Renewable and Sustainable Energy Reviews 33 (2014) 400–411

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

Real-time detection of anomalous power consumption

Jui-Sheng Chou*, Abdi Suryadinata Telaga

Department of Civil and Construction Engineering, National Taiwan University of Science and Technology, 43, Sec. 4, Keelung Rd., Taipei, 106, Taiwan

ARTICLE INFO

Article history:
Received 19 October 2013
Received in revised form
22 January 2014
Accepted 31 January 2014
Available online 5 March 2014

Keywords:
Power consumption
Big data analytics
Anomaly detection
Pattern recognition
Real time detection
Time series prediction

ABSTRACT

Effective feedback can reduce building power consumption and carbon emissions. Therefore, providing information to building managers and tenants is the first step in identifying ways to reduce power consumption. Since reducing anomalous consumption can have a large impact, this study proposes a novel approach to using large sets of data for a building space to identify anomalous power consumption. This method identifies anomalies in two stages: consumption prediction and anomaly detection. Daily real-time consumption is predicted by using a hybrid neural net ARIMA (auto-regressive integrated moving average) model of daily consumption. Anomalies are then identified by differences between real and predicted consumption by applying the two-sigma rule. The experimental results for a 17-week study of electricity consumption in a building office space confirm that the method can detect anomalous values in real time. Another contribution of the study is the development of a formalized methodology for detecting anomalous patterns in large data sets for real-time of building office space energy consumption. Moreover, the prediction component can be used to plan electricity usage while the anomaly detection component can be used to understand the energy consumption behaviors of tenants.

Contents

1. Introduction

The OECD has predicted that global energy consumption will increase 53% from 505 quadrillion Btu in 2008 to 770 quadrillion

* Corresponding author. Tel.: +886 2 2737 6321; fax: +886 2 2737 6606.
E-mail addresses: jschou@mail.ntust.edu.tw, rayson.chou@gmail.com (J.-S. Chou), D9915802@mail.ntust.edu.tw (A.S. Telaga).

http://dx.doi.org/10.1016/j.rser.2014.01.088
1364-0321

Btu in 2035 [1]. Compared to the transportation and industrial sectors, the building sector consumes more energy (approximately 40% of global energy use) and generates 30% more CO_2 [2]. Therefore, a critical step in lowering carbon is reducing energy consumption in buildings. Given the particularly high dependence of Taiwan on imported fossil fuels, developing an economical, low-carbon, and highly efficient green energy system is imperative [3].

Studies performed in the United Kingdom [4] and in the United States [5] show that the growing use of energy-consuming equipment beat efficiency gains in green building technology. The increased energy consumption is mainly due to equipment for maintaining comfort in residential and commercial buildings, such as air conditioners, heaters and other modern appliances [5]. However, energy consumption in commercial buildings is more complex than that in residential buildings [6].

While residential buildings mainly provide a sanctuary for people, commercial buildings have widely varying purposes. Nevertheless, commercial buildings are mainly designed for business activities and expected to generate income for building owners and their tenants. Therefore, energy-saving strategies are needed to reduce operating costs on both sides. Specifically, electricity consumption by commercial buildings is the highest during 9:00–17:00, which is usually the highest price in time-price based schemes. Moreover, Popescu et al. also found that energy-efficient buildings benefit owners by increasing the property values [7].

The building manager is responsible for managing building performance, and one of the main building performance measures is electricity consumption [8]. Additionally, in countries that have recently increased requirements for green building certification, the building manager must minimize energy consumption. Thus, to reduce electricity consumption and CO_2 emissions, building managers must understand energy consumption from the tenant perspective. Therefore, building electricity consumption is both a social problem and a technical problem [5].

Analyzing electricity consumption from the tenant perspective requires very detailed data. To acquire such data, researchers have proposed using sensors for detecting movement [9], thermostats [10], cameras [11] or combinations of sensors that detect light, CO_2, temperature, *etc.* [12]. In practice, however, implementing this approach in commercial buildings is highly impractical. For privacy reasons, some tenants may reject the idea of sensors installed in their offices. Moreover, wiring costs are 45% and 75% of total installation cost for new buildings and retrofitted buildings, respectively [13]. Analyzing data streams from numerous real-time sensors can also be a heavy burden on building energy managers [14].

Smart meter use can reduce the required number of sensors and eventually reduces data stream volume. A smart meter is an electrical meter that records electrical energy consumption at intervals of an hour or less and sends the information back to the utility center for monitoring and billing purposes [15]. Therefore, smart meters provide more information compared to conventional meters, which only provide data for billing purposes [15]. Moreover, a smart meter management system is needed for an efficient smart grid system [16]. Finally, customers benefit from improved reliability of utility networks [17] and improved responsiveness of services, which eventually improve and sustain the customer relationship [18].

Additionally, smart meter data can be utilized to provide power quality (PQ) information to customers and utility companies. As the quality is susceptible to any disturbance in power transmission network, PQ is an important measure for customers [19]. Particularly, for the buildings that use electricity from different companies, the companies could develop PQ index [20] to provide fair information to customer and use the index to monitor any disturbance in power quality production. Consequently, for a fairer energy price, the price can be adjusted in terms of power quality [21].

Smart meters can provide detailed data for the electricity consumption of a customer in real-time or near real-time. Further, in-home implementations combining smart meter and enabling technologies such as in-home display have shown that smart meters can reduce energy consumption [22]. Studies show that the highest reductions occur when people are already at home at 17:00 (5 pm), which indicates that, with the right feedback, people can reduce their electricity consumption [23]. For example, a study by the Energy Saving Trust in 2009 showed that feedbacks that had the largest contribution to smart meter use were those that helped to reduce electricity use [24].

Anomalous electricity consumption data can help tenants identify extraordinary consumption patterns [25]. In commercial buildings, anomalous consumption may also result from activities such as over-lighting [6], inefficient equipment or overtime work. Therefore, anomalous feedback data can be further used to warn tenants to minimize electricity use and to help them identify ineffective equipment or over-lighting in office spaces. However, extracting meaningful information from smart meter data is a formidable task [26].

Although several anomaly detection methods have been researched, the primary objective has been detecting anomalous consumption in automated building systems such as heating, ventilation, and air conditioning (HVAC) systems [14,26–28]. However, the building must also support random use of office equipment, lightings, heating, and air condition. Since HVAC systems consume almost 50% of energy in a building [8], reduction of energy use by non-HVAC systems can potentially reduce total consumption by 50%. Office equipment consumes 15% of the total energy consumed by an office. By 2020, this figure is expected to increase twofold [29]. Therefore, potential savings in electricity consumption by office spaces are also large.

Because no studies have considered anomaly detection in office spaces, this study performed an experiment to develop a real-time system for detecting anomalous electricity consumption in an office space from the perspective of occupant activities. All experimental data were retrieved from smart meters used to monitor electricity consumption in an office space in a university building. The main objective was to develop an anomaly detection methodology that is applicable in large data stream of smart meter data and real time environment. Therefore, the research results have potential applications in a web-based early warning system. Notably, the results application is not only limited to building energy consumption domain, but also applicable to any anomaly detection system that use time based sensor data as input. Furthermore, the potential application includes gas flow detection, water flow detection, and comfort level detection. The main contributions of this research are the following:

- A formalized methodology for detecting anomalous patterns in large real-time datasets for building office space energy consumption.
- The method is performed in two stages. The prediction stage helps building managers plan their electricity demand while the anomaly detection stage helps building managers identify tenant consumption patterns. In the case of a building that generates its own electricity and has abnormally low energy consumption, the building manager can connect to a smart grid and sell the excess electricity to gain profit.
- Anomaly detection benefits tenants by helping them understand how their office activities consume energy. They can then modify their anomalous activities, analyze energy consumption costs and benefits, and eventually reduce their wasteful activities.

The remainder of this paper is organized as follows. Section 2 briefly introduces the study context by reviewing related

literature, including studies of anomaly detection and some well-known demand prediction techniques. Section 3 then describes the research methodology, and the evaluation methods of proposed models. Section 4 further explains the experiment performed in this research. Section 5 then presents the experimental results, and Section 6 analyzes the results and compares model performance. Finally, Section 7 summarizes the findings and conclusions.

2. Review of literature on research problem

Analyzing building electricity consumption data is important because failure to do so can jeopardize building management by causing excessive energy use and potential increases in carbon taxes [30]. Although building energy consumption has been studied intensively, further studies are needed to optimize artificial intelligence (AI) for prediction models and to integrate the models in a Building Energy Management System (BEMS) [31]. In the future, most smart meter systems will be AI-based to enable independent management of power consumption [32].

Electricity consumption in an office space is usually estimated for a typical working day and working week. However, electricity demand signatures differ according to occupant behavior and during different periods, *e.g.*, lunch time, regular workdays, and seasonal holidays [26]. Therefore, an anomaly detection system for an office space must detect anomalous consumption based on these signatures and pattern changes during seasonal holidays.

Here, an anomalous condition is defined as an abnormal power consumption usage. An anomalous state is defined as a deviation from the normal electricity consumption of the tenant. Therefore, the proposed anomaly detection model has two stages: power consumption prediction method and anomaly detection. Because electricity consumption data is a time series domain, the objective was to develop a suitable time series anomaly detection method. By defining an anomalous state as two standard deviations (SDs) above or below the predicted power consumption, an anomaly can be easily computed and flagged. The definition is based on empirical rule of normal distribution that 95% of values lie within roughly two SDs of the mean. Therefore, another 5% value outside two SDs can be considered as an anomaly as depicted in Fig. 1. Similar definition was used in Brown et al. that they defined anomalous activity as consumption exceeding an SD of three in the prediction results [33].

Studies of anomaly detection methods in the energy consumption domain include Yi et al., who compared regression, entropy, and clustering methods. The regression methods obtained the best detection results [34]. Brown et al. further used *K*-nearest

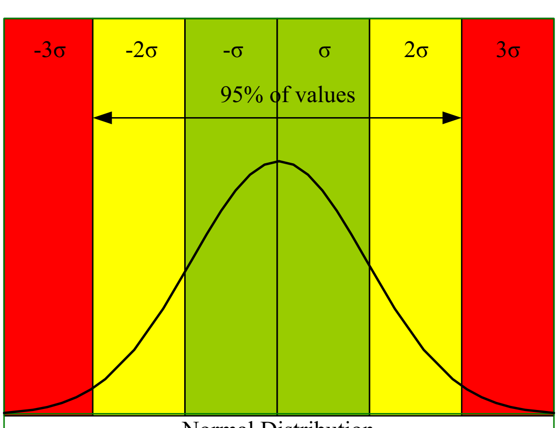

Fig. 1. 2-Sigma rule of normal distribution.

neighborhood (*K*-NN) in fast kernel regression to predict electricity consumption [33]. However, both methods require huge datasets, and the resulting models are static. Since they are prone to pattern change, they are not the preferred models for on-line prediction [35].

When using large data sets to solve problems and identify pattern changes, researchers have combined sliding window datasets with other techniques such as adaptive artificial neural network (ANN) [36]. Even when the dataset is small, adaptive ANN outperforms static ANN when using a real measurement dataset and performs comparably to static ANN when using a static dataset. Wrinch et al. applied Fourier transform with a sliding window and found that a weekly window provided faster fault detection compared to a monthly window [26]. Li et al. performed a time-series auto-regressive integrated moving average (ARIMA) analysis of a dataset for real-world daily shifts in water consumption to detect meter stilting [27].

However, anomaly detection by Fourier transformation has a high false positive rate due to the assumption of constant periodicity of data [35]. The ARIMA method does not obtain a good model if the duration of anomaly data is long [37]. Optimizing the hidden layer and time lag is also problematic when applying ANN in time series domain [38]. Therefore, a suitable method is needed to address these limitations.

Recent studies suggest that a combination of several individual models can compensate for deficiencies of a model. Theoretically, hybridization of unrelated models also reduces generalization variance or error [39]. Since forecasting problems in real-world time series data usually contain both linear and nonlinear components [38], the hybrid model usually combines linear and nonlinear models. In the energy demand prediction domain, ARIMA and ANN are the most common forecasting methods [40,41]. Individual ARIMA models have been widely used for linear time series forecasting [39,41], and ANN has been successfully used to solve nonlinear problems [42].

Zhang confirmed that a hybrid model combining ARIMA and ANN is better than either of the models used independently [38]. Hybrid models have been successfully applied in economic time series forecasting [43], fuel wood price [44], wind speed forecasting [45], and electricity price [46]. Although there is still no consensus of best approach in combining ANN and ARIMA [42], hybridization of ARIMA and ANN or other soft computing has been proved to improve energy demand forecasting [40]. Until now, however, these methods have only been applied to small data streams. Their effectiveness in large data streams for electricity consumption is still unknown and deserved further investigation. In this sense, this study proposes a hybrid ANN and ARIMA model with a sliding window for analyzing large data streams in the power consumption domain.

3. Methodology

The experimental methodology was based on a case study. First, a K-means algorithm, a data mining cluster analysis [47], was used to categorize daily consumption patterns in a week. The analysis of *K*-means algorithm results suggested that electricity consumption patterns differ each day except on weekends. A correlation analysis was then performed to determine whether training data are better suited for a weekly or daily analysis. The results were consistent with our observation that training data were best presented by weekly window since the office occupants were students who attended class on a weekly basis.

After arrangement of the consumption data, the data were entered into an ANN–ARIMA hybrid model using NNAR (Neural Network Auto Regressive), a function in forecast library of R

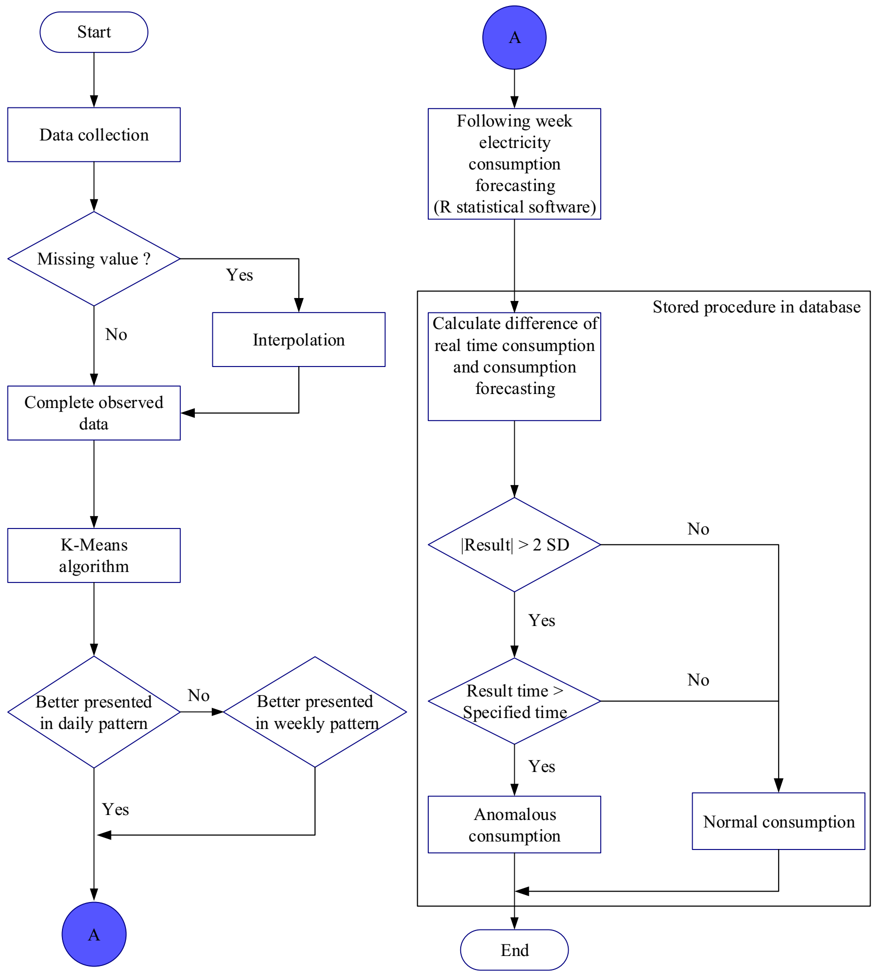

Fig. 2. Anomaly detection flowchart.

statistical software. The R statistical software is a comprehensive software applicable to handle data manipulation, calculation and graphical visualization [48]. Sliding windows for 4-week and 8-week datasets were used for training sets in the NNAR model to form dynamic models. The resulted models then were stored in database as a predicted consumption for next week electricity consumption. Fig. 2 shows the flow chart of anomaly detection process.

The anomalous states computed by calculating differences between actual consumption and predicted consumption were then flagged when time duration exceeded anomalous time allowed. The prediction results were then evaluated using mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean square error (RMSE) for accuracy evaluation purpose. Therefore, three evaluation methods were utilized to represent deviation between actual electricity consumption and predicted consumption.

3.1. *K*-means algorithm

The *K*-means algorithm is one of the simplest unsupervised learning algorithms for solving clustering problems [47]. This algorithm is also one of the most popular and widely used partition clustering methods. The procedure is simple; firstly the algorithm classifies a set through a definite number of clusters. After finding the cluster centers, the algorithm positions the centers as remotely as possible. Finally, the algorithm affiliates all data points from the dataset with the closest centers. The first iteration is complete when no data points remain. The iterations continue until all centers are determined.

The *K*-means algorithm searchers for the cluster centers $(c_1, c_2, \ldots, c_k)$ such that the sum of the squared distances (called distortion) of each data point (x_i) to its nearest cluster center (c_k) is minimized (Eq. (1)), where d is the distance function of the Euclidean distance [47]:

$$D = \sum_{i=1}^{n} [\min d(x_i c_k)^2] \quad k = 1, 2, \ldots, \mathrm{K} \tag{1}$$

- Step 1: Assign cluster number K and initialize the centroids of each cluster, $(c_1^{(0)}, c_2^{(0)}, \ldots, c_k^{(0)})$; each cluster center is an m-dimensional vector 1, 2, …, $c_i^0 = \{c_{i1}^{(0)}, c_{i2}^{(0)}, \ldots, c_{im}^{(0)}\}$.
- Step 2: Start the iterative procedure. Set iteration count t to 1.
- Step 3: Calculate the distance measure $d_{ki}^{(t-1)}$ between the kth cluster center and the ith data set (data point in m space). Here, the distance is defined as the Euclidean distance given by the following equation:

$$d_{ki}^{(t-1)} = ||x_i - c_k^{t-1}|| = \sqrt{\sum_{j=1}^{m} (x_{ij} - c_{kj}^{t-1})^2} \tag{2}$$

- Step 4: Assign each data object x_i to its nearest cluster center c_k,

- Step 5: Update each cluster center $c_k^{(t)}$ as the mean of all x_i that are closest according to the following equation:

$$c_k^t = \frac{\sum x_{j \in k} x_i}{n_k} \tag{3}$$

where n_k is number of data items in the kth cluster.
- Step 6: Use Eq. (3) to calculate distortion D, which depicts the sum of all intra cluster distances. A low D is preferable.
- Step 7: If the value of D has converged, return the final cluster centers $(c_1^{(0)}, c_2^{(0)}, \ldots, c_k^{(0)})$. Otherwise, set $t = t+1$, and return to step 3.

3.2. Artificial neural networks

Artificial neural networks (ANNs) are information-processing units that function similarly to neurons in the human brain except that a neural network consists of artificial neurons [49]. The structure of an ANN contains many such neurons connected systematically. The feed-forward neural networks used here are also known as multilayer perceptrons (MP). The quantifiable data used for problem solving are fed into the input layer and then processed by the self-updating and self-learning model in the hidden layer. The resulting solution is then sent to the output layer. The mathematical model for MP is:

$$z_j = \varphi\left(b_j + \sum_{i=1}^{n} w_{i,j} x_i\right) \tag{4}$$

where

z is the forecast value;
φ is the activation function;
w is the vector of weights;
b is the bias; and
n is the number of neurons.

In the hidden layer, the activation function is often selected as the logistic sigmoid function.

$$s(z) = \frac{1}{1+e^{-z}} \tag{5}$$

3.3. Auto-regressive integrated moving average

The ARIMA (auto-regressive integrated moving average) is a time series forecasting model [38,50] that uses time-series stationary data. Therefore, the data must be made stationary by differencing d times. Auto-regression is a forecasting equation term that explains lags of the time series. Furthermore, lag forecast errors are explained by a moving average term in the forecasting equation. Lastly, integration explains the addition of those two series. Eq. (6) below depicts the non-seasonal ARIMA model as "ARIMA(p,d,q)":

$$r_t = \varphi_o + \sum_{i=1}^{p} \varphi r_{t-i} + a_t - \sum_{i-1}^{q} \theta_i a_{t-i} \tag{6}$$

where

p is the number of autoregressive terms;
d is the number of non-seasonal differences;
q is the number of lagged forecast errors in the prediction equation;
φ is the autoregressive constant;
θ is the moving average constant;
t is the number of time series data items;
r is the forecast value; and
a is the moving average value.

3.4. Neural network auto regressive

The NNAR forecasting model is a hybrid ANN–ARIMA model in which the neural network uses lagged values of the time series as inputs. Since the model uses one hidden layer feed-forward network in which the inputs are lags 1 to p, the model uses p last observations [50]. In the ANN modeling stage, the model starts with a random weight and then applies the adjusted weight when performing the forecasting computation. The network is trained for one-step forecasting and uses a recursive calculation for multi-step forecasting. Therefore, the mathematical formula for NNAR becomes:

$$y_t = w_0 + \sum_{j=1}^{k} b_j \cdot g\left(w_{0j} + \sum_{i=1}^{p} w_{ij} \cdot y_{t-1}\right) + \varepsilon_t \tag{7}$$

where

$j = 1, 2, \ldots, k$ is the number of neurons;
$i = 1, 2, \ldots, p$ is the lag;
w_o is a constant;
w_j is the connection weight where $j = 1, 2, \ldots, k$;
g is the activation function in Eq. (5);
w_{oj} is a constant at neuron j;
w_{ij} is the connection weight where $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, k$; and
ε_t is an error term.

In practice, the ANN performs a nonlinear model of the last p observations with k neurons where

$$y_t = f(y_{t-1}, \ldots, y_{t-p}, w) + \varepsilon_t \tag{8}$$

y_t is the predicting value;
p is the lag number;
w is the weight for all parameter; and
ε_t is the error term.

The resulting NNAR(p,k) model resembles a model that uses p lagged inputs and k nodes in hidden layer. For example NNAR(8,9) indicates a neural network that uses the previous eight values $(y_{t-1}, y_{t-2}, \ldots, y_{t-8})$ as inputs for the neural network and uses nine neurons in hidden layer. A NNAR(p,0) model resembles an ARIMA (p,0,0) model but does not have the parameter restrictions used to ensure stationarity.

3.5. Evaluation method

To evaluate the accuracy of electricity demand, several criterions are used. There are different alternative methods for this purpose: MAPE, MAE and RMSE are defined as follows:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|\widehat{Y}_i - Y_i|}{Y_i} \times 100 \tag{9}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} [\widehat{Y}_i - Y_i]^2} \tag{10}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\widehat{Y} - Y_i| \tag{11}$$

where

$\widehat{Y}$, predicted value;
Y_i, observed values; and
n, sample size.

The MAPE is useful for evaluating the performance of predictive models because of its relative values. Since the MAPE is unaffected by the size or the unit of the observed and predicted values, it indicates their relative difference. A MAPE value lower than 10% indicates high forecasting accuracy. Values of 10–20%, 20–50%, and over 50% indicate good, reasonable, and inaccurate forecasting accuracy, respectively.

The RMSE formula calculates the square error of the prediction compared to observed values and calculates the square root of the summation value. Since the errors are squared before calculating the average, values with large errors are weighted most heavily. Therefore, this measure is effectively reveals unacceptably large differences. In contrast, MAE by definition calculates the average magnitude of errors between predicted and observed values without considering the direction of errors. That is, since all individual differences are weighted equally, MAE can measure continuous variables.

4. Experimental design

An experiment was designed to measure actual power consumption data. Although the experiment was performed in a laboratory located in a university office building, the experiment was performed in a real-life setting. The office occupants performed daily activities without being given suggestions for improving energy efficiency. Therefore, the power consumption data reflected the real-life activities of office occupants. Variations in power consumption resulted from people activities inside the office. The commercially available smart meters used for the experiment were connected to the server operated by the manufacturer.

Fig. 3 shows the steps of the anomaly detection process. Power consumption data in this experiment were obtained only from smart meter 1, which was used to measure electricity consumption in an office space without centralized HVAC. The data were then transferred to the application server for further processing. The application-server then stored the data and automatically uploaded the data to a database. The database preparation stage restored missing values by interpolating two adjacent data items and then formatting them for further modeling.

The forecast package in R language was used to predict electricity consumption. Finally, after saving the prediction results in the database, the anomaly detection procedure was performed. The database then received the real-time power consumption data from the smart meter and calculated the difference by two-sigma rule to

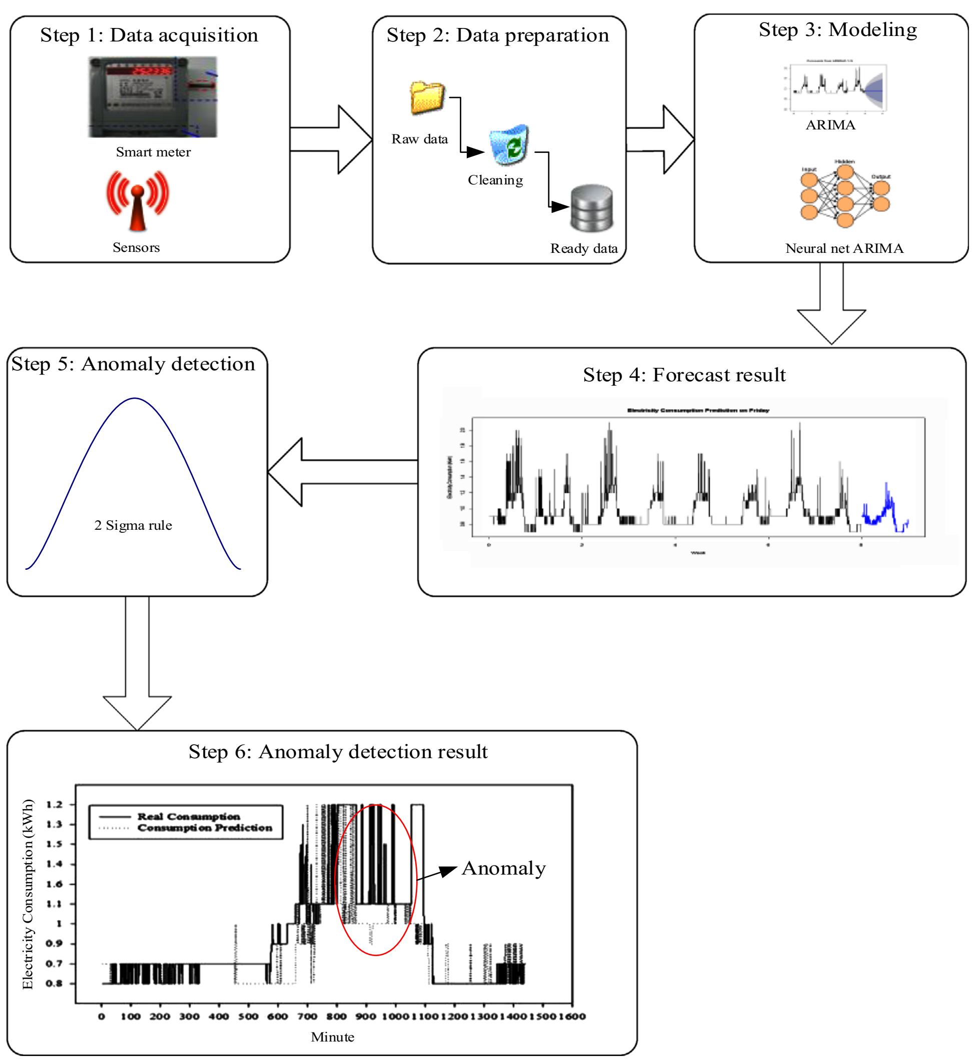

Fig. 3. Stages of anomaly detection process.

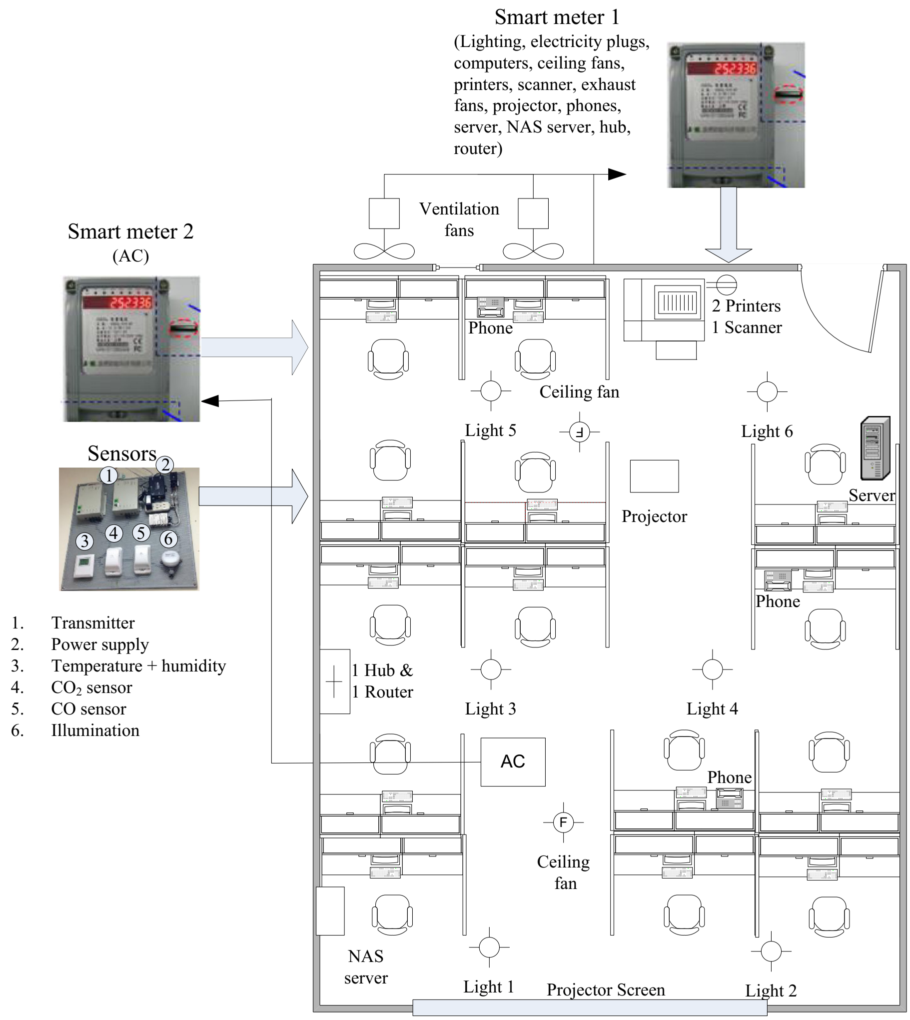

Fig. 4. Office space layout.

detect anomalous electricity consumption behavior. The current and predicted energy consumption data could then be accessed by users.

Fig. 4 shows how the smart meters used in the experiment were installed in an office space. Smart meter 1 recorded the electricity consumed by the computers, printers, scanners, ventilation fans, projectors, phones, servers, NAS servers, electricity plugs, ceiling fans, hub, and lamps inside the office space. In the office space used for the experiment, 28 out of 30 main electricity sockets were in use, and 50 out of 55 additional electricity sockets were in use. The room also had 20 LAN sockets. Thirteen people used the office from 9:00 to 18:00. Since the electricity consumption data were obtained for the real-life activities of office occupants during each minute, data were collected 1440 times daily.

5. Results

5.1. Data set

The experiment was performed using office power consumption data collected from a smart meter and stored in a database. The dataset comprised power consumption data samples obtained once per minute during the 17-week period from October 22, 2012 to October 7, 2013. The forecasting performance of the models was assessed by dividing each dataset into training and testing sets. The datasets were also rolled every week as new data arrived to obtain a sliding window. Therefore, the 8-week and 4-week sliding windows were the training dataset, respectively. The testing dataset was one week after training.

The ARIMA calculations for both hybrid models were performed in a similar environment, *i.e.*, the R library. The hybrid model was performed using NNAR function and ARIMA predictions were implemented using auto ARIMA function. The same environment was used for both models because the research goal was to develop an applicable real-time detection system. Therefore, a suitable method identified in this research could later be used in a complete system. Predictions of electricity consumption during the following week were based on 8 or 4 weeks of rolling training data.

For example, data for weeks 1–8 were used to predict electricity consumption for week 9 while data for weeks 2–9 were used to predict consumption for week 10. The same procedure was used for 4 weeks of rolling data, *i.e.*, data for weeks 5–8 were used to predict consumption for 9, and so on. The data used to build the forecasting model for each of the considered weeks included electricity consumption for each minute of the 8 weeks and 4 weeks previous to the first day of the week whose consumption are to be predicted. Standard ARIMA predictions based on 8-week and 4-week training data were also used for comparison with the NNAR method.

5.2. Training results

Table 1 shows both the NNAR and ARIMA models accuracy based on the training data set. Particularly, the NNAR's MAPE

Table 1
Comparison of accuracy using training dataset.

Method	Descriptive statistics	Monday			Tuesday			Wednesday			Thursday			Friday			Overall		
		MAPE (%)	RMSE (kWh)	MAE (kWh)	MAPE (%)	RMSE (kWh)	MAE (kWh)	MAPE (%)	RMSE (kWh)	MAE (kWh)	MAPE (%)	RMSE (kWh)	MAE (kWh)	MAPE (%)	RMSE (kWh)	MAE (kWh)	MAPE (%)	RMSE (kWh)	MAE (kWh)
NNAR 8-week data stream	Mean	2.793	0.063	0.025	3.078	0.119	0.035	2.729	0.121	0.029	2.488	0.061	0.026	2.333	0.054	0.023	2.684	0.084	0.028
	Min.	2.713	0.052	0.020	2.227	0.070	0.026	2.360	0.067	0.027	2.229	0.053	0.021	1.852	0.046	0.020	1.852	0.046	0.020
	Max.	2.928	0.070	0.030	3.597	0.178	0.040	3.177	0.286	0.034	2.738	0.064	0.028	2.895	0.063	0.027	3.597	0.286	0.040
NNAR 4-week data stream	Mean	2.601	0.070	0.022	2.852	0.136	0.032	2.572	0.131	0.025	2.368	0.059	0.024	2.055	0.044	0.018	2.490	0.088	0.024
	Min.	2.087	0.040	0.010	2.054	0.062	0.021	1.527	0.040	0.010	1.786	0.049	0.020	1.252	0.026	0.006	1.252	0.026	0.006
	Max.	3.125	0.168	0.031	3.752	0.245	0.045	3.554	0.288	0.038	2.924	0.068	0.029	2.742	0.052	0.025	3.752	0.288	0.045
ARIMA 8-week data stream	Mean	2.708	0.077	0.023	2.927	0.135	0.033	2.653	0.155	0.027	2.422	0.059	0.025	2.188	0.047	0.020	2.580	0.095	0.026
	Min.	2.087	0.040	0.010	2.054	0.062	0.021	1.527	0.040	0.010	1.786	0.049	0.020	1.252	0.026	0.006	1.252	0.026	0.006
	Max.	3.125	0.168	0.031	3.752	0.245	0.045	3.554	0.288	0.038	2.924	0.068	0.029	2.895	0.063	0.027	3.752	0.288	0.045
ARIMA 4-week data stream	Mean	2.685	0.083	0.022	2.921	0.139	0.033	2.628	0.157	0.026	2.407	0.059	0.025	2.163	0.047	0.019	2.561	0.097	0.025
	Min.	2.087	0.040	0.010	2.054	0.062	0.021	1.527	0.040	0.010	1.786	0.049	0.020	1.252	0.026	0.006	1.252	0.026	0.006
	Max.	3.125	0.168	0.031	3.752	0.245	0.045	3.554	0.288	0.038	2.924	0.068	0.029	2.895	0.063	0.027	3.752	0.288	0.045

values for 8 weeks of training data have a minimum value of 1.852%, a maximum value of 3.597% and an average value of 2.684%. The MAPE results for the 8-week NNAR training model revealed a satisfactory fit to the data. The 4-week training data had a minimum value of 1.252%, a maximum value of 3.752% and an average value of 2.490%. The results for the 4-week NNAR training model also showed a good data fit.

Moreover, RMSE evaluation showed that the 8-week model had a minimum value of 0.046 kWh, a maximum value of 0.286 kWh, and an average value of 0.083 kWh. Similarly, the 4-week model had a minimum value of 0.026 kWh, a maximum value of 0.288 kWh, and an average value of 0.088 kWh. The experimental results showed small RMSE values for both models. As for MAE, the 8-week model had a minimum value of 0.02 kWh, a maximum value of 0.04 kWh, and an average value of 0.0276 kWh. The 4-week model had a minimum value of 0.006 kWh, a maximum value of 0.045 kWh, and an average value of 0.024 kWh. Likewise, the results indicated a very good data fit in the 4-week model.

Further comparisons between the NNAR results and the ARIMA results showed that the NNAR results were slightly better for both 8-week data and 4-week data. All RMSE, MAE and MAPE results for the NNAR were smaller than the ARIMA in the 4-week model.

Specifically, standard ARIMA modeling using 8-week and 4-week data both gave MAPE result between 1.252% and 3.752%, RMSE between 0.026 kWh and 0.288 kWh and MAE between 0.006 and 0.045. The modeling results confirmed that the 4-week NNAR method had a lower error rate compared to ARIMA.

5.3. Prediction results

Table 2 presents the prediction results obtained using both the NNAR and ARIMA models for real-time prediction during the following week. The MAPE for the 8-week NNAR model had a minimum value of 0.72%, a maximum value of 18.39% and an average value of 10.69%. The results showed that the prediction accuracy fell within acceptable MAPE that is around 10%. Furthermore, considering random office occupants activities, MAPE consumption range are all below 20%. Therefore, the prediction results showed a satisfactory data fit in the 8-week NNAR model. Similarly, 4-week model obtained a minimum value of 0.38%, maximum value of 18.64%, and an average value of 10.66%. Therefore, the 4-week model also had a very good data fit.

The RMSE evaluation showed that the predictions obtained by the 8-week NNAR model had a minimum value of 0.020 kWh, a maximum value of 0.35 kWh and an average value of 0.156 kWh. Likewise, prediction results for the 4-week model had a minimum value of 0.02 kWh, a maximum value of 0.51 kWh and an average value of 0.158 kWh. The maximum value of 8-week NNAR was 0.35 kWh which occurred on Tuesday as depicted in Table 2. The power consumption in the office increased because that is the day before weekly meeting. Everyone is preparing for meeting presentation and working progress report. Similarly, the maximum value of 4-week NNAR is 0.51 kWh which occurred on Wednesday, the day of weekly meeting. Furthermore, the values indicate the office occupants' activities increased during those two days. In view of the daily consumption of approximating 1 kWh, the average difference of 0.156 kWh and 0.158 kWh are considered satisfactory. Therefore, the NNAR models based on 4-week and 8-week test data had a very good data fit.

Further analysis of the MAE results in the 8-week NNAR model showed a minimum value of 0.01 kWh, a maximum value of 0.18 kWh and an average value of 0.096 kWh, which indicated a very good data fit. In the 4-week NNAR model, the MAE results had a minimum value of 0.00 kWh, a maximum value of 0.18 kWh and an average value of 0.098 kWh. Therefore, the model fit very well in both NNAR 4-week and 8-week data streams.

Table 2
Comparison of accuracy using test dataset.

Method	Descriptive statistics	Monday			Tuesday			Wednesday			Thursday			Friday			Overall		
		MAPE (%)	RMSE (kWh)	MAE (kWh)	MAPE (%)	RMSE (kWh)	MAE (kWh)	MAPE (%)	RMSE (kWh)	MAE (kWh)	MAPE (%)	RMSE (kWh)	MAE (kWh)	MAPE (%)	RMSE (kWh)	MAE (kWh)	MAPE (%)	RMSE (kWh)	MAE (kWh)
NNAR 8-week data stream	Mean	10.60	0.17	0.10	11.53	0.20	0.11	11.59	0.15	0.10	9.27	0.13	0.08	10.48	0.13	0.09	10.69	0.16	0.10
	Min.	6.93	0.11	0.08	7.52	0.13	0.07	5.85	0.11	0.05	6.70	0.10	0.06	0.72	0.02	0.01	0.72	0.02	0.01
	Max.	15.56	0.31	0.17	16.9	0.35	0.18	18.39	0.19	0.15	12.27	0.15	0.10	17.48	0.19	0.15	18.39	0.35	0.18
NNAR 4-week data stream	Mean	10.48	0.17	0.10	10.40	0.16	0.10	11.90	0.19	0.11	10.00	0.14	0.09	10.50	0.13	0.09	10.66	0.16	0.10
	Min.	5.67	0.11	0.06	0.38	0.02	0	5.95	0.12	0.05	6.810	0.10	0.060	0.67	0.02	0	0.38	0.02	0.00
	Max.	15.55	0.29	0.17	16.99	0.35	0.18	18.64	0.51	0.15	14.990	0.20	0.160	17.41	0.19	0.15	18.64	0.51	0.18
ARIMA 8-week data stream	Mean	13.20	0.19	0.04	13.74	0.21	0.05	13.79	0.23	0.07	13.72	0.17	0.03	12.62	0.15	0.03	13.41	0.19	0.04
	Min.	3.50	0.05	0.00	7.52	0.13	0.02	1.28	0.03	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Max.	19.87	0.34	0.12	22.60	0.36	0.13	21.61	0.53	0.28	18.60	0.24	0.06	18.31	0.26	0.07	22.60	0.53	0.28
ARIMA 4-week data stream	Mean	12.05	0.18	0.04	14.91	0.22	0.05	13.79	0.23	0.07	13.96	0.17	0.03	9.26	0.12	0.02	12.79	0.18	0.04
	Min.	3.50	0.05	0.00	9.55	0.13	0.02	1.28	0.03	0.00	0.01	0.00	0.00	0.65	0.02	0.00	0.01	0.00	0.00
	Max.	16.83	0.34	0.12	23.02	0.36	0.13	21.61	0.53	0.28	20.54	0.24	0.06	15.02	0.18	0.03	23.02	0.53	0.28

As in the training data set, NNAR results were compared with ARIMA results in both the 8-week and 4-week models. The test results were identical to the training results, which confirmed that the performance of the NNAR model was slightly superior. However, the indicators are quite deceptive since the main advantage of the NNAR model over standard ARIMA is its capability to reveal consumption patterns (Fig. 5), which is not possible in standard ARIMA (Fig. 6).

6. Analytical results and discussions

6.1. Anomaly detection results

The anomaly detection procedure applied the two sigma rule which classifies any points outside of 2 SD from the mean as anomalous data. However, further observation of daily activities showed that the occupants occasionally used the printer, which instantaneously increased electricity consumption by over 2 SD. Therefore, another rule was included so that printing was not defined as anomalous activity. Observations also showed that printing activity usually lasted less than 5 min. The final rule then considers electricity consumption anomalous if the consumption exceeds the prediction by 2 SD at least 5 min. Although the rule is only applicable in this case, the building manager or tenants can later adjust the rule according to their own requirements. Likewise, Fig. 7 shows examples of anomalous activity, which are indicated by large gaps between predicted and actual consumption.

Table 3 depicts the anomaly detection results obtained by NNAR method based on 8-week data and 4-week data. The detection results show that NNAR can identify consumption patterns and detect anomalies. Furthermore, the method can also differentiate normal and anomalous consumption precisely. After excluding Monday of week 12, the 8-week and 4-week models achieved only 51.25% and 51.32% accuracy, respectively, due to incomplete previous historical data. The NNAR can predict power consumption with 76.46–99.65% accuracy for the 8-week data window (average 89.1–96.5%). For 4-week data window, the average accuracy is 86.8–94.72% with normal consumption between 76.25% and 100% except for Thursday on week 9.

However, Table 4, which contains consecutive ARIMA results for 8-week data and 4-week data, shows that the ARIMA models could not identify consumption patterns. The methods were also unable to differentiate between normal and anomalous consumption when 2-sigma rule was applied. Thus, Table 4 shows that most incorrectly categorized data were classified as anomalous. Moreover, 60.47% of 8-week results and 55.81% of 4-week results were categorized as anomalous consumption. The result shows that, after 9 weeks, the 8-week data provided normal consumption values between 0 and 100% with average values between 33.45% and 38.31% except for Thursday, which revealed an average value of 86.31%. The 4-week data obtained normal consumption values between 0% and 100% with average values between 32.21% and 58.92% except for Thursday, which had an average value of 76.2%.

6.2. Discussions

The main objective of this study was to develop a method of predicting anomalous power consumption in real time. Therefore, the method must be able to process large data streams quickly in a real-time environment. Predictions were obtained for both 8-week data and 4-week data. Standard ARIMA was also compared. The experiment was motivated by the parsimony principle. That is, for an IBM server with a quad-core processor and 64 GB memory, predictions based on 4-week data can be obtained in 3 min, which is 2 min faster than predictions based on 8-week data. Therefore, using less data reduces prediction time, which is suitable for a real-time environment.

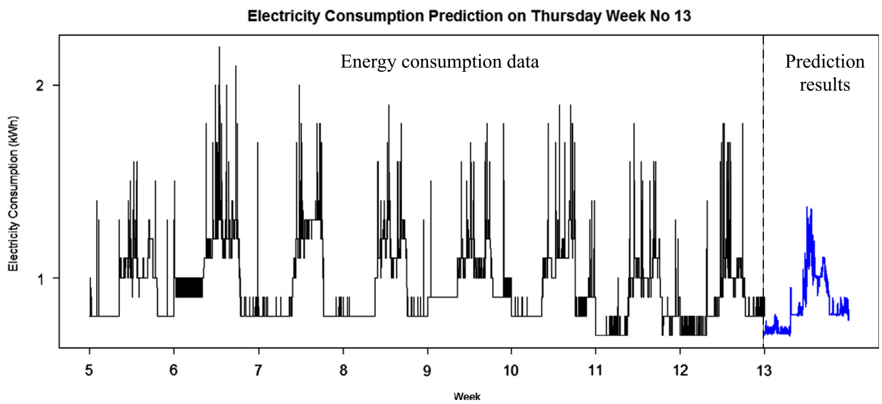

Fig. 5. Electricity consumption data and prediction results of NNAR on Thursday of week 13 based on 8 weeks of historical data.

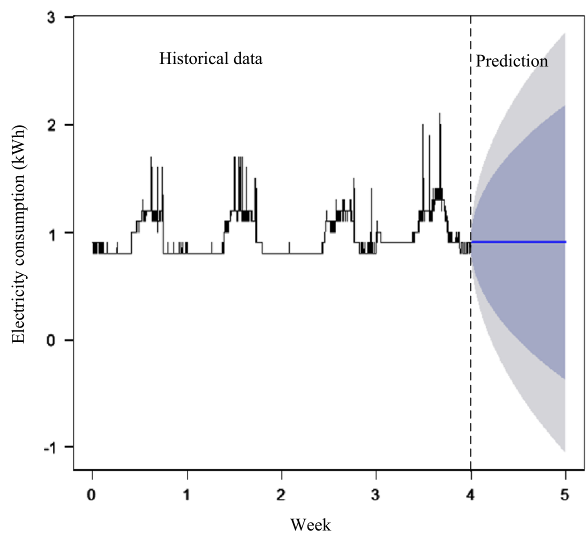

Fig. 6. Electricity consumption data and prediction result of ARIMA(3,1,0) on Friday of week 9 based on 4 weeks of historical data.

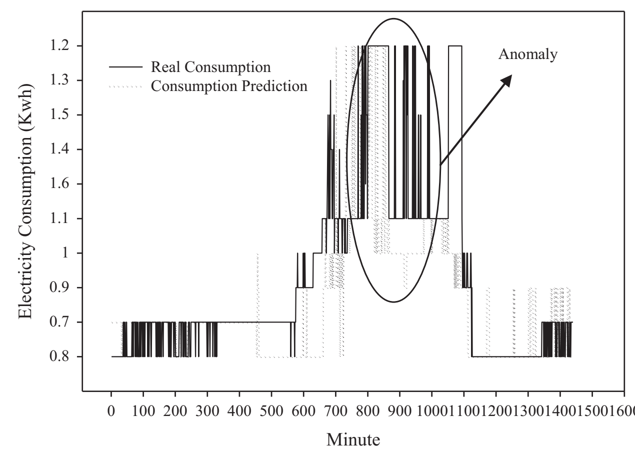

Fig. 7. Electricity consumption on Thursday of week 13.

The MAPE results for 9-week electricity consumption prediction data showed that 44.18% of predictions based on 8-week data had MAPEs below 10% while 37.78% of predictions based on 4-week data had MAPEs below 10%, which are considered satisfactory. The remainder had values between 10% and 20%, which are considered good. Moreover, comparisons of minimum, maximum and average values showed that the 4-week models had lower minimum and average values but higher maximum values compared to the 8-week models. These facts indicate that 4-week models are slightly more volatile than 8-week models. Further analysis based on observation showed that MAPE values larger than 10% indicate changing consumption patterns while MAPE values lower than 10% indicate that consumption is similar to that in the previous week.

The average RMSE values for training models based on 8-week and 4-week data were 0.084 kWh and 0.088 kWh, respectively, which were small enough to compare daily consumption approximating 1 kWh. Further comparisons of minimum and maximum values also showed no significant differences between the models. Likewise, the average and minimum predicted values were almost identical, but the maximum values obtained by the 4-week model were significantly larger than those obtained by the 8-week model (0.51 kWh and 0.35 kWh, respectively. The values indicate that the 4-week models are more error-prone compared to 8-week models.

Comparison of MAE values in the two training models showed that the 4-week models had a lower minimum value and a larger maximum value compared to the 8-week models. However, average values were similar (0.024 kWh and 0.028 kWh in the 4-week model and 8-week model, respectively). Moreover, the two models obtained similar maximum and minimum values in the testing datasets. Also, average values were the same, *i.e.*, 0.1 kWh for both 4-week and 8-week models.

Further comparison with standard ARIMA models showed that 4-week NNAR models performed better in terms of MAPE, RMSE and MAE. Although both NNAR and standard ARIMA produced comparable prediction results, standard ARIMA method did not perform well in the anomaly detection stage. Since the ARIMA predictions also tended to converge to the mean value, they could not reveal power consumption patterns. Consequently, when 2-sigma rule was applied, data were categorized as anomalous data. In contrast, the NNAR method revealed power consumption patterns.

Comparisons of 8-week and 4-week NNAR model performance using training data showed similar performance in terms of average MAPE, RMSE, and MSE values. Further evaluations using real-world data also yielded similar performances for both models. Although the 4-week models are more volatile and more error-prone compared to 8-week models, the difference is small, and the 4-week models have the advantage of lower computational load. Thus, in terms of the trade-off between performance and computational load, 4-week models are better than 8-week models.

7. Conclusions

The objective of this study was to develop a fast and accurate method of real-time anomaly detection. Anomaly detection is essential for effective building power demand management.

Table 3
Anomaly detection by NNAR using 8-week and 4-week data.

Dataset	Day	Consumption prediction	Week									Average (%)
			9	10	11	12	13	14	15	16	17	
8-Week data stream	Monday	Normal consumption	98.47	–	–	51.25	98.82	95.14	99.65	87.57	92.78	89.1
		Anomalous consumption	1.53	–	–	48.75	1.18	4.86	0.35	12.43	7.22	10.9
	Tuesday	Normal consumption	96.87	94.03	98.06	86.18	90.14	99.58	98.4	93.61	91.94	94.31
		Anomalous consumption	3.13	5.97	1.94	13.82	9.86	0.42	1.6	6.39	8.06	5.69
	Wednesday	Normal consumption	95.35	95.56	91.11	99.03	97.57	96.04	89.79	95.28	87.57	94.14
		Anomalous consumption	4.65	4.44	8.89	0.97	2.43	3.96	10.21	4.72	12.43	5.86
	Thursday	Normal consumption	97.36	97.15	98.96	98.61	96.11	97.85	99.65	98.96	83.82	96.5
		Anomalous consumption	2.64	2.85	1.04	1.39	3.89	2.15	0.35	1.04	16.18	3.5
	Friday	Normal consumption	99.51	93.12	91.46	99.65	94.86	97.08	99.03	76.46	94.93	94.01
		Anomalous consumption	0.49	6.88	8.54	0.35	5.14	2.92	0.97	23.54	5.07	5.99
4-Week data stream	Monday	Normal consumption	99.1	–	–	51.32	98.89	95.83	99.65	99.17	93.26	91.03
		Anomalous consumption	0.9	–	–	48.68	1.11	4.17	0.35	0.83	6.74	8.97
	Tuesday	Normal consumption	97.57	94.1	97.57	85.42	90	100	91.81	93.47	84.65	92.73
		Anomalous consumption	2.43	5.9	2.43	14.58	10	0	8.19	6.53	15.35	7.27
	Wednesday	Normal consumption	95.69	95.62	97.99	99.03	97.29	96.04	88.68	94.86	87.29	94.72
		Anomalous consumption	4.31	4.38	2.01	0.97	2.71	3.96	11.32	5.14	12.71	5.28
	Thursday	Normal consumption	12.65	96.94	98.89	98.54	95.97	96.04	99.72	99.17	83.26	86.8
		Anomalous consumption	87.35	3.06	1.11	1.46	4.03	3.96	0.28	0.83	16.74	13.2
	Friday	Normal consumption	99.51	92.85	91.18	99.58	94.86	97.22	98.61	76.25	95.28	93.93
		Anomalous consumption	0.49	7.15	8.82	0.42	5.14	2.78	1.39	23.75	4.72	6.07

Table 4
Anomaly detection by ARIMA using 8-week and 4-week data.

Dataset	Day	Consumption prediction	Week									Average (%)
			9	10	11	12	13	14	15	16	17	
8-Week data stream	Monday	Normal consumption	20.76	–	–	30.28	24.86	12.36	27.85	56.04	72.01	34.88
		Anomalous consumption	79.24	–	–	69.72	75.14	87.64	72.15	43.96	27.99	65.12
	Tuesday	Normal consumption	47.64	25.07	14.1	19.24	9.79	30.28	67.71	52.64	78.33	38.31
		Anomalous consumption	52.36	74.93	85.9	80.76	90.21	69.72	32.29	47.36	21.67	61.69
	Wednesday	Normal consumption	34.37	26.81	15.28	12.22	25.35	41.53	13.19	62.43	89.72	35.66
		Anomalous consumption	65.63	73.19	84.72	87.78	74.65	58.47	86.81	37.57	10.28	64.34
	Thursday	Normal consumption	0	100	100	98.61	97.71	94.17	89.79	96.53	99.93	86.3
		Anomalous consumption	100	0	0	1.39	2.29	5.83	10.21	3.47	0.07	13.7
	Friday	Normal consumption	5.76	12.92	58.89	42.08	36.94	32.64	6.87	4.93	100	33.45
		Anomalous consumption	94.24	87.08	41.11	57.92	63.06	67.36	93.13	95.07	0	66.55
4-Week data stream	Monday	Normal consumption	20.76	–	–	67.43	24.86	12.36	27.85	56.04	72.01	40.19
		Anomalous consumption	79.24	–	–	32.57	75.14	87.64	72.15	43.96	27.99	59.81
	Tuesday	Normal consumption	1.81	25.14	14.1	20.69	9.79	30.28	67.71	52.64	67.71	32.21
		Anomalous consumption	98.19	74.86	85.9	79.31	90.21	69.72	32.29	47.36	32.29	67.79
	Wednesday	Normal consumption	34.37	26.81	15.28	12.22	25.35	41.53	13.54	62.43	89.72	35.69
		Anomalous consumption	65.63	73.19	84.72	87.78	74.65	58.47	86.46	37.57	10.28	64.31
	Thursday	Normal consumption	0	7.64	100	100	97.71	94.17	89.79	96.53	99.93	76.2
		Anomalous consumption	100	92.36	0	0	2.29	5.83	10.21	3.47	0.07	23.8
	Friday	Normal consumption	99.51	32.99	31.04	42.08	94.86	32.64	6.87	95	95.28	58.92
		Anomalous consumption	0.49	67.01	68.96	57.92	5.14	67.36	93.13	5	4.72	41.08

Increased government regulation and new certification schemes also require improved energy efficiency in buildings [7]. Likewise, anomalous power consumption indicates inefficient building power consumption. Therefore, inefficiencies can be quickly recognized so that the building manager and/or tenants can take appropriate action.

Several researchers have proposed real-time anomaly detection methods. However, methods proposed so far require huge training data sets, produce static models [33] and have unsatisfactory accuracy [34]. Therefore, this study proposes a combination of prediction and two-sigma rule for detecting anomalous pattern in real time. The proposed prediction model uses a hybrid method of univariate time series ARIMA and ANN based on per-minute electricity consumption in an office space.

The prediction accuracy of 4-week and 8-week NNAR models were evaluated and compared to standard ARIMA. The comparison results confirm that the hybrid NNAR method obtains more accurate predictions of electricity consumption compared to standard ARIMA. Moreover, the models based on 4 weeks and 8 weeks of rolling training data performed similarly. Specifically, average real-time predictions were between 89.1% and 96.5% for the 8-week data window, and between 86.8% and 94.72% for the 4-week data window.

Furthermore, the model for predicting daily consumption during the following week could be developed in only 3 min using the previous 4-week consumption dataset. Therefore, anomalies could be detected in real time using a simple and quick calculation of differences between real-time and predicted consumption data. Consequently, the approach is suitable for use in a real-time environment and large data environment. Finally, considering the parsimony principle and trade-off between performance and computational load, 4 weeks of training data are adequate for achieving good results.

The research contributes to the formalization of a methodology for real-time detection of anomalous patterns in large data sets. Particularly, as the method has two stages, the prediction part helps building managers plan their energy consumption while the anomaly detection part helps building managers to identify unusual consumption of electricity by tenants. Moreover, the tenants can understand how their business activities consume energy through their anomalous energy consumption.

As power consumption feedback is essential for maximizing energy savings, implementing optimization algorithms and adding explanatory variables into a building management system are the next stage of this research. Two other research directions are possible. Firstly, as time-based pricing policy becomes widely implemented in many countries, providing end users in residential buildings with feedback by suggesting operation time scheduling options for minimizing energy consumed by appliances is essential. Secondly, implementing smart meters and sensors will rapidly accumulate electricity consumption data. Therefore, big data infrastructure must be applied to accommodate and analyze unstructured data resulted from the smart grids.

References

[1] IEA. World energy outlook 2011 executive summary. Paris: International Energy Agency; 2011.
[2] Costa A, Keane MM, Raftery P, O'Donnell J. Key factors methodology—a novel support to the decision making process of the building energy manager in defining optimal operation strategies. Energy Build 2012;49:158–63.
[3] Ning SK, Yeh SC, Chen JC, Hung MC, Lin QG, Cai YP, et al. Identification of optimal strategies for sustainable energy management in Taiwan. Int J Energy Res 2013;37:15.
[4] Palmer J, Cooper I. Great Britain's housing energy fact file. Cambridge: Department of Energy and Climate Change; 2011.
[5] Janda KB. Buildings don't use energy: people do. Archit Sci Rev 2011;54:15–22.
[6] Janda K. Overview of energy standards for buildings. Report for United Nations forum on energy efficiency and energy security for sustainable development: taking collaborative action on climate change. Seoul, Republic of Korea; 2007.
[7] Popescu D, Bienert S, Schützenhofer C, Boazu R. Impact of energy efficiency measures on the economic value of buildings. Appl Energy 2012;89:454–63.
[8] Costa A, Keane MM, Torrens JI, Corry E. Building operation and energy performance: monitoring, analysis and optimisation toolkit. Appl Energy 2013;101:310–6.
[9] Yu T. Modeling occupancy behavior for energy efficiency and occupants comfort management in intelligent buildings. In: 4th international conference on machine learning and applications. Washington, USA: IEEE; 2010. p. 726–31.
[10] Lu J, Sookoor T, Srinivasan V, Gao G, Holben B, Stankovic J, et al. The smart thermostat: using occupancy sensors to save energy in homes. In: Proceedings of the 8th ACM conference on embedded networked sensor systems. Zurich, Switzerland: ACM; 2010. p. 211–24.
[11] Erickson VL, Lin Y, Kamthe A, Brahme R, Surana A, Cerpa AE, et al. Energy efficient building environment control strategies using real-time occupancy measurements. In: Proceedings of the 1st ACM workshop on embedded sensing systems for energy-efficiency in buildings. Berkeley, California: ACM; 2009. p. 19–4.
[12] Galasiu AD, Newsham GR. Energy savings due to occupancy sensors and personal controls: a pilot field study. In: European lighting conference. Istanbul; 2009. p. 745–52.
[13] Jang W-S, Healy WM, Skibniewski MJ. Wireless sensor networks as part of a web-based building environmental monitoring system. Autom Constr 2008;17:729–36.
[14] Fontugne R, Ortiz J, Tremblay N, Borgnat P, Flandrin P, Fukuda K, et al. Strip, bind, and search: a method for identifying abnormal energy consumption in buildings. In: Proceedings of the 12th international conference on information processing in sensor networks. Philadelphia, Pennsylvania, USA: ACM; 2013. p. 129–40.
[15] Depuru SSSR, Wang L, Devabhaktuni V. Smart meters for power grid: challenges, issues, advantages and status. Renewable Sustainable Energy Rev 2011;15:2736–42.
[16] Markovic DS, Zivkovic D, Branovic I, Popovic R, Cvetkovic D. Smart power grid and cloud computing. Renewable Sustainable Energy Rev 2013;24:566–77.
[17] Joung M, Kim J. Assessing demand response and smart metering impacts on long-term electricity market prices and system reliability. Appl Energy 2013;101:441–8.
[18] Accenture. Ten leading practices for smart grid analytic. Accenture; 2011.
[19] Mohod SW, Aware MV. A novel PWM-inverter scheme for grid connected WEGS for power quality measure. In: India conference (INDICON) annual IEEE 2009; 2009. p. 1–4.
[20] Naik CA, Kundu P. Power quality index based on discrete wavelet transform. Int. J. Electr. Power Energy Syst. 2013;53:994–1002.
[21] Katsaprakakis DA, Christakis DG, Zervos A, Voutsinas SA. Power-quality measure. IEEE Trans Power Delivery 2008;23:553–61.
[22] Ivanov C, Getachew L, Fenrick SA, Vittetoe B. Enabling technologies and energy savings: the case of energywise smart meter pilot of connexus energy. Util Policy 2013;26:76–84.
[23] Allen D, Janda K. The effects of household characteristics and energy use consciousness on the effectiveness of real-time energy use feedback: a pilot study American council for an energy-efficient economy 2006 summer study, human and social dimensions of energy use. Washington, DC, USA: ACEE; 2006; 1.
[24] Darby S. Smart metering: what potential for householder engagement? Build Res Inf 2010;38:442–57.
[25] Fei L, Lee YM, Huijing J, Snowdon J, Bobker M. Statistical modeling for anomaly detection, forecasting and root cause analysis of energy consumption for a portfolio of buildings. New York: IBM; 2011.
[26] Wrinch M, El-Fouly THM, Wong S. Anomaly detection of building systems using energy demand frequency domain analysis. In: 2012 IEEE Power and Energy Society General Meeting. IEEE; 2012. p. 1–6.
[27] Hongfei L, Dongping F, Mahatma S, Hampapur A. Usage analysis for smart meter management. In: 8th international conference and expo on emerging technologies for a smarter world (CEWIT). New York, USA; 2011. p. 1–6.
[28] Wang L, Greenberg S, Fiegel J, Rubalcava A, Earni S, Pang X, et al. Monitoring-based HVAC commissioning of an existing office building for energy efficiency. Appl Energy 2013;102:1382–90.
[29] Office based companies maximising savings in an office environment. London, UK: Carbon Trust; 2010.
[30] Vikhorev K, Greenough R, Brown N. An advanced energy management framework to promote energy awareness. J Clean Prod 2013;43:103–12.
[31] Zhao H, Magoulès F. A review on the prediction of building energy consumption. Renewable Sustainable Energy Rev 2012;16:3586–92.
[32] Aziz AFA, Khalid SN, Mustafa MW, Shareef H, Aliyu G. Artificial intelligent meter development based on advanced metering infrastructure technology. Renewable Sustainable Energy Rev 2013;27:191–7.
[33] Brown M, Barrington-Leigh C, Brown Z. Kernel regression for real-time building energy analysis. J Build Perform Simul 2011;5:263–76.
[34] Yi Z, Weiwei C, Black J. Anomaly detection in premise energy consumption data. In: 2011 IEEE power and energy society general meeting. San Diego, USA: IEEE; 2011. p. 1–8.
[35] Fontugne R, Tremblay N, Borgnat P, Flandrin P, Esaki H. Mining anomalous electricity consumption using ensemble empirical mode decomposition. In: The 38th international conference on acoustics, speech, and signal processing. Canada; 2013.
[36] Yang J, Rivard H, Zmeureanu R. On-line building energy prediction using adaptive artificial neural networks. Energy Build 2005;37:1250–9.
[37] Yao Y, Sharma A, Golubchik L, Govindan R. Online anomaly detection for sensor systems: a simple and efficient approach. Perform Eval 2010;67:1059–75.
[38] Zhang PG. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 2003;50:159–75.
[39] Khashei M, Bijari M. A new class of hybrid models for time series forecasting. Expert Syst Appl 2012;39:4344–57.
[40] Suganthi L, Samuel AA. Energy models for demand forecasting—a review. Renewable Sustainable Energy Rev 2012;16:1223–40.
[41] Sumer KK, Goktas O, Hepsag A. The application of seasonal latent variable in forecasting electricity demand as an alternative method. Energy Policy 2009;37:1317–22.
[42] Khashei M, Bijari M. An artificial neural network (p,d,q) model for timeseries forecasting. Expert Syst Appl 2010;37:479–89.
[43] Tseng F-M, Yu H-C, Tzeng G-H. Combining neural network model with seasonal time series ARIMA model. Technol Forecast Soc Change 2002;69:71–87.
[44] Koutroumanidis T, Ioannou K, Arabatzis G. Predicting fuelwood prices in Greece with the use of ARIMA models, artificial neural networks and a hybrid ARIMA–ANN model. Energy Policy 2009;37:3627–34.
[45] Liu H, Tian H-q, Li Y-f. Comparison of two new ARIMA–ANN and ARIMA–Kalman hybrid methods for wind speed prediction. Appl Energy 2012;98:415–24.
[46] Areekul P, Senjyu T, Urasaki N, Yona A. Next day price forecasting in deregulated market by combination of artificial neural network and ARIMA time series models. In: The 5th IEEE conference on industrial electronics and applications (ICIEA). Taichung, Taiwan; 2010. p. 1451–6.
[47] Kalyani S, Swarup KS. Particle swarm optimization based K-means clustering approach for security assessment in power systems. Expert Syst Appl 2011;38:10839–46.
[48] Venables WN, Smith DM, Team RC. An introduction to R: notes on R: a programming environment for data analysis and graphics version 3.0.22013.
[49] Chou J, Chiu C, Farfoura M, Al-Taharwa I. Optimizing the prediction accuracy of concrete compressive strength based on a comparison of data-mining techniques. J Comput Civil Eng 2011;25:242–53.
[50] Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice. OTexts; 2013.