Energy and Buildings 111 (2016) 184–194

Contents lists available at ScienceDirect

Energy and Buildings

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

Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings

Young Tae Chae [a,*], Raya Horesh [b], Youngdeok Hwang [b], Young M. Lee [b]

[a] *Department of Architectural Engineering, Cheongju University, 298 Daesung-ro, Cheongju-si, Republic of Korea*
[b] *IBM, T.J. Watson Research Center, 1101 Kitchwan Road, Yorktown Heights, United States*

ARTICLE INFO

Article history:
Received 1 October 2014
Received in revised form 13 October 2015
Accepted 14 November 2015
Available online 19 November 2015

Keywords:
Building electricity consumption
Short-term load forecasting
Artificial Neural Network
Bayesian regularization

ABSTRACT

Short-term load forecasting of building electricity usage is of great importance for anomaly detection on electricity usage pattern and management of building energy consumption in an environment where electricity pricing is dynamically determined based on the peak energy consumption. In this paper, we present a data-driven forecasting model for day-ahead electricity usage of buildings in 15-minute resolution.

By using variable importance analysis, we have selected key variables: day type indicator, time-of-day, HVAC set temperature schedule, outdoor air dry-bulb temperature, and outdoor humidity as the most important predictors for electricity consumption. This study proposes a short-term building energy usage forecasting model based on an Artificial Neural Network (ANN) model with Bayesian regularization algorithm and investigates how the network design parameters such as time delay, number of hidden neurons, and training data effect on the model capability and generality.

The results demonstrate that the proposed model with adaptive training methods is capable to predict the electricity consumption with 15-minute time intervals and the daily peak electricity usage reasonably well in a test case of a commercial building complex.

1. Introduction

Forecasting electricity load has become one of the more important topics recently, not only for the electricity power suppliers, but also for the consumers of the electricity, especially for commercial and industrial buildings. Building energy administrators, such as building owners or engineers, need to have an accurate forecast of energy demand in the near future or one day ahead to be able to better manage energy usage. Although the forecast time horizon can range from minutes to years, the short-term load forecast (STLF), especially for a period shorter than a day, has been more of an interest in the perspective of buildings because the utility prices may change by seasonality, time-of-use in on/off peak period, and contract demand [1,2].

A large variety of forecasting models and approaches, such as regression models, time series models, and machine-learning-based models have been developed and used. Regression models, in which the load forecasting is formulated usually as a linear function of input variables, have been effective to predict building energy consumption with number of experiments [3,4]. These models are attractive because the model components can have direct physical interpretation, such as the total load, base load, heating load, cooling load, and weather components, etc.

Time series models assume that the current and future energy usage is a function of the past observed energy usage. Examples include the autoregressive integrated moving average (ARIMA) [5], multiplicative autoregressive models [6], autoregressive moving average with exogenous input model (ARMAX) [7], Kalman filtering [8], and Fourier series model [9] among many. Although these models can predict short-term load pattern, they may be unstable in the case of nonlinear load or non-stationary conditions, due to its restrictive assumptions.

Other machine learning techniques have received significant attention in the context of STLF problems since the late 1980's including Support Vector Machine (SVM) model [10,11], neuro-fuzzy system [12], and Artificial Neural Networks (ANNs) for both electricity supply and demand side [13–15]. Several studies have shown that ANNs have produced better results compared with other approaches [16,17].

Although STLF models have been used in practice with different degree of success, these were mostly designed and evaluated with the hourly load forecasting. However, if the utility price depends on time of use, season, and the characteristics of building energy usage pattern in sub-hourly time resolutions [18], building

* Corresponding author. Tel.: +82-43-229-8479; fax: +82-43-229-8479.
E-mail addresses: ychae@cju.ac.kr, y.t.chae@gmail.com (Y.T. Chae).

http://dx.doi.org/10.1016/j.enbuild.2015.11.045
0378-7788/

Nomenclature

MSE Mean squared error, $\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(S_i - M_i)^2$ where,M: measured energy consumption in the period (n), S:simulated energy consumption in the period (n).

MBE Mean bias error, $\mathrm{MBE} = \frac{\sum_{i=1}^{n}(S_i - M_i)}{\sum_{i=1}^{n} M_i}$ where, M:measured energy consumption in the period (n), S:simulated energy consumption in the period (n).

APE Absolute percentage error, $\mathrm{APE} = \frac{|S-M|}{M} \times 100$ where, M:measured energy consumption in a time step, S:simulated energy consumption in a time step.

CV(RMSE) Coefficient of variation of the root-mean-squared error, $\mathrm{CV(RMSE)} = \frac{\mathrm{RMSE}}{A}$, $\mathrm{RMSE} = \sqrt{\mathrm{MSE}}$, and $A = \frac{\sum_{i=1}^{n} M_i}{n}$, where RMSE: root-mean-square-error, A:mean of measured data

engineers need a precise estimation of the peak electricity load in a day and the load shape to better control the utility costs. In addition, the energy forecast can provide useful information in making an electricity purchase plan when the building has an on-site electricity generation system with renewable energy sources under a dynamic electricity pricing grid [19]. For this problem, the hourly load forecast model may be insufficient to estimate the peak demand and to plan on-site energy generation.

Previous studies on STLF for the sub-hourly electricity consumption of buildings are limited. Escriva-Escriva et al. [14] proposed STLF model using ANNs to forecast the building energy consumption using 96 time steps for a day using an independent algorithm that first searches for four or more days which have similar outdoor air temperature patterns and types of day in the previous year for model training. Then the algorithm trains 96 ANNs for each time interval. This model requires an entire whole year's data set and the performance may not be stable when the energy consumption pattern has large daily or annual variability. Therefore, it is useful to explore a model that can perform well under more general setting, in particular, not requiring a large amount of input data for forecasting electricity usage of buildings.

For this need, we develop a short-term load forecasting model using data mining and machine learning technique while assuming limited availability of data. In particular, we investigate ANNs model to predict the energy consumption of a commercial building complex. This paper is structured as follows. The building complex for the case study and data processing approaches are illustrated in Section 2. The section also presents a feature extraction identifies key predictors that impacts the electricity consumption of the case, as well as a model selection procedure that selects the most suitable machine learning algorithm for this problem. Section 3 describes the sensitivity analysis study on the configuration variables of the model and implementation results, which are compared with the actual measurement of energy use. Lastly, in Section 4 summary of the results and future research directions are presented.

2. Methodological approach

2.1. Description of a case study: a building complex

건물 단지

All data set for this study was obtained from a building management system (BMS) of a commercial office building complex, and the data are periodically pulled into a relational database (IBM DB2™). The site consists of three office buildings in urban area, each of which has different number of floors; five in building 1 (BLDG1), four in building 2 (BLDG2), and two in building 3 (BLDG3). A total floor area of 15,224 m^2 spreads over typical office area, small laboratories, cafeteria, parking garage, and small gymnasium (Fig. 1). Although the buildings are separated, they are all managed by one utility billing system.

Two absorption chiller systems provide chilled water in summer and hot water in winter for a constant air volume (CAV) system for each floor of BLDG1 and fan coil units (FCU) of perimeter zones of each floor of BLDG2. All three buildings have electric heat pump (EHP) systems with multi-indoor units. EHP systems are supplementary system incorporated with CAV systems for BLDG1, but operate as the main air-conditioning system for BLDG2 and BLDG3. BMS system monitors operational conditions of both primary/secondary system and EHPs. The system also controls all secondary system operation, whilst EHPs are locally controlled by occupants' indoor thermal demand. For the electricity usage monitoring, one main electric meter and several sub-meters are installed as illustrated in Fig. 2. The main meter measures electricity usage, both the instantaneous power in kW with a minute interval and aggregated electricity usage at every 15 minutes in kWh.

The short-term monitoring of electricity usage is very important to the building engineers since 15-minute peak energy consumption in the current year impacts the annual utility expenditure of the following year. According to the actual electricity price of the building complex in Table 1 from the electricity grid supplier [18], a monthly electricity bill is composed of a demand and meter charge for commercial buildings. The monthly demand charge, except in residential buildings, is set by a peak usage in three summer months, July, August, and September of the previous year. The supplier sets a building's peak usage as the maximum electricity usage for 15-minute intervals during the three months. If the peak electricity usage exceeded that of the last year at a certain time interval, the monthly demand charge is reset with the new peak usage for the next 12 months.

2.2. Data collection and processing

The relational database management system (RDBMS) has been used to collect and store environmental variables, BMS data, and electricity usage as illustrated in Fig. 3. Software engineers, building engineers, and the predictive model developers worked together to create data schema for RDBMS. Although the data set has over 1000 data points with 20 different measurement types, some variables may not eventually be used for electricity usage forecast. For example, the supply and return air temperature of an air handling unit are important in estimating the space thermal load and the electricity usage by the primary/secondary system, but they are not available at the time when forecast is made. Thus after taking the data availability into account, the predictors are divided into three categories: environmental data, time indicator, and operational condition, as illustrated in Table 2.

For the environmental variables, the database retrieves data from local weather forecasting service. The data include outdoor air temperature, relative humidity, wind speed and direction, sky condition, and precipitation type in every 3-hour interval for the next 72 hours. The forecast data is updated eight times a day. RDBMS system converts string variables like sky condition and wind

3시간마다.
하루 8번

Fig. 1. Typical floor plans of the case study building complex.

Table 1
Electric rate for the target building complex[a] [18].

Classification		Demand charge [$/kW]	Energy charge [$/kWh]			
			Time period[c]	Summer (Jun.1–Aug. 31)	Spring/Fall (Mar.1–May.31/Sep.1–Oct.31)	Winter (Nov.1–Feb. 28)
High-voltage A	Option I	7.22	Off-peak	0.0627	0.0627	0.0714
			Mid-load	0.1139	0.0701	0.1018
			Peak-load	0.1364	0.0814	0.1166
	Option II	8.23	Off-peak	0.0574	0.0574	0.0661
			Mid-load	0.1086	0.0648	0.0965
			Peak-load	0.1311	0.0761	0.1113
	Option III[b]	9.81	Off-peak	0.0552	0.0552	0.0625
			Mid-load	0.1084	0.0773	0.1086
			Peak-load	0.1787	0.1010	0.1555

[a] Category: General (B)-Commercial, contract demand 300 kW or more.
[b] Option III: Use electricity for 500 h or more per month.
[c] Time period:
Off-peak: 23:00~09:00 (Summer, Spring/Fall, Winter)
Peak: 11:00~12:00/13:00~17:00 for Summer and Spring/Fall,
10:00~12:00/17:00~20:00/22:00~23:00 for Winter.
Mid-load: Out of Off-peak and Peak time period.

Table 2
Potential predictor variables.

Category	Variables	Unit/index
Environment	Outdoor dry-bulb temp. (ODT)	°C
	Outdoor relative humidity (ODH)	%
	Precipitation probability (PPT)	%
	Rain indicator (RAN)	0: No rain 1: Rain 2: Snow 3: Rain and snow
	Wind speed (WSD)	km/h
	Sky condition (SKC)	0: Clear sky1: Partially cloud2: Cloud3: Overcast
Time indicator	Day indicator (DTF)	0: Weekdays 1: Saturday 2: Sunday
	Interval stamp (TIF)	0-95
Operational condition	HVAC operation schedule(OPC)	Discretized

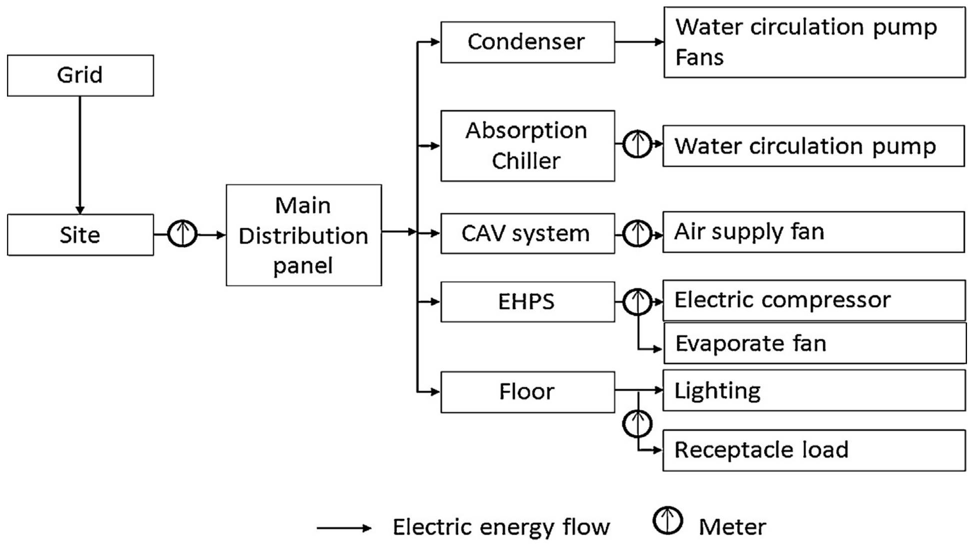


Fig. 2. Electricity delivery hierarchy and meter installation at the site.

direction to numeric variables. Each parameter is interpolated to obtain higher time resolution, as desired for the electricity prediction. The time indicator variables were transformed to integer values from 0 to 95 and day types (e.g., normal working day of Monday to Friday, Saturday, Sunday) were interpreted into integer variables. The categorical variables like day types and wind directions can be transformed to a set of indicator variables; for example, (1 = 1,0,0) for Weekdays, (2 = 0,1,0) for Saturday, (3 = 0,0,1) for Sunday. An averaged system operation schedule of five CAV systems was used as the HVAC operational condition.

2.3. Feature extraction for dimension reduction

As described earlier, the potential predictors are nine independent variables. However, those variables influence the electricity usage of the building in a different way. If some variables in the input data set are irrelevant to the output, it decreases model accuracy, stability, and effectiveness. Therefore, it is necessary to pre-screen the variables by identifying the important variables from the input data set.

Random forests algorithm [20] was used to assess the importance of variables by measuring the candidate parameters in terms of their impacts on the response of prediction. This algorithm is one of the most popular ensemble learning methods for classification and regression. To train the model, first bootstrap samples based on tree size (m_{tree}) are generated from original sample. Then each bootstrap sample grows an un-pruned decision tree by splitting of randomized selection (m_{try}) of the input variables, and then it predicts new data with combining the prediction of number of trees.

To assess the importance of a variable, it measures variable importance in a data with initial fitting a random forest to the data

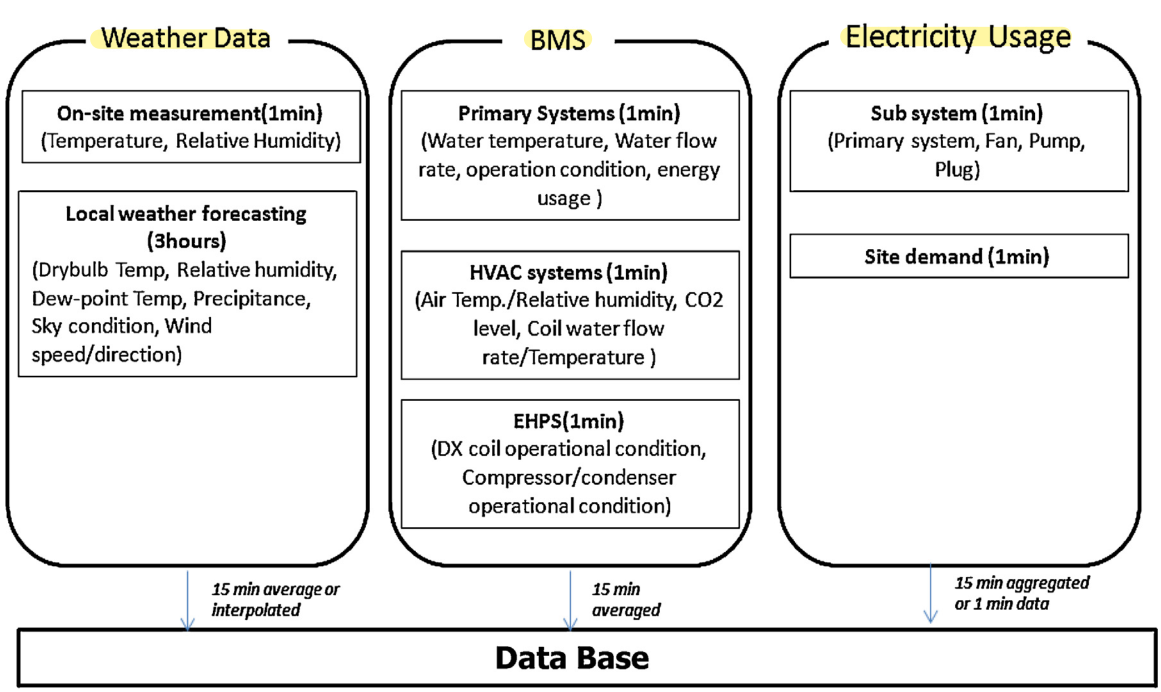


Fig. 3. Data collecting system.

Table 3
Variable importance ranked by the conditioned permutation-importance.

Variables	Conditioned permutation-importance
TIF	482.27
DTF	354.03
OPC	296.50
ODT	222.01
ODH	212.01
PPT	92.56
WSD	90.42
SKC	81.95
RAN	32.36

as a training condition. During the process, the out-of-bag error is collected over the forest. After training period, each variable is randomly permuted along the training data, and the out-of-error is calculated again. Then the importance score is computed by difference of the error before and after the shuffle. The permutation importance and node impurity were used to select variables in this study. The permutation importance estimates the importance of each predictor by considering prediction error variation when the variable is permuted but others are kept the same. Node impurity is measured by the total decrease in residual sum of squares from splitting on the variable, averaged over all trees. Both provide a relative significance of a certain variable in the predictor group. In addition, it also investigates a conditional permutation-importance measure, because the predictors are of different type and some of them are correlated with others [21]. In this study, m_{try} and m_{tree} were increased until mean square error (MSE) in the prediction was minimized.

Variables are ranked by both relative importance metrics in Fig. 4. It can be seen that the time indicator (day type and time interval stamp) and the operation condition have a strong influence on the electricity usage of the building in July. From the environmental variables, outdoor air temperature and relative humidity contribute to energy usage while other weather variables are of less importance. It has a similarity to the conditional permutation-importance in Table 3. According to permutation importance and node impurity, the operational condition was one of the most important factors. It shows that the operational condition of the secondary system is useful to capture the actual activity in the building such as occupancy and electricity consumptions of lighting systems and receptacles.

As shown in Fig. 4 and Table 3, five variables are ranked highly in their importance; the operational condition, time indicator, day type, outdoor dry-bulb temperature, and outdoor relative humidity. These five variables are selected as input attributes, together with previous electricity usages.

2.4. *Predictive model selection*

Although many machine-learning algorithms are available, the choice of the specific method to use is not trivial and depends strongly on the specific application and data availability and type. In this study, we examined nine different machine-learning algorithms and chose the one that performed best. It should be noted that in case where the model is to be used as part of a model predictive control (MPC) optimization problem, preferably differentiable algorithms should be considered, such as regression type algorithms.

Gaussian process assumes the observed data points as a realization of a distribution over functions, where a prediction is a weighted average of the observed data points; different covariance function calculates the distance differently, hence produces different prediction [22]. Linear regression is modelling the relationship between scalar dependent variable and independent variables. The

Table 4
Evaluation of different machine-learning algorithms.

Algorithm	Correlation coefficient	CV(RMSE)
Gaussian process with radial basis function (RBF) kernel	0.94	0.11
Gaussian process with polynomial kernel	0.87	0.15
Linear regression	0.83	0.16
Artificial neural network	0.96	0.08
Support vector machine (SVM) with normalized polynomial kernel	0.92	0.13
SVM with RBF kernel	0.88	0.14
K-Star classifier	0.92	0.12
Nearest neighbour ball tree	0.94	0.11
Simple model	0.81	0.18

data is modelled using linear predictor functions, with associated parameters, which are interpreted from the data. The parameters are estimated using least square methods [23]. Support Vector Machine (SVM) is a supervised learning method based on kernel function and it is used for classification and function approximation. Using specific kernel function, original parameter space is transformed into a high-dimensional space where a separated hyperplane is constructed with the maximum-margin [24]. K-Star classifier is an instance-based learner that uses entropy as a distance measure. Instance-base learners classify an instance by comparing it to a database of pre-classified examples. Simple instances are similarly classified, where the similarity is measured by the distance metric [25]. Nearest-neighbour ball tree is a nearest neighbour algorithm performed on a ball tree dataset space partitioning. Ball tree is a binary tree in which every note defines a D-dimensional hypersphere. Each point is assigned to one ball (or hypersphere) according to the distance from the ball's center [26].

Since the model is not part of an optimization scheme in this study, differentiability consideration is not important and we can account for all types of machine learning algorithms, including tree-based methods that are not differentiable. In addition to machine-learning-based short-term load forecasting models, this study also tested a very naive model, which just assumes that the forecasted energy consumption profile of the target day in 15-minute intervals would be identical to the previous weekday's profile. In Table 4, we summarize the results for different algorithms that we have compared. We quantify the quality of the results based on two measures: correlation coefficient, and coefficient variance of root mean square error (CV(RMSE)). The training set was selected to be the first two weeks in July 2012 and the prediction of the total electricity usage was performed on the following three days. The input attributes for the model were time indicators; outdoor dry-bulb temperature, outdoor relative humidity, and HVAC operation schedule. In the case of the naive model, the forecasts are simply replicates of the profile of the corresponding previous weekday.

Although the naive approach yields R-square value of 0.81 and CV(RMSE) of 0.18 for test days, model-based forecasting approaches outperforms the naive model in both metrics. In particular, the artificial neural network (ANN) performs better than any other machine-learning algorithms that were considered for this problem. For the rest of the paper, we focus on the ANN method only.

2.5. *ANN model architecture*

The ANN model may have benefits to reduce run time with relatively high accuracy compared to other methods [27]. In addition, it may not be necessary to establish the relationship between prediction and dependent (output) variables before the model

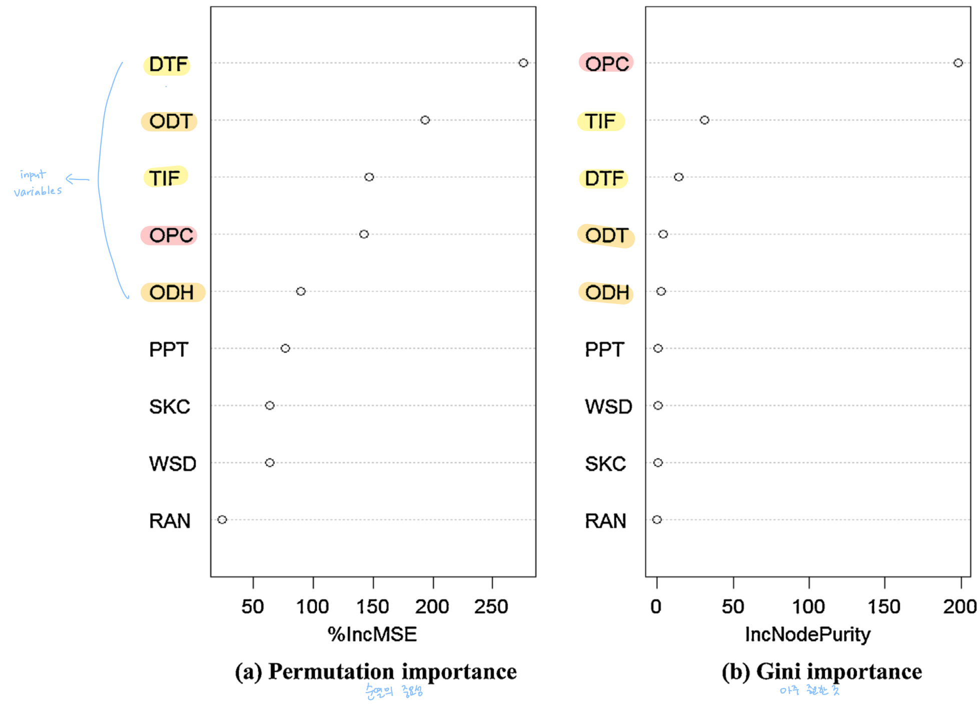


Fig. 4. Relative variable importance.

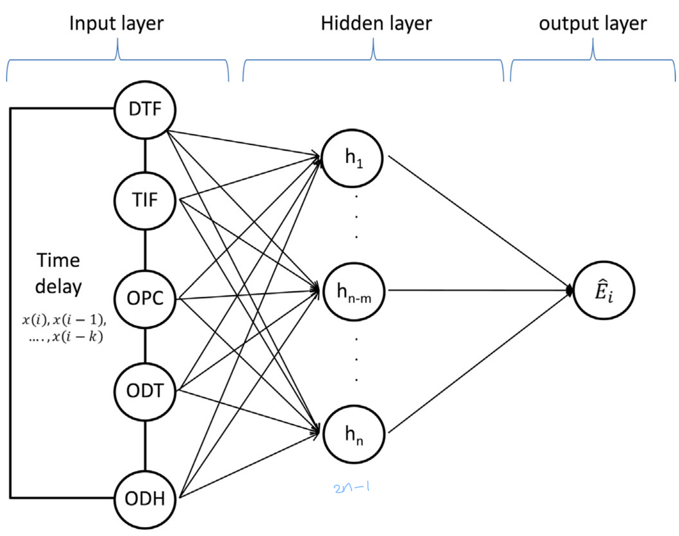


Fig. 5. Artificial neural network structure.

implementation since this relationship can be identified by a supervised learning process of the neural network.

The ANN model in this study has a conventional multi-layered feedforward network using a backpropagation algorithm as illustrated in Fig. 5. It has three layers: input, hidden, and output [28]. The input layer includes n neurons depending on the number of input data variables, and neurons' interconnecting weights are adjusted based on an error signal, which is derived by the difference of the actual energy consumption at a time instance, E_i, and estimated consumption, $\hat{E}_i$ at the same instance ($i = 0$ to 95) [29,30]. The output layer has a single neuron for the dependent variable. In the hidden layer, there are $2n+1$ neurons for a preliminary structure [31].

Even if the neural network configuration has been successfully implemented into various prediction problems with a time series data, it fails frequently to provide a generalized result. When an

Table 5
The evaluation result with different training data sizes.

Training size	CV(RMSE) [%]	MBE [%]	Peak usage APE range [%]
1 week	11.98	−10.4	1.5–5.2
2 weeks	8.2	−5.6	2.4–5.9
3 weeks	8.0	−2.3	2.2–4.3
4 weeks	7.3	0.03	1.8–3.7

ANN model is trained by a noisy data with the short-term intervals, for example, the actual forecasting performance can be unstable or very poor even if it can fit the training data set well. This situation is called "overfitting" and it happens often in various applications [32].

A Bayesian regularized neural network model with Levenberg–Marquart (LM) backpropagation algorithm is employed for the training process to improve the generalization of model. In this approach, the objective function includes both the conventional error function and the weight decay components or penalty term. The weights and biases in the model are assumed to be random variables with Gaussian distribution and the regularization parameters in the objective function can be optimized by using Bayesian rules [33,34].

The network was built by using MATLAB's *trainbr* function. The logistic sigmoid functions are used for the activation function in each neuron and a linear transfer function is used to calculate the network output.

A test and validation procedure of the ANN model was conducted under several predictor conditions and data implementations with time delays during the training period. The 15-minute interval data set of E_i and highly ranked five predictor variables (Table 3) were collected from July 1st to July 31st, 2012. By removing data from one day, which has sensor and meter malfunction, 30 days, of which 22 are weekdays and 8 are weekend days, were used. Total of 2880 data points for six input variables, including the electric usage, were used to train the model. The network model selected the training input data set in a random, while three weekdays with a new data set (August 1–3, 2012), due to the weather

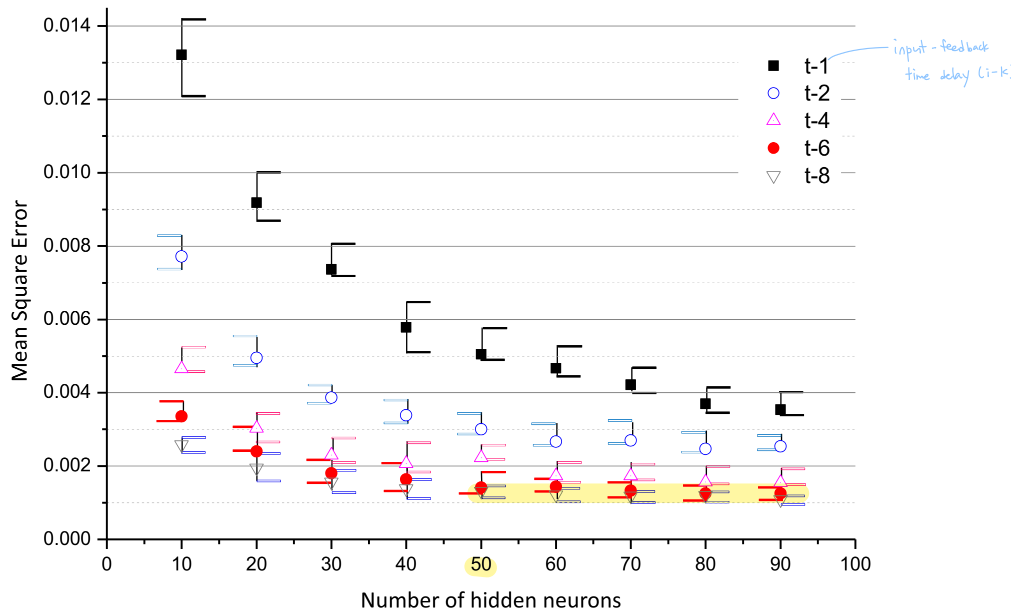


Fig. 6. Average MSE in the training with neuron numbers and time delays.

forecasting time scope, were used to evaluate the out-of-sample testing.

3. Results and discussion

3.1. Network design parameters

3.1.1. Hidden layer architecture and time-delayed inputs

In addition to the initial network configurations such as layer design and training algorithm, the number of neurons in hidden layer and time delay of input variables are highly important factors on the network performance in the training and forecasting stages. If too few parameters are selected, the network cannot capture the complex model dynamics, whilst a model with too many parameters neurons has poor predictive performance as it is easily overfitted by a minor fluctuation in the data. Therefore, it is important to investigate the model performance with varying design parameters.

Figs. 6 and 7 show average (mark) and min/max error range (horizontal bar) of MSE between the actual energy consumption and the model, presented values in the training and evaluation procedures. The total neuron number in the hidden layer is varied from 10 to 90. The input-feedback time delay (i-k) is also parameterized, for the time step k = 1, 2, 4, 6, 8 (up to 2 hours). Each network configuration has been tested by repeating 50 runs to evaluate the stability and robustness of the model performance.

As illustrated in Fig. 6, additional neurons and/or time delay steps improve training performance drastically, reducing not only the average MSE but also min/max error range. In the case of 50 neurons and k = 6, the average and min-max MSE range were 0.0015 and 0.001–0.0018, respectively. But the improvement is stagnant after 50 neurons and k = 6.

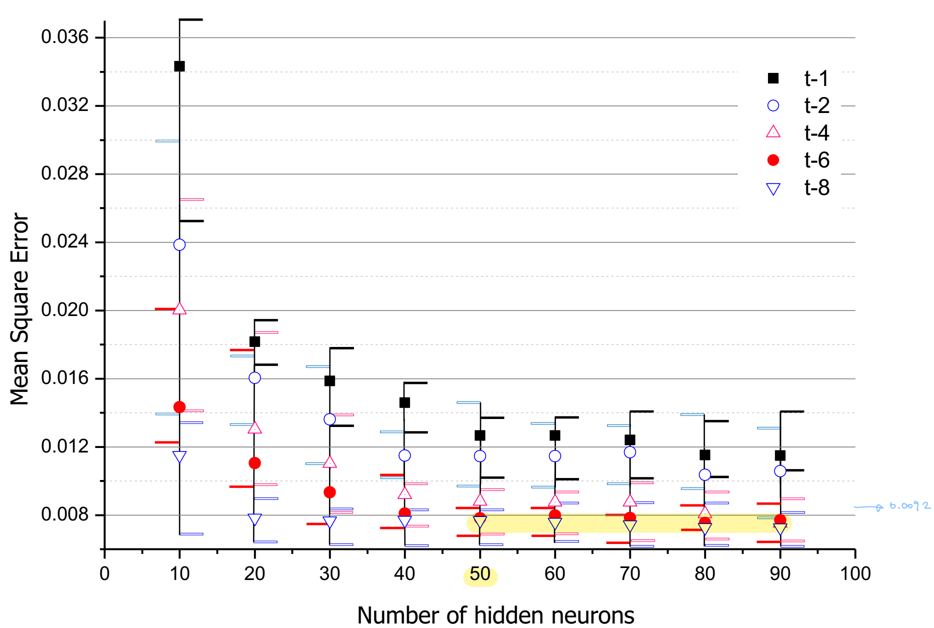
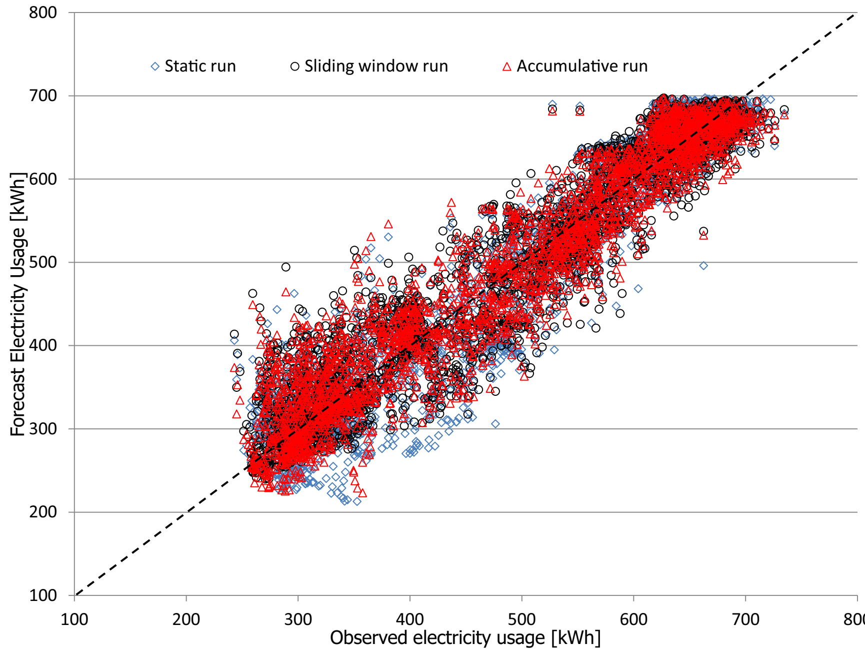
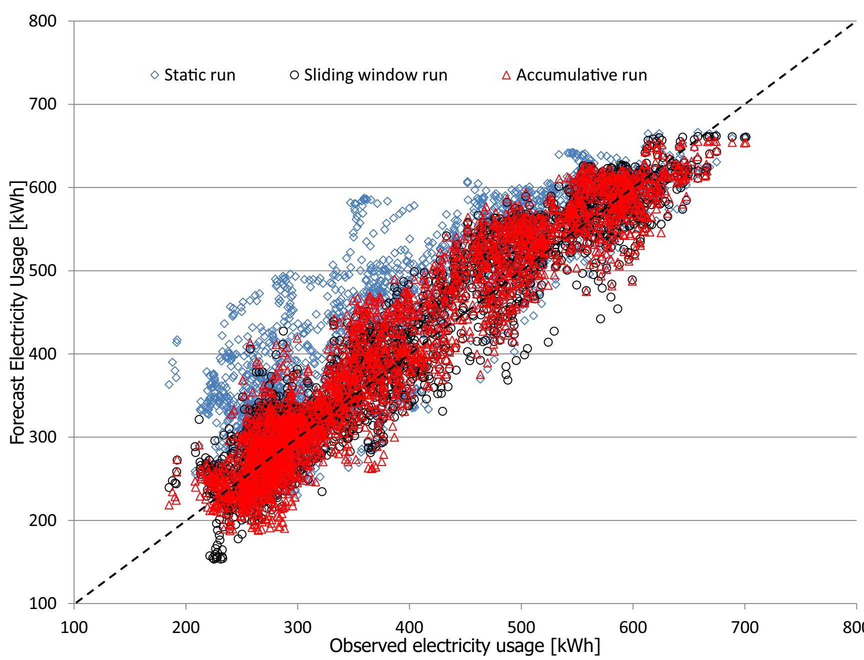


Fig. 7. Average MSE in the evaluation with neuron numbers and time delays.



(a) August



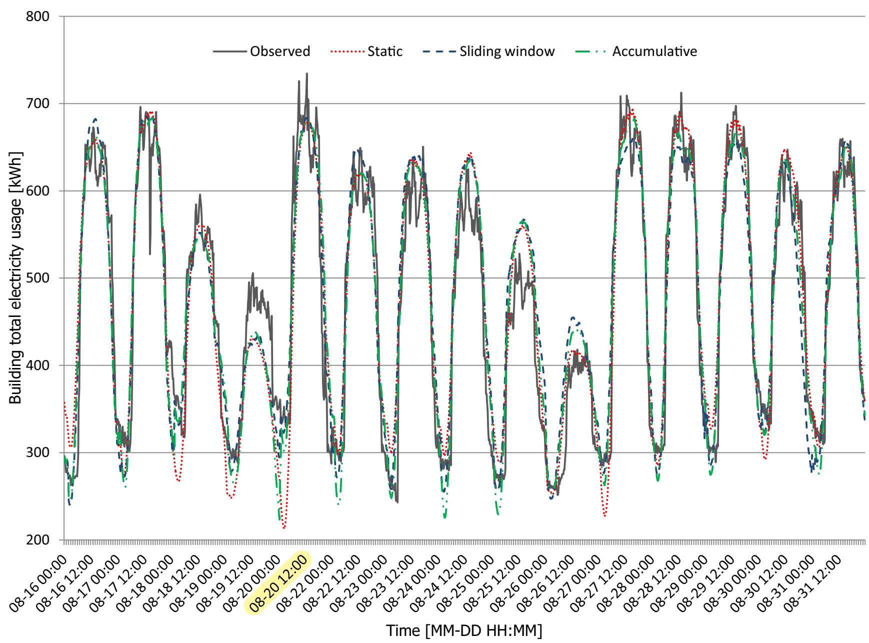
(b) September

Fig. 8. Comparisons of actual observed and forecasted electricity usage.

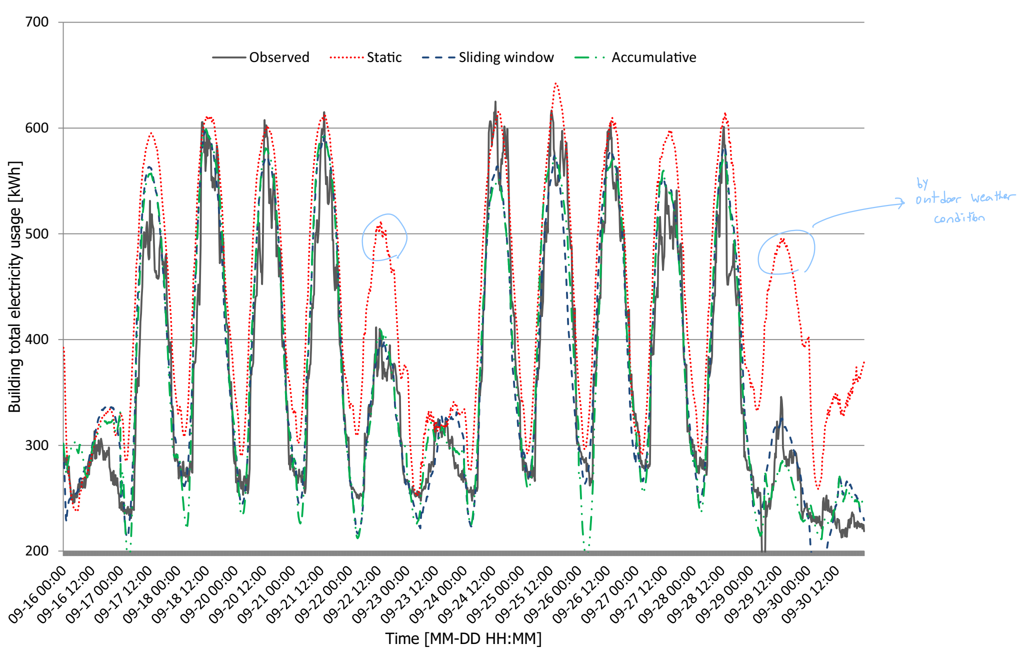
In the evaluation test as shown in Fig. 7, the average MSE for each case and the range of min/max MSE are slightly higher than the case of training procedure because the evaluation procedure uses the calculated electricity usage as a feedback input instead of the actual usage in the training. Similar to what was observed in the training test, the addition of neurons and time delays can improve the evaluation performance and generalization. However, it is observed that the evaluation performance is converged at average MSE of 0.0092. The MSE ranges from 0.0071 to 0.0096 when neuron number and time delay is 50 and $k = 6$, respectively. Although the training algorithm has the regularization function, the network requires a relatively large number of neurons and time delays to provide a stable performance. Considering the model complexity and computation time, it is reasonable to have 50 neurons in the hidden layer and have t-6 time delay for input variables and feedback for the network model in this study.

3.1.2. *Training size for the prediction model*

When constructing the prediction model by training with the historical data set, the size of data set for training plays an important role in model accuracy. If the training data size is small, it takes less time to conduct the training procedure but the model may not be good enough to capture the actual usage pattern, as it might not be general enough. When the historic data volume is large, it requires more time to train the model but the model is more robust and can be more accurate. Table 5 shows the model performance



(a) August



(b) September

Fig. 9. Actual forecast result.

with varying size of training data, from one to four weeks from the validation period. The daily mean bias error (MBE) and (CV(RMSE)) [35] measured at 15-minute interval results are used to measure the model performance. In addition, the absolute percentage error (APE) of daily peak usage is used as another criterion that measures the prediction performance of the model.

As illustrated in Table 5, the larger the training data set, the more accurate the model performance is. Although the model performance for the validation period is reasonable with one week of training data, the model provides the best performance with four weeks' training data set. It suggests that the model can use less training data, especially compared to the other approaches in the same forecasting resolution, to achieve an acceptable performance [14]. There is no significant difference between the model trained with two and three weeks of training data.

3.2. *Model implementation*

The network model parameters such as time lagged measurements, total neuron numbers in the hidden layer, input data type, and training data set size were investigated and established. To verify the forecasting performance of the developed model, two months of data, August and September in year 2012, were used.

Three training methods were considered: static, accumulative, and sliding windows. For the static training, the model is trained using four weeks of data in July, and it forecasts for August and

Table 6
15 min prediction results with each training type.

Month	Day type (days)	Static		Accumulative		Sliding window	
		CV(RMSE) [%]	σ	CV(RMSE) [%]	σ	CV(RMSE) [%]	σ
August	Weekday (21)	8.44	2.41	7.97	2.53	8.74	2.53
	Weekend (8)	11.16	3.61	9.91	2.11	9.62	2.31
September	Weekday (19)	13.76	5.03	9.35	2.14	9.20	2.25
	Weekend (10)	26.74	11.97	11.06	2.44	11.20	3.60

Table 7
Daily peak prediction results with each training type.

Month	Day type (days)	Static			Accumulative			Sliding window		
		Max/min APE [%]	Averaged APE [%]	σ	Max/min APE [%]	Averaged APE [%]	σ	Max/min APE	Averaged APE [%]	σ
August	Weekday (21)	7.57/0.16	2.84	2.79	7.62/0.05	3.37	2.38	8.58/0.02	3.30	2.81
	Weekend (8)	20.33/2.0	8.77	6.67	13.28/3.03	7.28	3.78	13.67/3.75	8.10	2.83
September	Weekday (19)	13.17/0.2	4.40	3.92	11.6/0.2	4.52	2.71	9.7/0.6	4.26	2.64
	Weekend (10)	59.4/4.4	24.9	1.77	17.4/0.6	6.48	3.01	7.7/0.8	4.25	2.80

September without retraining the model with newly available data. Accumulative training, an adaptive training method, uses accumulated data set from the first day of July to the day before the target day and retrained on a daily basis.

The sliding windows method uses a fixed training data window size (four weeks) and the window is shifted by a day by removing the first day of the old training set and adding the new measurements into the data set. The networks are also retrained daily with the new training data.

3.2.1. *Overall prediction result*

Fig. 8 shows the prediction results of 59 days (41 weekdays and 18 weekends) from August to September in 2012 with 15-minute interval resolution. Three days were removed from the evaluation data set; 15th, 21st of August and 19th of September due to the electric meter failure. These outlier days were also removed from the training data set for the next day forecasting.

As illustrated in Fig. 8(a), the overall prediction performance of each training type is similar to each other in August. The coefficient of determination (*R*-square) values is 0.904, 0.912, and 0.902 for the static, accumulative, and sliding windows methods, respectively. During this month, the monthly CV(RMSE) values are 9.3% (Static), 8.5% (Accumulative), and 9.0% (Sliding windows).

The result of September, as shown in Fig. 8(b), illustrates that the model performance of the accumulative and sliding windows training is similar to the results of August, with *R*-square value of 0.914 and 0.919, respectively. However, the static training with the trained network based on July data was not as good as in August (*R*-square = 0.798). The forecast accuracy falls for both weekend and weekdays, but more substantially for weekend. The accuracy falls especially for the last week of the month as described in Fig. 9(b). Accordingly, monthly CV(RMSE) of the static training (18.8%) is much higher than those of the accumulative (10.1%), and the sliding windows (9.9%) training for the month.

3.2.2. *Daily electricity usage pattern*

Table 6 presents the statistical summary of CV(RMSE) and its standard deviation (σ) on daily level. All three run types of training set have a similar performance in August. The average value of the daily CV(RMSE) and standard deviation in weekdays are around 8% and 2.5%, respectively. In September, CV(RMSE) and standard deviation of the adaptive training, sliding window and accumulative, are 9% and 2% in weekdays but the average daily forecast error (11.6%) and variance (5.03) by the static training is higher than for the other training types. As presented in Fig. 9(b), the static training often overestimated the electricity consumption in weekdays. It may indicate that the energy consumption pattern of the site has changed by the outdoor weather condition.

3.2.3. *Daily peak usage prediction*

In addition to the daily electricity usage profile, the daily peak demand forecasting accuracy is one of the important factors in model evaluation. Table 7 summarizes the daily peak forecasting performance on weekdays and weekends for the two months. All training types predicted the daily peak demand with average APE of approximately 3% and 4.5% for weekdays of August and September, respectively. As similar to the daily prediction performance, the adaptive training method provides more accurate results with smaller variance than the static training in September.

accumulative + sliding window

As shown in Fig. 9(a), the annual peak usage of the building was 734.60 kWh at 13:00 of August 20th. The model forecasted the daily peak electricity usage as 683.44 kWh at 12:30 by the sliding windows training (APE = 6.96%) and 678.62 kWh at 14:15 by the accumulative training (APE = 7.62%). Although the forecast error is higher than during the other weekdays in this week, it may suggest that the day has an unusual electricity usage pattern comparing to the actual and model forecast results of the adjacent weekdays.

With the regard of the uncertainty of actual electricity consumption in the short-term intervals, the day-ahead peak demand forecasting performance by the developed model is reasonable enough to make a management plan for the daily electricity consumption and peak electricity demand in the building.

4. Conclusion

Many studies have been done on model development for short-term electricity load forecast for electricity supply and demand, and have reported successful results in practical tests over hourly resolution. However, the sub-hourly electricity usage forecast is still a challenging problem due to the complexity of usage pattern and the highly noisy input data.

In this study, we demonstrated a new approach using a feature extraction and an artificial neural network to make a-day-ahead forecast of the electricity usage profile for a commercial building complex in a high temporal resolution of 15 minutes. By using feature extraction process, important variables to forecast the actual electricity consumption were selected. The conventional predictors such as time indicator, weather variables and HVAC operational schedule still turn out to be very important predictors for building's electricity usage in our selection. With the specified predictors and electricity data, several algorithms were compared, and the artificial neural network model is found to be the most suitable for this problem. The artificial neural network with a regularization algorithm for training was adopted and the design parameters are investigated to select the best structure of the prediction model.

The implementation results of the developed model for two months illustrate that the daily error with 15-minute resolution forecasting is stable around averaged CV(RMSE) of 10% for weekdays and the model predicts the daily peak demand of weekdays within averaged APE of 5% for the period. It implies that the model can provide a day-ahead electricity usage profile with sub-hourly intervals and daily peak electricity consumption with a reasonable accuracy.

A good predictive model of energy consumption in buildings is useful not only for accurately forecasting future energy consumption but also in developing a good model predictive control (MPC) method that can reduce energy costs in buildings. For example, the model would be essential to control the daily load profile of a building by using on-site energy sources like an electricity storage system (ESS) with renewable energy, and to dispatch demand response (DR) for saving utility cost. It is also applicable to control the priorities of the end-used energy consumption with a smart switch panel or sub-meters during on-peak period.

As a future research, real-time electricity forecasting model with a smart meter which can be adaptive to weather and building operation changes for both the total meter and sub-meter level, and an anomaly detection model in short-term electricity usage pattern will be investigated.

References

[1] K. Herter, P. McAuliffe, A. Rosenfeld, An exploratory analysis of California residential customer response to critical peak pricing of electricity, Energy 32 (1) (2007) 25–34.

[2] P. Mandal, et al., A new recursive neural network algorithm to forecast electricity price for PJM day-ahead market, Int. J. Energy Res. 34 (6) (2010) 507–522.

[3] M.L. Goldberg, F.F. Margaret, Refraction of PRISM results into components of saved energy, Energy Build. 9 (1) (1986) 169–180.

[4] D. Ruch, D.E. Claridge, A four-parameter change-point model for predicting energy consumption in commercial buildings, J. Sol. Energy Eng. 114 (2) (1992) 77–83.

[5] A. Kimbara, et al., On-line prediction for load profile of an air-conditioning system, ASHRAE Trans. 101 (1995) 198–207.

[6] G.A.N. Mbamalu, M.E. El-Hawary, Load forecasting via suboptimal seasonal autoregressive models and iteratively reweighted least squares estimation, Power Syst. IEEE Trans. 8 (1) (1993) 343–348.

[7] B. Wang, et al., A new ARMAX model based on evolutionary algorithm and particle swarm optimization for short-term load forecasting, Electr. Power Syst. Res. 78 (10) (2008) 1679–1685.

[8] D.G. Infield, D.C. Hill, Optimal smoothing for trend removal in short term electricity demand forecasting, Power Syst., IEEE Trans. 13 (3) (1998) 1115–1120.

[9] A. Dhar, T.A. Reddy, D.E. Claridge, A Fourier series model to predict hourly heating and cooling energy use in commercial buildings with outdoor temperature as the only weather variable, J. Sol. Energy Eng.—Trans. ASME 121 (1) (1999) 47–53.

[10] P. de Wilde, et al., Building simulation approaches for the training of automated data analysis tools in building energy management, Adv. Eng. Inform. 27 (4) (2013) 457–465.

[11] L. Wu, et al., Improving efficiency and reliability of building systems using machine learning and automated online evaluation, in: Proceedings of the Systems, Applications and Technology Conference (LISAT), IEEE, Farmingdale, NY, USA, 2012.

[12] K. Li, H. Su, J. Chu, Forecasting building energy consumption using neural networks and hybrid neuro-fuzzy system: a comparative study, Energy Build. 43 (10) (2011) 2893–2899.

[13] H.S. Hippert, C.E. Pedreira, R.C. Souza, Neural networks for short-term load forecasting: a review and evaluation, Power Syst. IEEE Trans. 16 (1) (2001) 44–55.

[14] G. Escrivá-Escrivá, et al., New artificial neural network prediction method for electrical consumption forecasting based on building end-uses, Energy Build. 43 (11) (2011) 3112–3119.

[15] P.A. González, J.M. Zamarreno, Prediction of hourly energy consumption in buildings based on a feedback artificial neural network, Energy Build. 37 (6) (2005) 595–601.

[16] M. Aydinalp-Koksal, V.I. Ugursal, Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector, Appl. Energy 85 (4) (2008) 271–296.

[17] A.H. Neto, F.A.S. Fiorelli, Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption, Energy Build. 40 (12) (2008) 2169–2176.

[18] Korea Electric Power Corporation (KEPCO), Electric rates table. ⟨http://cyber.kepco.co.kr/kepco/EN/F/htmlView/⟩, (accessed 15.11.14).

[19] A. Chabaud, J. Eynard, S. Grieu, Optimal design of energy production and storage systems in buildings, in: Proceedings of the 4th Inverse Problems, Design and Optimization Symposium (IPDO-2013), 2013, Albi, France, Thematic-5.

[20] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.

[21] C. Strobl, et al., Conditional variable importance for random forests, BMC Bioinform. 9 (1) (2008) 307.

[22] D.J.C. MacKay, Introduction to Gaussian processes, NATO ASI Ser. F Comput. Syst. Sci. 168 (1998) 133–166.

[23] R.-E. Fan, et al., LIBLINEAR: a library for large linear classification, J. Mach. Learn. Res. 9 (2008) 1871–1874.

[24] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. (TIST) 2 (3) (2011) 27.

[25] J.G. Cleary, L.E. Trigg, K^*: an instance-based learner using an entropic distance measure, ICML 5 (1995) 108–114.

[26] S.M. Omohundro, Five Balltree Construction Algorithms, International Computer Science Institute, Berkeley, 1989.

[27] H.-X. Zhao, F. Magoulès, A review on the prediction of building energy consumption, Renew. Sustain. Energy Rev. 16 (6) (2012) 3586–3592.

[28] M. Daneshdoost, et al., Neural network with fuzzy set-based classification for short-term load forecasting, Power Syst., IEEE Trans. 13 (4) (1998) 1386–1391.

[29] S.V. Kartalopoulos, Understanding Neural Networks and Fuzzy Logic: Basic Concepts and Applications, Wiley-IEEE Press, New York, NY, 1997.

[30] S.S.A.K. Javeed Nizami, A.Z. Al-Garni, Forecasting electric energy consumption using neural networks, Energy Policy 23 (12) (1995) 1097–1104.

[31] R. Hecht-Nielsen, Theory of the backpropagation neural network, in: Proceedings of the International Joint Conference on Neural Networks, 1989. IJCNN, IEEE, Washington D.C., 1989, pp. 593–605.

[32] J. Sietsma, R.J.F. Dow, Creating artificial neural networks that generalize, Neural Netw. 4 (1) (1991) 67–79.

[33] D.J.C. MacKay, Bayesian interpolation, Neural Comput. 4 (3) (1992) 415–447.

[34] F. Dan Foresee, M.T. Hagan, Gauss–Newton approximation to Bayesian learning, in: Proceedings of the International Conference on Neural Networks, IEEE, Houston, TX, USA, 1997, pp. 1930–1935.

[35] ASHRAE Guideline 14-2002, Measurement of Energy and Demand Savings, American Society of Heating, Ventilating, and Air Conditioning Engineers, Atlanta, Georgia, 2002.