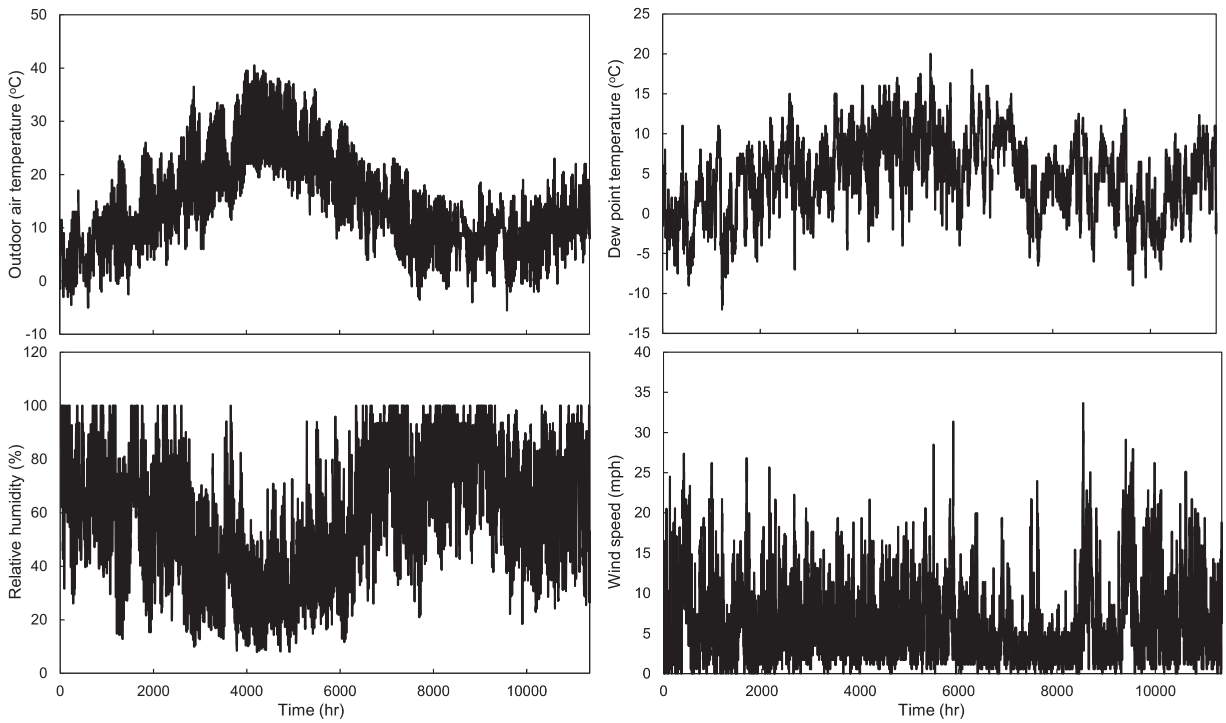

Fig. 4. Weather data of Madrid, Spain. Note: This data is from 14/01/2015 00:00 until 30/04/2016 23:00.

dew point temperature, wind speed and relative humidity were collected from a nearby weather station at Adolfo Suárez Madrid-Barajas International Airport. The weather station is located at a latitude of 40.466 and longitude of −3.5556. The climate of Madrid is classified as continental, with hot summers and moderately cold winters (it has an annual average temperature of 14.6 °C). The hourly values of outdoor air temperature, dew point temperature, relative humidity and wind speed are shown in Fig. 4. The training and validation datasets contain data from 14/01/2015 00:00 until 30/04/2016 23:00 (10,972 data samples after removing outliers and missing values).

To improve ANN model's accuracy, all input and output parameters were normalized between 0 and 1 as follow:

$$x'_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{11}$$

$$y'_i = \frac{y_i - y_{min}}{y_{max} - y_{min}} \tag{12}$$

where $\boldsymbol{x}_i$ represents each input variable, $\boldsymbol{y}_i$ is the building's HVAC electricity consumption, and $\boldsymbol{x}_{min}$, $\boldsymbol{x}_{max}$, $\boldsymbol{y}_{min}$, $\boldsymbol{y}_{max}$ represent their corresponding minimum and maximum values, x'_i and y'_i are normalised input and output variables.

4.2. Evaluation metrics

To assess models' performance, we used different metrics: the mean absolute percent deviation (MAPD), mean absolute percentage error (MAPE), root mean squared errors (RMSE), coefficient of variation (CV) and mean absolute deviation (MAD). CV has been used in the previous studies e.g. [34] and measures the variation in error with respects to the actual consumption average and is defined by:

$$CV = \frac{\sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{\mathbf{y}}_i)^2}{N}}}{\bar{\mathbf{y}}} \times 100 \tag{13}$$

The MAPE metric calculates average absolute error as a percentage and has been used in previous studies for evaluation the performance of a model [35,36]. It is calculated as follows:

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{y_i - \hat{y}_i}{y_i} \times 100 \tag{14}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{\mathbf{y}}_i)^2}{N}} \tag{15}$$

$$\text{MAD} = \frac{1}{N}\sum_{i=1}^{N}|\hat{\mathbf{y}}_i - \mathbf{y}_i| \tag{16}$$

where $\hat{y}_i$ is the predicted value, $\boldsymbol{y_i}$ is the actual values, $\bar{y}$ is the mean of the observed values and N is the total number of samples. In this work, we have used root mean squared error (RMSE) as our primary metric and other metrics were only used as tie-breakers. All three tie-breakers were only considered when the RMSE did not provide a statistical difference between two models.

We used the implementation of random forests included in the scikit-learn [37] module of python programming language, and neurolab [38] for developing artificial neural networks. All development and experimental work was carried out on a personal computer (Intel Core i5 2.50 GHz with 16 GB of RAM).

5. Results and discussion

The values of performance metrics are calculated while considering some or all of the ten input variables (i.e. outdoor air temperature, dew point temperature, relative humidity, wind speed, hour of the day, day of the week, month of the year, number of guests for the day, number of rooms booked). Table 1 shows the sum of squared errors at 1000 epochs, RMSE, CV, MAPE, MAD and R^2 values of reduced artificial neural networks in order to evaluate their performances. First, the performance metrics are shown for a model which considers all of the input variables and then metrics

Table 1
Full and reduced neural networks for predicting HVAC energy consumption.

Input variables	SSE @1000 epochs	RMSE	CV	MAPE	MAD	R^2
DBT, DPT, RH, WS, hr, day, Mon, Occupants, Rooms booked, y_{t-1}	2.943	4.605	9.599	7.761	3.357	0.9639
DPT, RH, WS, hr, day, Mon, Occupants, Rooms booked, y_{t-1}	2.915	4.617	9.624	7.736	3.347	0.9637
DBT, RH, WS, hr, day, Mon, Occupants, Rooms booked, y_{t-1}	2.957	4.626	9.642	7.757	3.357	0.9635
DBT, DPT, WS, hr, day, Mon, Occupants, Rooms booked, y_{t-1}	3.030	4.602	9.593	7.677	3.325	0.9639
DBT,DPT, RH, hr, day, Mon, Occupants, Rooms booked, y_{t-1}	2.953	4.594	9.576	7.690	3.334	0.9640
DBT, DPT, RH, WS, day, Mon, Occupants, Rooms booked, y_{t-1}	2.959	4.621	9.632	7.798	3.368	0.9636
DBT, DPT, RH, WS, hr, Occupants, Rooms booked, y_{t-1}	2.956	4.608	9.603	7.713	3.327	0.9638
DBT, DPT, RH, WS, hr, day, Mon, y_{t-1}	2.957	4.627	9.645	7.718	3.339	0.9635
DBT,DPT, RH, WS, hr, day, Mon, Occupants, Rooms booked	10.153	8.400	17.507	15.473	6.554	0.8798
hr, day, Mon, Occupants, Rooms booked, y_{t-1}	3.170	4.710	9.817	7.775	3.375	0.9622

Notes: DBT: outdoor air temperature, DPT: dew-point temperature, RH: relative humidity, hr: hour of the day, day: day of the week, Mon: month of the year, occupants: total number of occupants booked, rooms booked: total rooms booked on the day, $y_t - 1$: previous hour value of energy consumption. The network architecture of these ANNs was; number of inputs:10:1; where 10 is the number of hidden layer neurons and 1 is the number of output layer neuron.

are listed for networks considering fewer inputs. It is also worth mentioning that all of the networks were trained and tested on same datasets. From results, it is clear that the model without wind speed as an input variable provided better results on all of the performance metrics. There was a small difference between networks considering all inputs and reduced networks. However, Fig. 5 shows that the performance of the network trained without considering previous hour value reduced significantly as compared to the model developed by using all inputs.

Sensitivity of ANN model was studied for different number of hidden layer neurons in the range of 10–15. For artificial neural networks, there is no general rule for selecting the number of hidden layer neurons. However, some researchers have suggested that this number should be equal to one more than twice the number of input neurons (input variables) i.e. $2n+1$ [39], where n is the number of input neurons. It is also reported that the number of neurons obtained from this expression may not guarantee generalization of the network [40]. It is worth mentioning that the number of hidden layer neurons vary from problem to problem and, depends on the number and quality of training patterns [41]. If too few neurons in the hidden layers are selected, the network can result in large errors (under-fitting). Whereas, if too many hidden layer neurons are included, then the network can result in overfitting (i.e. it learns the noise in the dataset). We first tried 10 number of neurons and then used the stepwise searching method to find the optimal value of hidden layer neurons. It was found that for our problem the higher number of neurons were not making a significant difference in the accuracy of the models and therefore we chose 10 neurons to reduce network's complexity. We used Broyden–Flatcher–Goldfarb–Shano (BFGS) as training algorithm as it provided better results and requires few tuning parameters. Also, only one hidden layer was used as the use of more than one hidden layers did not improve model's performance substantially. Generally, one hidden layer should be adequate for most of the applications [42]. Considering the limited space, the details about searching the neurons in the hidden layer, no. of hidden layers and training algorithms are not shown here.

The effect of tree depth of the performance of a random forest is demonstrated in Figs. 6 and 7 and Table 2. The results in Table 2 show that a forest constructed with deeper trees resulted in better performance. A maximum depth of 1 led to under-fitting, whereas the performance of the forest started to deteriorated with max depth more than 10. Tree deeper than 10 levels started to under-fit i.e. all performance metrics started to increase. Fig. 7 shows that a forest with maximum depth = 1 resulted in an under-fit model and any value of energy consumption below 66 kWh was predicted as 38.559 kWh. This was the reason that the models resulted in higher values of RMSE (13.219), CV (27.551%), MAPE (24.622%), MAD (10.511) and lower value of R^2 (0.7028).

Figs. 8 and 9 and Table 3 show the performance of RF while varying the number of features. It represents the number of randomly selected variables for splitting at each node during the tree

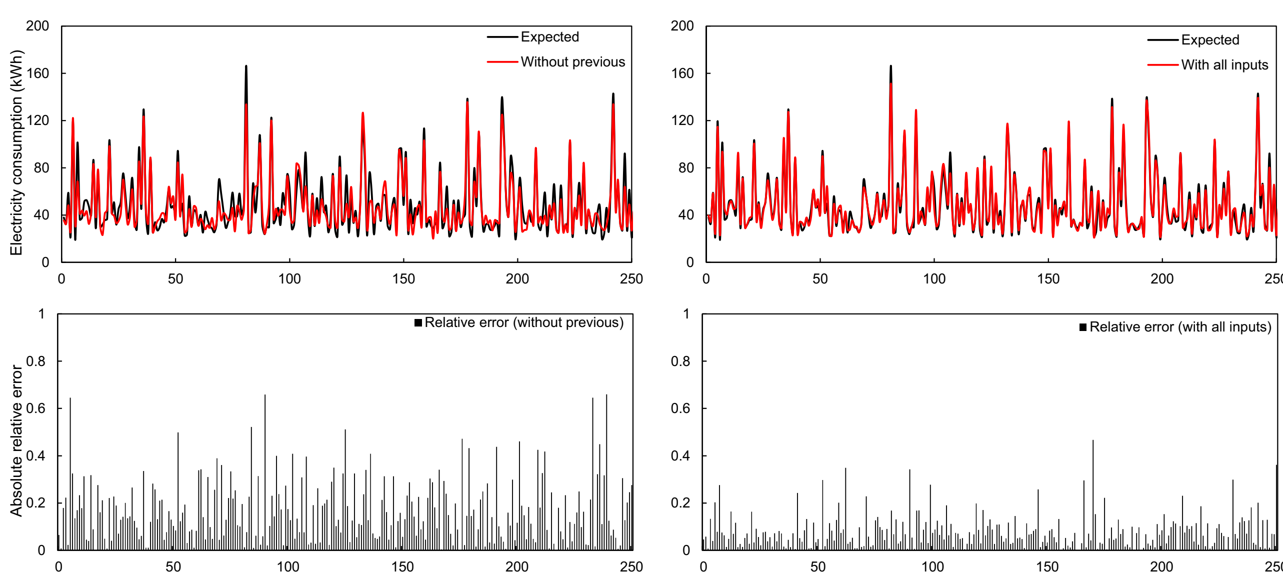

Fig. 5. Results from ANN models developed without previous hour value and with all variables as inputs.

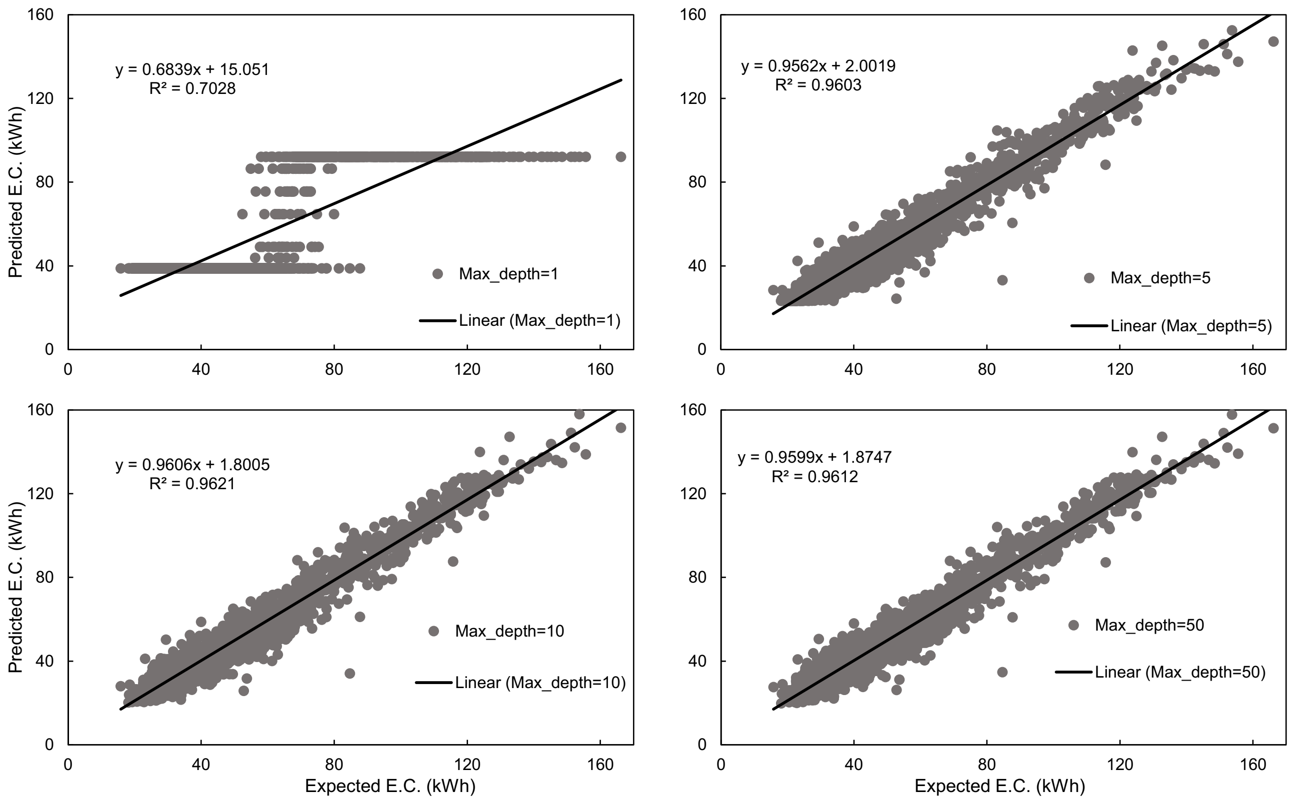

Fig. 6. RF models with different maximum depths.

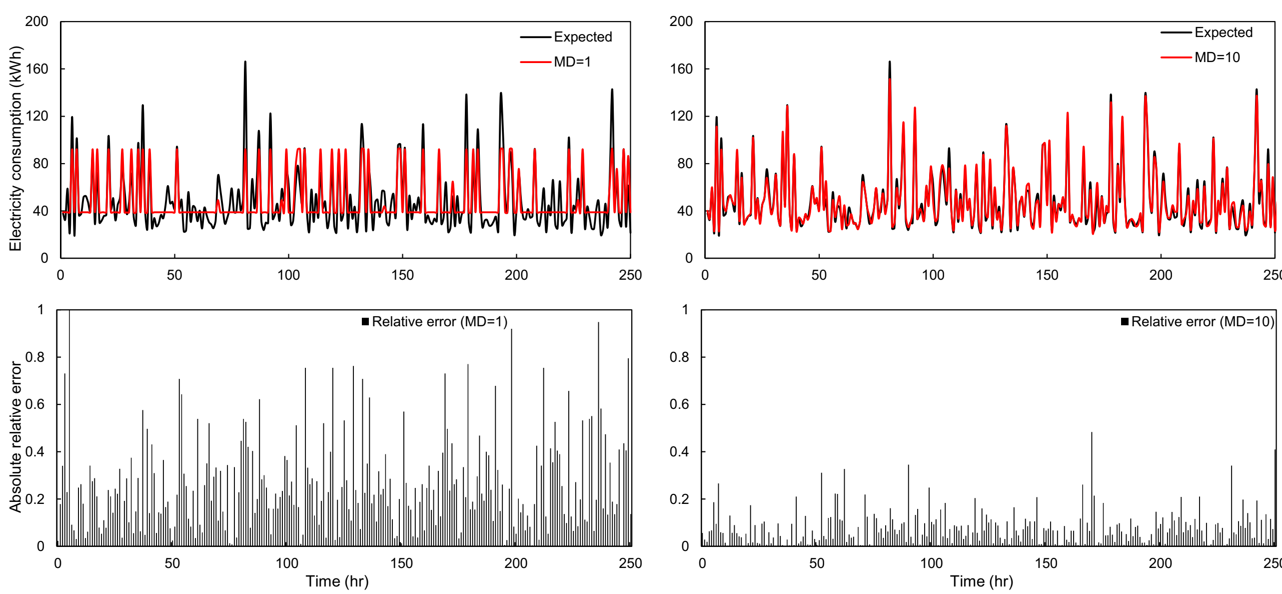

Fig. 7. RF models with different maximum depths.

Table 2
Random forest models with different max depth parameter.

Random forest models	RMSE (kWh)	CV (%)	MAPE (%)	MAD (kWh)	R^2 (–)
RF with max depth = 1	13.219	27.551	24.622	10.511	0.7028
RF with max depth = 3	5.792	12.072	9.681	4.253	0.9429
RF with max depth = 5	4.831	10.069	7.969	3.474	0.9603
RF with max depth = 7	4.734	9.866	7.831	3.397	0.9618
RF with max depth = 10	4.714	9.825	7.814	3.387	0.9621
RF with max depth = 15	4.755	9.910	7.927	3.428	0.9615
RF with max depth = 50	4.772	9.945	7.982	3.447	0.9612

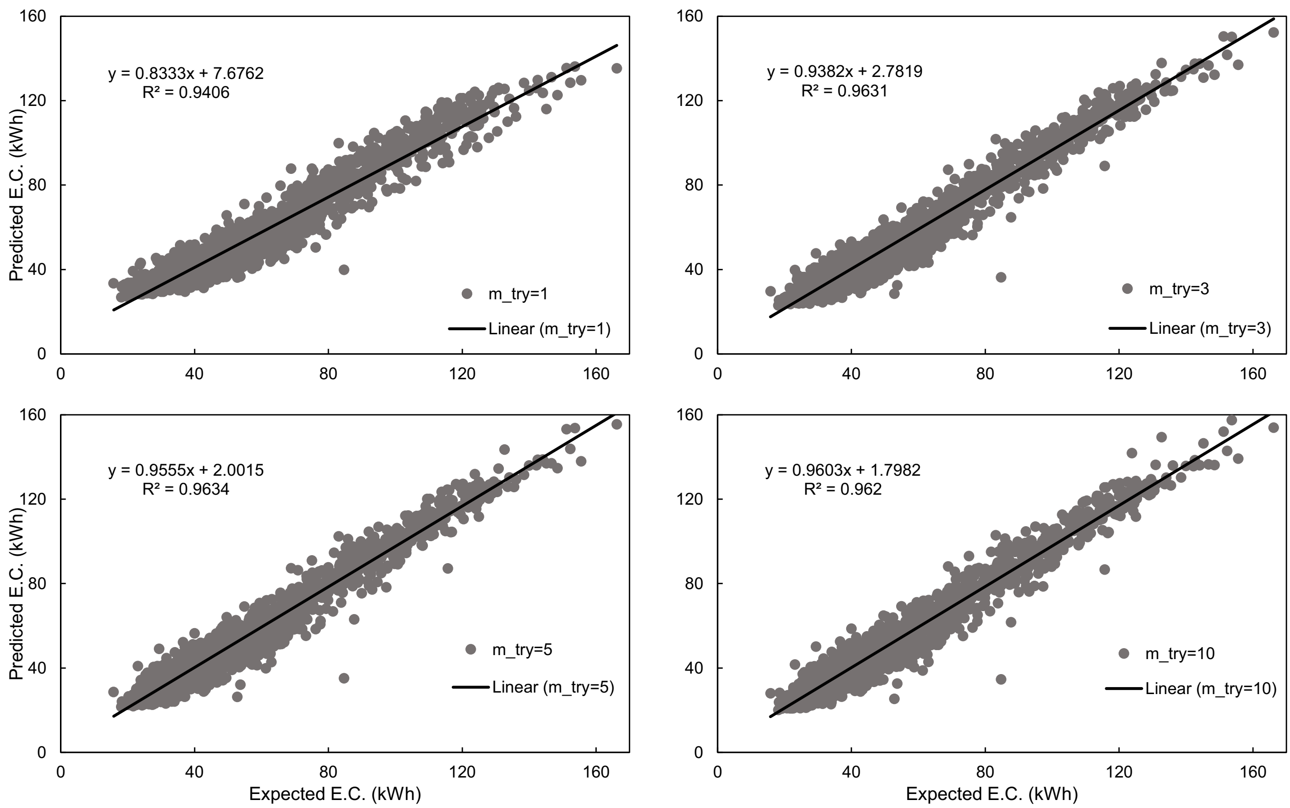

Fig. 8. RF models with different maximum features and max depth = 10.

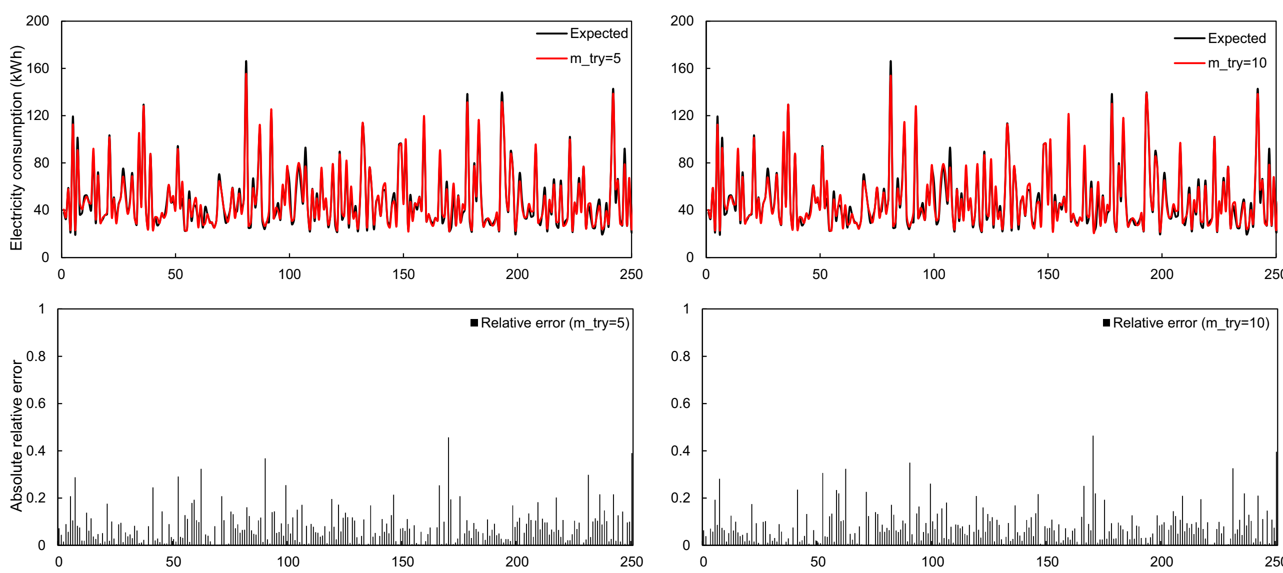

Fig. 9. RF models with m_{try} = 5 and m_{try} = = 10.

Table 3
Random forest models with different max features parameter.

m_{try}	RMSE(kWh)	CV (%)	MAPE (%)	MAD (kWh)	R^2(–)
1	6.493	13.535	12.193	4.998	0.9406
3	4.700	9.796	8.078	3.447	0.9631
5	4.640	9.671	7.733	3.345	0.9634
7	4.679	9.753	7.760	3.364	0.9627
9	4.714	9.824	7.820	3.389	0.9622
10	4.725	9.848	7.835	3.396	0.9262

Table 4
Comparison of the prediction errors of full and reduced models.

Model	Training dataset			Validation dataset		
	RMSE (kWh)	CV (%)	R^2 (–)	RMSE (kWh)	CV (%)	R^2 (–)
RF with all variables	3.31	6.93	0.981	4.66	9.60	0.965
RF with important variables	3.39	7.10	0.98	4.72	9.84	0.962
ANN with all variables	4.38	9.18	0.97	4.84	10.07	0.961
ANN with important variables	4.44	9.31	0.966	4.60	9.59	0.964

induction [33]. Generally, increasing the number of maximum features at each node can increase the performance of an RF, as each node of the tree has a higher number of features available to consider. However, this may not be true for all cases, in our case the performance of RF improved quite significantly while considering more than 1 feature but the performance started to decrease when we tried more than 5 features. As shown in Table 3, the resulting CV values for the model with max features equal to 5 was 9.671%, which was higher than considering minimum feature (max features = 1, CV = 13.535%) and max features = 10 (CV = 9.848%). The results showed same behaviour for other performance metrics. Also, Fig. 8 demonstrates that a random forest constructed with max features = 1, is under fitted as compared to random forested constructed with more features. Increasing the number of features also reduces the diversity of individual tree in the forest and therefore the performance of the RF reduced after considering more than 5 features. It is also worth mentioning that the construction of a RF with more features is computationally intensive and slower as compared to the one with fewer features. We also used 5-fold grid search (on training data only) to find the best $m_{try} \in \{1,3,5,7\}$ and MD $\in \{1,3,5,7,10,15,50\}$. It was found that the best hyper-parameters were MD = 15 and $m_{try} = 3$. However, we found that the parameters found from step-wise search method resulted in marginally better performance and hence were used for developing the RF models.

Fig. 10 shows the variable importance plot, which was produced by replacing each input variable in turn by random noise and analysing the deterioration of the performance of the model. The importance of the input variable is then measured by this resulting deterioration in the performance of the model. For regression problem, the most widely used score is the increase in the mean of the error of a tree (Mean squared error) [43]. It is found that the previous hour's electricity consumption is the most important variable, followed by outdoor air temperature, relative humidity, the month of the year and hour of the day. Among social variables, number of occupants is the most important variable to enhance prediction accuracy of the model. Wind speed, day of the month, number of rooms booked and outdoor air dew point temperature are the least important variables. Table 4 compares the performance of models developed by using important and all input variables. The

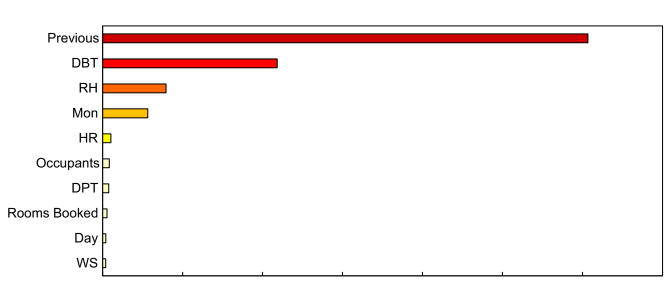

Fig. 10. Variable importance for predicting Hotel's HVAC electricity consumption. Notes: DBT: outdoor air temperature, DPT: dew-point temperature, RH: relative humidity, HR: hour of the day, Mon: month of the year, Day: day of the week, WS: wind speed.

Table 5
Comparison of the prediction errors of RF and ANN models.

Model	RMSE (kWh)	CV (%)	MAPE (%)	MAD (kWh)	R^2(–)
Random forest	6.10	6.03	4.60	4.70	0.92
Artificial neural network	4.97	4.91	4.09	4.02	0.95

table indicates that in our case, the RF models developed by using important variables has marginally lower performance than the model developed by using all input variables. For ANN model, the performance was improved on validation dataset while using important input variables only. Variable importance plot is a useful tool for dimensionality reduction in order to improve model's performance on high-dimensional datasets – the performance could be enhanced by reducing the training time and/or enhancing the generalization capability of the model.

For comparing ANN and RF, both models were used to predict HVAC electricity consumption on a recently acquired data (from 06/07/2016 00:00 until 11/07/2016 23:00). This dataset was not used during the training and validation phases and is used to assess the actual generalisation and prediction power of the models. The predicted HVAC electricity consumption by these two models, the absolute relative errors between predicted and actual electricity consumption values, comparison of prediction and expected electricity consumption, and the histogram of percentage error obtained are illustrated in Figs. 11 and 12. Moreover, RMSE, CV, MAPE, MAD and R^2 of the testing samples using RF and ANN models are compared, which are shown in Table 5.

From Table 5, it is evident that ANN performed slightly better with better performance metrics values (lower RMSE, CV, MAPE and MAD, and higher R^2 values). Both of these models showed strong non-linear mapping generalization ability, and it can be concluded that both of these models can be effective in predicting hotel's HVAC energy consumption. Figs. 11 and 12 and Table 5 showed that ANN outperforms the RF model by a small margin on testing dataset. However, the RF model can effectively handle any missing values during the training and testing phases. As it is an ensemble based algorithm, it can accurately predict even when some of the input values are missing. Also, less accurate results does not mean that RF model did not capture the relationships between input and output variables. RF results are within the acceptable range and can also be utilised for prediction purposes. Figs. 11 and 12 show that RF performed better in predicting the lower values, whereas ANN performed better in predicting the higher values of the electricity consumption. ANN closely followed the electricity consumption pattern and therefore performed slightly better.

We demonstrate that both RF and ANN are valuable machine learning tools to predict building's energy consumption. It is a well known fact that best machine learning techniques for predicting energy consumption cannot be chosen a priori and different computational intelligence techniques need to be compared to find best techniques for a particular regression problem. However, as a comprehensive evaluation of different machines learning techniques is beyond the scope of our work, we only compared ANN and RF

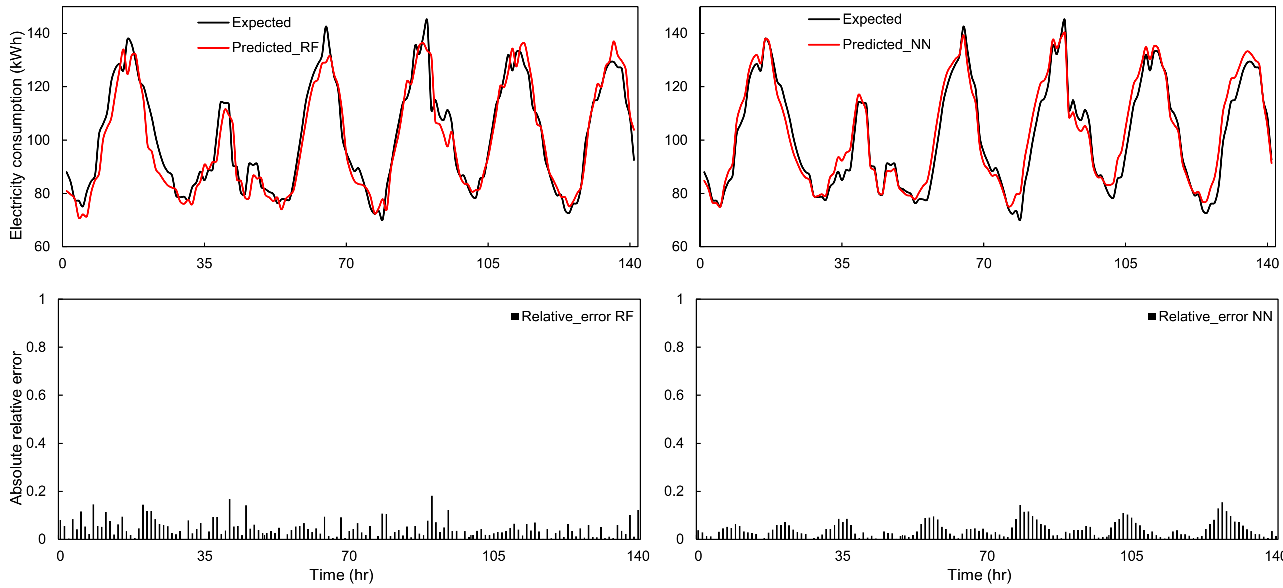

Fig. 11. Predicted energy consumption and absolute errors from RF and ANN models.

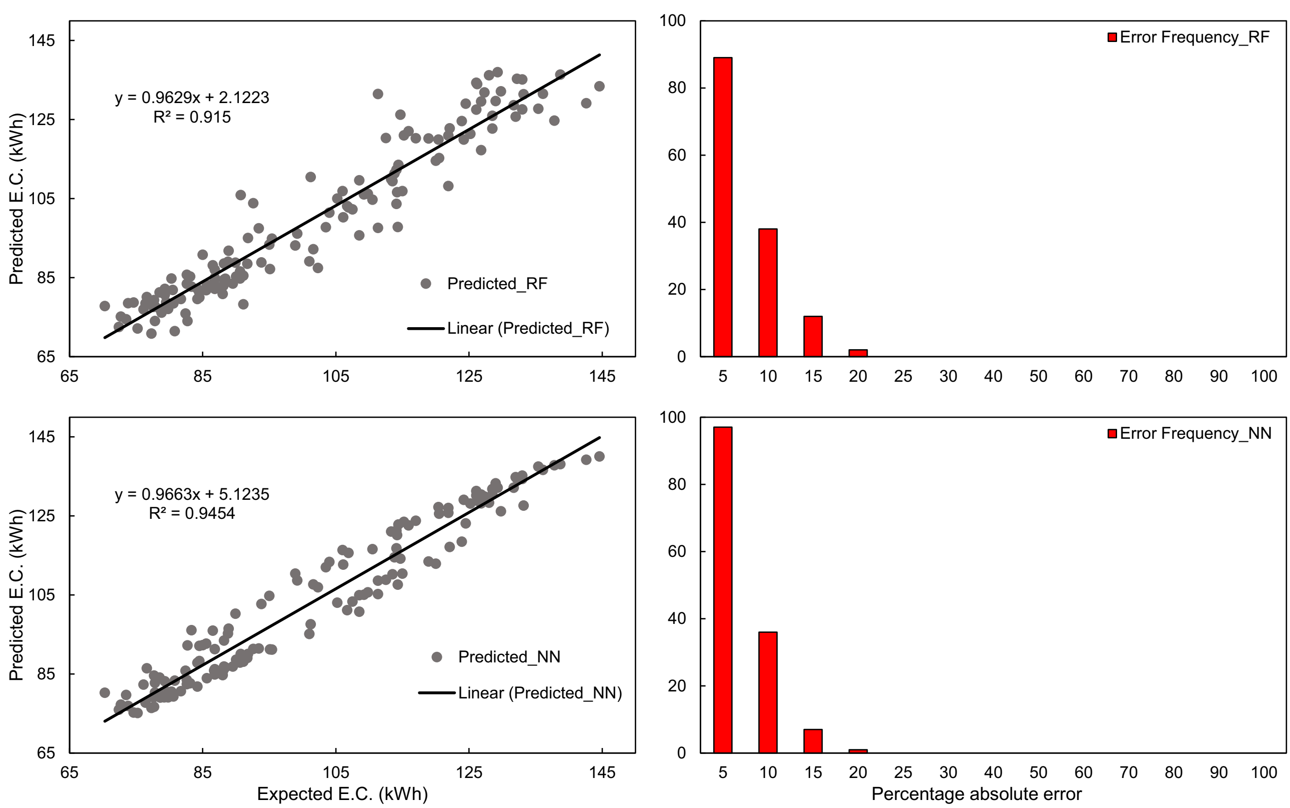

Fig. 12. Comparison between actual and predicted energy consumption and histogram of percentage error plots.

to investigate their suitability for predicting building energy consumption. According to Siroky et al. [44], random forest are fast to train and tune as compared to other ML techniques. For this work, the training time for random forest was much less than for ANN (few seconds compared to minutes). The training time for RF was 9.92 s for 4 jobs running in parallel and 17.8 s for running one job at a time. On the other hand, the training time for ANN was 11.1 min. The training time could depend on many factors e.g. how the algorithm is implemented in the programming library, number of inputs used, model complexity, input data representation and sparsity, and feature extraction. For random forest models, training time could vary from problem to problem and other factors e.g. number of trees used to generate a random forest can influence the training time.

6. Conclusions

Energy prediction plays a significant role in making informed decisions by facility managers, building owners, and for making planning decisions by energy providers. In the past, linear regression, ANN, SVM, etc. were developed to predict energy consumption in buildings. Other machine learning techniques (e.g. ensemble based algorithms) are less popular in building energy research domain. Recent advancements in the computing technologies have led to the development of many accurate and advanced ML techniques. Also, the best machine learning technique cannot be chosen a priori, and different algorithms need to be evaluated to find their applicability for a given problem. This study compared the performance of the widely-used feed-forward back-propagation artificial neural network (ANN) with random forest (RF), an ensemble-based method gaining popularity in prediction – for predicting HVAC electricity consumption of a hotel in Madrid, Spain. The paper compared their performance to evaluate their applicability for predicting energy consumption. Based on the performance metrics (RMSE, MAPE, MAD, CV, and R^2) used in the paper, it is found that ANN performed marginally better than the decision tree based algorithm (random forest). Both of these models performed better on training and validation datasets. ANN showed higher accuracy on a recently acquired data (testing dataset). However, from results, it is concluded that both of the models can be feasible and effective for predicting hourly HVAC electricity consumption.

In built environment research community, ensemble-based methods (including random forest, extremely randomised tree, etc.) have often been ignored despite being gained considerable attention in other research fields. This paper explored RF as an alternative method for predicting building energy use and prompted the readers to explore the usefulness of RF and other tree based algorithms. Random forests have been developed to overcome the shortcomings of CART (classification and regression trees). The main drawback of CART methodology was that the final tree is not guaranteed to be the optimal tree and to generate a stable model. Decision trees are mostly unstable, and significant changes in the variables used in the splits can occur due to a small change in the learning sample values. RF can be used for handling high-dimensional data, performs internal cross-validation (i.e. using OOB (out-of-bag) samples) and only has a few tuning parameters. By default, RF uses all available input variables simultaneously and therefore we have to set a maximum number of variables (based on a variable selection procedure). The developed model will facilitate an understanding of complex data, identify trends and analyse what-if scenarios. The developed model will also help energy managers and building owners to make informed decisions. The developed model will be incorporated into a software module (Expert system module), which will enable the user to make informed decisions, identify gaps between predicted and expected energy consumption, identifying reasons for these gaps along with their probabilities, detect and diagnose any faults in the system. There is also a need to investigate the performance of different ensemble based algorithms e.g. Extremely randomised tree [45], Gradient Boosted Regression Trees (GBRT) [46] against random forest and other machine learning techniques (e.g. ANN, SVM, etc.) for energy predictions. Big Data technologies need to be explored for training and deploying future machine learning models. Future work will also explore the possibility of using random forest based prediction models for near real-time HVAC control and optimisation applications. Future work will also investigate the optimal number of previous hour's energy prediction to improve prediction accuracy. Future studies will also examine the impact of temporal and spatial granularity on model's prediction accuracy.

Acknowledgements

The authors acknowledge financial support from the European Commission Seventh Framework Programme (FP7), grant reference – 609154. They would also like to thank Francisco Diez Soto (Engineering Consulting Group) and Jose María Blanco (Iberostar) for providing valuable historical data for this project.

References

[1] M.W. Ahmad, M. Mourshed, D. Mundow, M. Sisinni, Y. Rezgui, Building energy metering and environmental monitoring – a state-of-the-art review and directions for future research, Energy Build. 120 (2016) 85–102, ISSN 0378-7788, doi: http://dx.doi.org/10.1016/j.enbuild.2016.03.059.

[2] IPCC, Climate Change 2014. Synthesis Report-Contribution of Working Groups I, II and III to the Fifth Assessment Report, Intergovernmental Panel on Climate Change, Geneva, Switzerland, 2014.

[3] F. McLoughlin, A. Duffy, M. Conlon, A clustering approach to domestic electricity load profile characterisation using smart metering data, Appl. Energy 141 (2015) 190–199, http://dx.doi.org/10.1016/j.apenergy.2014.12.039.

[4] M. Mourshed, S. Robert, A. Ranalli, T. Messervey, D. Reforgiato, R. Contreau, A. Becue, K. Quinn, Y. Rezgui, Z. Lennard, Smart grid futures: perspectives on the integration of energy and ICT services, Energy Procedia 75 (2015) 1132–1137, http://dx.doi.org/10.1016/j.egypro.2015.07.531.

[5] X. Li, J. Wen, Review of building energy modeling for control and operation, Renew. Sustain. Energy Rev. 37 (2014) 517–537.

[6] M. Mourshed, D. Kelliher, M. Keane, Integrating building energy simulation in the design process, IBPSA News 13 (1) (2003) 21–26.

[7] M.W. Ahmad, J.-L. Hippolyte, J. Reynolds, M. Mourshed, Y. Rezgui, Optimal scheduling strategy for enhancing IAQ, visual and thermal comfort using a genetic algorithm, in: Proceedings of ASHRAE IAQ 2016 Defining Indoor Air Quality: Policy, Standards and Best Paractices, 2016, pp. 183–191.

[8] M.W. Ahmad, M. Mourshed, J.-L. Hippolyte, Y. Rezgui, H. Li, Optimising the scheduled operation of window blinds to enhance occupant comfort, in: Proceedings of BS2015: 14th Conference of International Building Performance Simulation Association, Hyderabad, India, 2015, pp. 2393–2400.

[9] M.W. Ahmad, M. Mourshed, B. Yuce, Y. Rezgui, Computational intelligence techniques for HVAC systems: a review, Build. Simul. 9 (4) (2016) 359–398, http://dx.doi.org/10.1007/s12273-016-0285-4, ISSN 1996-8744.

[10] A. Statnikov, L. Wang, C.F. Aliferis, A comprehensive comparison of random forests and support vector machines for microarraybased cancer classification, BMC Bioinform. 9 (1) (2008) 1.

[11] L. Pérez-Lombard, J. Ortiz, C. Pout, A review on buildings energy consumption information, Energy Build. 40 (3) (2008) 394–398.

[12] K. Grolinger, A. L'Heureux, M.A. Capretz, L. Seewald, Energy forecasting for event venues: big data and prediction accuracy, Energy Build. 112 (2016) 222–233, http://dx.doi.org/10.1016/j.enbuild.2015.12.010, ISSN 0378-7788.

[13] P.A. González, J.M. Zamarreño, Prediction of hourly energy consumption in buildings based on a feedback artificial neural network, Energy Build. 37 (6) (2005) 595–601, http://dx.doi.org/10.1016/j.enbuild.2004.09.006, ISSN 0378-7788.

[14] S.J. Nizami, A.Z. Al-Garni, Forecasting electric energy consumption using neural networks, Energy Policy 23 (12) (1995) 1097–1104, http://dx.doi.org/10.1016/0301-4215(95)00116-6, ISSN 0301-4215.

[15] A.E. Ben-Nakhi, M.A. Mahmoud, Cooling load prediction for buildings using general regression neural networks, Energy Convers. Manag. 45 (13-14) (2004) 2127–2141, http://dx.doi.org/10.1016/j.enconman.2003.10.009, ISSN 0196-8904.

[16] S.A. Kalogirou, M. Bojic, Artificial neural networks for the prediction of the energy consumption of a passive solar building, Energy 25 (5) (2000) 479–491, http://dx.doi.org/10.1016/S0360-5442(99)00086-9, ISSN 0360-5442.

[17] J. Kreider, D. Claridge, P. Curtiss, R. Dodier, J. Haberl, M. Krarti, Building energy use prediction and system identification using recurrent neural networks, J. Sol. Energy Eng. 117 (3) (1995) 161–166.

[18] Y. Cheng-wen, Y. Jian, Application of ANN for the prediction of building energy consumption at different climate zones with HDD and CDD, in: Future Computer and Communication (ICFCC), 2010 2nd International Conference on, vol. 3, 2010, pp. 286–289, http://dx.doi.org/10.1109/ICFCC.2010.5497626.

[19] A. Azadeh, S. Ghaderi, S. Sohrabkhani, Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors, Energy Convers. Manag. 49 (8) (2008) 2272–2278, http://dx.doi.org/10.1016/j.enconman.2008.01.035, ISSN 0196-8904.

[20] Z. Yu, F. Haghighat, B.C. Fung, H. Yoshino, A decision tree method for building energy demand modeling, Energy Build. 42 (10) (2010) 1637–1646, http://dx.doi.org/10.1016/j.enbuild.2010.04.006, ISSN 0378-7788.

[21] G.K. Tso, K.K. Yau, Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks, Energy 32 (9) (2007) 1761–1768, http://dx.doi.org/10.1016/j.energy.2006.11.010, ISSN 0360-5442.

[22] T. Hong, C. Koo, K. Jeong, A decision support model for reducing electric energy consumption in elementary school facilities, Appl. Energy 95 (2012) 253–266, http://dx.doi.org/10.1016/j.apenergy.2012.02.052, ISSN 0306-2619, http://www.sciencedirect.com/science/article/pii/S0306261912001511.

[23] T. Hong, C. Koo, S. Park, A decision support model for improving a multi-family housing complex based on {CO2} emission from gas energy consumption, Build. Environ. 52 (2012) 142–151, http://dx.doi.org/10.1016/j.buildenv.2012.01.001, ISSN 0360-1323, www.sciencedirect.com/science/article/pii/S0360132312000029.

[24] S.S. Haykin, Neural Networks: A Comprehensive Foundation, Macmillan, 1994, ISBN 9780023527616.
[25] A. Krenker, J. Bester, A. Kos, Introduction to the Artificial Neural Networks. Artificial Neural Networks: Methodological Advances and Biomedical Applications, 2011, pp. 1–18.
[26] R.D. Reed, R.J. Marks, Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks, MIT Press, 1998.
[27] R. Rojas, Neural Networks: A Systematic Introduction, Springer Science & Business Media, 2013.
[28] K. Ermis, A. Erek, I. Dincer, Heat transfer analysis of phase change process in a finned-tube thermal energy storage system using artificial neural network, Int. J. Heat Mass Transf. 50 (15) (2007) 3163–3175.
[29] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, John Wiley & Sons, 2012.
[30] B. Larivière, D.V. den Poel, Predicting customer retention and profitability by using random forests and regression forests techniques, Expert Syst. Appl. 29 (2) (2005) 472–484, http://dx.doi.org/10.1016/j.eswa.2005.04.043, ISSN 0957-4174.
[31] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[32] R. Jiang, W. Tang, X. Wu, W. Fu, A random forest approach to the detection of epistatic interactions in case–control studies, BMC Bioinform. 10 (1) (2009) 1.
[33] V. Svetnik, A. Liaw, C. Tong, J.C. Culberson, R.P. Sheridan, B.P. Feuston, Random forest: a classification and regression tool for compound classification and QSAR modeling, J. Chem. Inf. Comput. Sci. 43 (6) (2003) 1947–1958.
[34] F.L. Quilumba, W.J. Lee, H. Huang, D.Y. Wang, R.L. Szabados, Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities, IEEE Trans. Smart Grid 6 (2) (2015) 911–918, http://dx.doi.org/10.1109/TSG.2014.2364233, ISSN 1949-3053.
[35] C. Fan, F. Xiao, S. Wang, Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques, Appl. Energy 127 (2014) 1–10.
[36] R.E. Edwards, J. New, L.E. Parker, Predicting future hourly residential electrical consumption: a machine learning case study, Energy Build. 49 (2012) 591–603.
[37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (Oct) (2011) 2825–2830.
[38] Neurolab, Neural Network Library for Python, 2011 http://pythonhosted.org/neurolab/.
[39] R. Hecht-Nielsen, Theory of the backpropagation neural network, in: International Joint Conference on Neural Networks-IJCNN, vol. 1, 1989, pp. 593–605, http://dx.doi.org/10.1109/IJCNN.1989.118638.
[40] M. Rafiq, G. Bugmann, D. Easterbrook, Neural network design for engineering applications, Comput. Struct. 79 (17) (2001) 1541–1552.
[41] Q. Li, Q. Meng, J. Cai, H. Yoshino, A. Mochida, Predicting hourly cooling load in the building: a comparison of support vector machine and different artificial neural networks, Energy Convers. Manag. 50 (1) (2009) 90–96.
[42] J.C. Principe, N.R. Euliano, W.C. Lefebvre, Neural and Adaptive Systems: Fundamentals Through Simulations with CD-ROM, John Wiley & Sons, Inc., 1999.
[43] S. Vincenzi, M. Zucchetta, P. Franzoi, M. Pellizzato, F. Pranovi, G.A.D. Leo, P. Torricelli, Application of a random forest algorithm to predict spatial distribution of the potential yield of Ruditapes philippinarum in the Venice lagoon, Italy, Ecol. Model. 222 (8) (2011) 1471–1478, http://dx.doi.org/10.1016/j.ecolmodel.2011.02.007, ISSN 0304-3800.
[44] D.S. Siroky, et al., Navigating random forests and related advances in algorithmic modeling, Stat. Surv. 3 (2009) 147–163.
[45] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, Mach. Learn. 63 (1) (2006) 3–42.
[46] J.H. Friedman, Stochastic gradient boosting, Comput. Stat. Data Anal. 38 (4) (2002) 367–378.