AIMS Energy, 8(5): 783–801.
DOI: 10.3934/energy.2020.5.783
Received: 27 May 2020
Accepted: 19 August 2020
Published: 27 August 2020

http://www.aimspress.com/journal/energy

*Research article*

# Data-driven predictive models for daily electricity consumption of academic buildings

**Bilal Akbar[1], Khuram Pervez Amber[1,2,*], Anila Kousar[3], Muhammad Waqar Aslam[4], Muhammad Anser Bashir[1] and Muhammad Sajid Khan[1]**

[1] Department of Mechanical Engineering, Mirpur University of Science and Technology (MUST), Mirpur 10250, AJK, Pakistan
[2] Faculty of Engineering, Science and the Built Environment, London South Bank University, London SE1 0AA, UK
[3] Department of Electrical Engineering, Mirpur University of Science and Technology (MUST), Mirpur 10250, AJK, Pakistan
[4] Department of Computer Systems Engineering, Mirpur University of Science and Technology (MUST), Mirpur 10250, AJK, Pakistan

* **Correspondence:** Email: khuram.parvez@must.edu.pk; Tel: +923415964460.

**Abstract:** Academic buildings in a typical university campus occupy 42% of the total space and are responsible for nearly 50 percent of the total energy use and carbon emissions of the campus. Forecasting of energy consumption in this energy intensive building category could help higher education institutions in taking energy saving initiatives and in revising their building operating strategies. Reliable predictive techniques does not only help in forecasting a building' energy consumption, but also help in identifying a variety of factors affecting the energy consumption of that building. This study attempts to forecast and benchmark the daily electricity consumption of an academic building situated in London, United Kingdom using two different data-driven modeling techniques, i.e., Multiple Regression and Artificial Neural Network. Hourly dataset for the electricity consumption was available for the period 2007 to 2011 from the smart meter whereas hourly data of different factors such as ambient temperature, relative humidity, wind speed and solar radiation were downloaded from the website of environmental research group of Kings College London. The performances of the two predictive models have been critically analyzed by comparing their predicted consumption with a real dataset of the same building for the year 2012. A comparison shows that both Multiple Regression (MR) and Artificial Neural Network (ANN) perform reasonably well with a Mean Absolute Percentage Error (MAPE) of 3.34% and 2.44% for working days and 5.12%

and 4.59% for non-working days respectively. ANN performs slightly better than MR. This energy consumption forecasting approach can easily be adapted for predicting energy use of similar buildings.

**Keywords:** data-driven forecasting models; electricity consumption; academic buildings; ANN; MR

## 1. Introduction

Buildings are responsible for about 30–45% of the global energy demand and are the major energy consuming sector [1]. Academic buildings in a typical university campus in the United Kingdom occupy 42% of the total space and are responsible for nearly 50 percent of the total energy use and carbon emissions of the campus. Forecasting building energy consumption has gained a rapid momentum in recent years as it helps the facility managers in scheduling and controlling energy usage in the buildings [2]. It has also emerged as an essential approach for energy planning, management, and conservation [3]. Researchers [4–6] consider that energy consumption prediction is an effective and helpful technique in order to regulate buildings energy consumption. Nonetheless, because of the diverse nature and intricacy of buildings in addition to the uncertain climatic conditions, forecasting of energy utilization becomes more difficult [7]. According to Ziyo and Ravi [8], major focus (57%) of researchers has been on the prediction of electricity usage in the buildings.

Energy consumption forecasting models could be classified into two main categories; (a) Engineering models and (b) Statistical models. Engineering models require comprehensive amount of building related data such as building materials, schedules, equipment, occupancy levels, etc. The engineering methods require application of physical principles in order to estimate the energy behavior of the complete building itself and or for its components and are highly complex [9].

Conversely, statistical models are constructed on the historical datasets of different variables and are therefore, classified as Data-Driven Models (DDM). This historical data could be obtained through a detailed energy audit of the building. Hence, the historical data is of great significance, the quantity and quality associated with it has a central role in constructing the statistical models. The DDM primarily governs analysis of a system especially finding relation between different state variables without mere dependency on physical behaviors as seen in previous case. The data-driven models are largely applied to predict energy usage of the building-structures of a large variety [10].

A range of energy consumption DDMs are available but these differ in complexity level, time consumption and ease to use. These techniques include; Artificial Neural Networks (ANN), Deep Neural Networks (DNN), Genetic Programming (GP), Genetic Algorithms (GA), and Multiple Regression (MR). Among these, DNN, GA and GP require huge amount of data for reliable results, Secondly, these three techniques are computationally expensive requiring a lot of time to regress a model [11]. On the other hand, ANN and MR are relatively two easier forecasting techniques which offer reliable results with minimum time consumption [12].

Artificial Neural Networks have received inspiration from biological neural network and present a computational model which works on the principle same as of human brain [13]. In the beginning ANN could not get attention of the researchers due to complexity associated with its functionality. However, ANN proved its significance by the availability of high-performance computing it contains. In a recent study, ANN application for energy consumption prediction was 41% in a review of prediction studies by Ahmad et al. [14]. ANN outperforms other techniques in solving input-output relationships for complex, multifaceted, non-linear quantities. It functions as an adaptive system that adapts the coming input and develops its structure according to the input it receives. The basic structure of ANN is composed of numerous neurons which are connected together in a specific and precised neural pattern. This pattern of neurons is further sub-divided into various layers. This division is very crucial in ANN as the number of layers define different architectures of ANN along with flow of data. The architectures are categorized as unsupervised (without training) and supervised (with training) ANN. Single layered and multi layered ANN are also among common known architectures of ANN. Figure 1 shows the simplified structure of ANN.

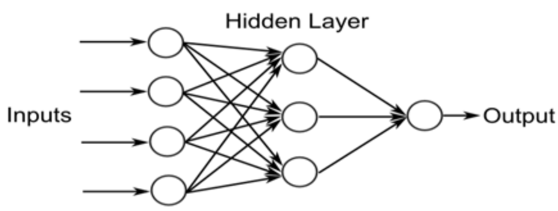


**Figure 1.** A Simple structure of ANN.

Interestingly, supervised ANN involves two-staged adaptive learning. These are training and testing stages. ANN follows the following steps during training. Training data is given as input to architecture, and random biases and weights for each neuron layer are calculated. The output of each neuron layer is calculated by an activation function and there is a sequential flow of output generated from one layer to next layer as its input. The final combined output obtained at the end of last layer is fed back to adjust deviations in biases and weights among intermediate layers in order to obtain the targeted output. Feedback mechanism is present in the ANN to match the required output with the finally generated output from the last layer of ANN.

$$h_i x = \frac{1}{1+e^{-\left(\sum w_i x_i^j + b\right)}} \tag{1}$$

where $x_i^j$ is the input with j variables, $w_i$ is weight and $h_i x$ is output at ith layer. In this research, the feedback process continuous until the output follows the maximum bounds. Over fitting problem is also resolved in this research using validation phase. The trained neural network is then tested with test data to get the accuracy metrics.

Multiple Regression (MR) is a simple and a common approach in predicting buildings energy consumption. The percentage of prediction studies involving MR is 26%, second to ANN in a review of prediction studies by [14]. A general form of MR model is shown in Eq 2.

$$\beta = \alpha_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + \cdots\cdots\cdots a_n x_n \tag{2}$$

where β is the output or dependent variable, $a_i$ are the regression coefficients (i = 0, 1, 2….n), $x_i$ are the independent variables (i = 1, 2, 3 …n).

The predictive power of a MR model is generally expressed in terms of Coefficient of Determination ($R^2$). Its value lies between '0' and '1'. A value of unity means there is a strong relationship between output and input variables whereas value below 0.5 represents a weaker relationship between the dependent and the independent variables [15].

In last two decades, ANN and MR have been widely used for the buildings energy prediction. Martellotta et al. [16] used sigmoid hidden layered and linear output neurons based double layered feed-forward ANN model to predict the hourly energy use of domestic sector in Italy. They selected 15 variables as input. Their results showed that ANN performed extremely well with very low relative errors, rarely exceeding a ±5% tolerance. Deb et al. [17] used ANN to estimate energy saving by HVAC system in 56 retrofitted organizational premises in Singapore. The hidden layer in their model was comprised of four neurons. The study included 14 variables as input to the ANN architecture. The model showed good results with MAPE and $R^2$ values of 14.80% and 0.70 respectively. In a similar study, Ekici and Aksoy [18] applied ANN technique to forecast energy consumption of buildings in Turkey and found that ANN with 3.43% variation shows outstanding forecasting of the order of 94.8 percent to 98.5 percent. Deb et al. [19] presented a study to predict energy consumption by cooling loads in three different academic buildings. This study was based on a two year data and ANN was applied to create a forecasting model. The investigation showed ANN performs outstanding with $R^2 > 0.94$.

Gul and Patidar [1] attempted to develop correlation of occupancy and electrical energy demand in academic building for period of January 2013 to May 2013 in UK. Their analysis presented interesting results showing the major electrical energy consumption is by building's own management system instead of occupants' activities. Kaytez et al. [5] compared energy usage prediction pattern of three distinct techniques viz, ANN, MR and support vector machine. The study was conducted in Turkey for a period from 1970 to 2009 using population, total installed capacity and generation as independent variables. Comparison showed that results obtained from ANN and MR deviated by 1.79% and 0.88% respectively as acquired by support vector machine. Park et al. [20] used ANN and MR to predict hourly performance of ground source heat pump in Korea in 2018. Comparison revealed that ANN offers 1.75% higher accuracy based on coefficient of variation of root mean squared error than the MR. Aranda et al. [21] developed MR models to forecast energy use of 55 banks in Spain. They found outstanding performance by the model in estimating the energy use of the branches. Bianco et al. [22] developed MR model to estimate electricity usage using historical gross domestic product (GDP), electricity consumption, population and GDP per capita in Italy for thirty-seven years (1970–2007). They found $R^2$ values equal to 0.961, 0.990 and 0.981 for non-residential, residential and total consumption respectively. Bingchun et al. [23] established three different prediction models for forecasting energy consumption in China by the end of 2021. Their models included ANN, multi-variable linear regression, and support vector machine considering imports & exports, gross domestic product and population as independent variables. The predicted results showed 47% (2954.04–5618.67 MTOE) variations in energy consumption in 2021. Joseph et al. [24] studied the impact of different climates in forecasting energy usage using MR technique. They had used building load (shading coefficient, window to wall ratio, lightning), HVAC system (fresh air, set point temperature for summer and winter) and HVAC plant (Chiller COP, boiler efficiency) as the inputs. Their proposed MR model elaborated $R^2$ ranging from 0.89 to 0.97 for very cold to moderate regions. Mohammad et al. [25] described the impact of building shape (rectangle, triangle, H, L, T and U shape) in forecasting the energy consumption along with climate impacts as done in [24] in US. They used DOE-2 and eQUEST softwares to simulate the buildings and performing R statistical analysis developed multi-linear regression model. The comparison of the simulated and the model predicted showed 95%

compatibility in the results with a difference of 5% in annual energy usage.

In a series of studies by Amber et al. [2,15,26] performance capabilities of various techniques are thoroughly evaluated for energy consumption in academic and administrative buildings. In [15] authors evaluated prediction pattern of two different techniques; MR and genetic programming; and compared the obtained results for administrative structures of London South Bank University using data of 2007–2012. Counter comparison of predicted results with actual values for the year 2013 confirmed the accuracy of the proposed models with 6% and 7% total absolute error for genetic programming and MR respectively. Forecasting competency of MR model for university sector energy usage was evaluated by Amber et al. [2]. They considered building type, ambient temperature, weekday index, solar radiation, wind speed and humidity as explanatory variables. They found building type, ambient temperature and weekday index had the most impact on sector's energy usage. They concluded the study with 12% and 13% normalized root mean square error for administrative and academic portions in the sector. In another study by Amber et al. [26] investigated the performance capabilities of five different forecasting techniques namely ANN, MR genetic programming, support vector machine and deep neural network. These techniques were trained for five year data of various weather parameters with weekday index explaining working and non-working days. The results of this study revealed outstanding performance by ANN with 6% Mean Absolute Percentage Error followed by 8.5% by MR.

There are many more studies [27–31] which have used ANN and MR for energy consumption prediction. These studies are summarized in Figure 2. It highlights the location, building type and error analysis for each study. It also indicates the technique used in these studies.

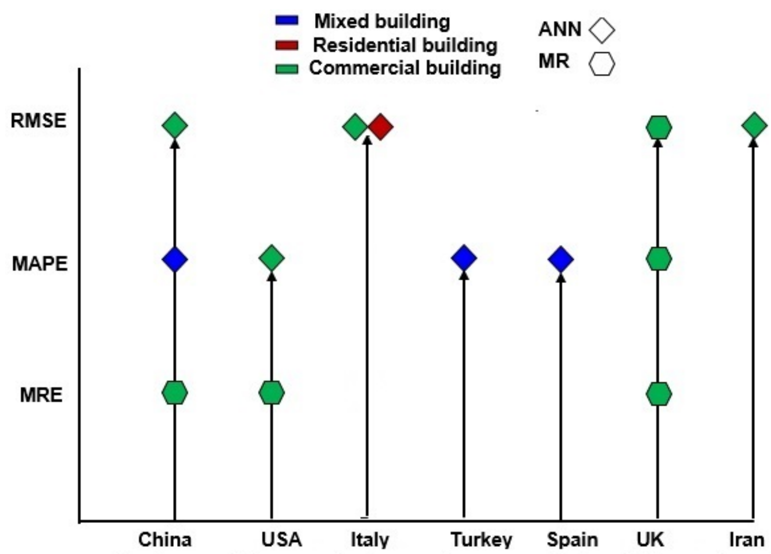


**Figure 2.** Buildings energy consumption forecasting studies.

This detailed literature review clearly shows that energy consumption forecasting is an important task for the energy management teams and a very limited literature has been published for the energy prediction of academic buildings category that takes into account various parameters. To fill this gap, this study aims to predict daily electricity consumption of an academic building using two forecasting techniques, i.e., ANN and MR based on the historical datasets. The rest of the paper is structured as follows: Section 2 describes Materials and Methods, Section 3 gives Results and Discussion whereas conclusions are drawn in Section 4.

## 2. Materials and method

This section describes different methods for collecting dataset and building related information. Dataset of different variables such as building occupancy, temperature, relative humidity, solar radiation and wind speed, have been plotted to observe daily and monthly variance whereas building's half-hourly electricity consumption data is transformed into hourly, daily and monthly consumption profiles. Different datasets cover a period 2007 to 2011 whereas dataset of 2012 is used for the validation of models.

### *2.1. Site selection*

For the purpose of this study, an academic building called 'Keyworth Centre' situated in the Southwark campus of London South bank University (LSBU) is marked as the most suited. The accessibility of building's electricity consumption data led to the selection of this building. University has a dedicated energy management team led by the energy manager. This team is responsible for observing, examining and evaluating energy usage data of the university buildings and to introduce energy efficiency by identifying energy saving opportunities, by installing renewable and clean energy technologies.

#### 2.1.1. Building description

The Keyworth Centre was built in 2003 and is situated at the Southwark campus of London South Bank University in London, UK. The building has six floors covering a total Gross Internal Area (GIA) of 8,588 m² and mainly consists of lecture theatres, class rooms, offices and a cafeteria.


**Figure 3.** A view of Keyworth Centre building.

Figure 3 shows a view of the Keyworth Centre (KWC) building. It operates between 7.30 am to 8 pm and closes at 10 pm every day. The building is a naturally ventilated building. Basement in the building has a Plant Room which houses six gas fired boilers (each of 160 kW) for providing space heating through radiators to all six floors. The building has two elevators found near the main reception for transportation to upper floor. Compact Fluorescent Lamps are used as lightning source in the complete building. Moreover, it has a single low voltage electricity connection. As per data collected from the office of the energy manager, the Keyworth Centre consumes nearly 7% of the total annual electricity consumption of London South Bank University.

### *2.2. Electricity consumption data*

The building has a single low voltage electricity connection. Hourly electricity usage dataset was available for the period of five years (i.e., 2007–2011) from the office of the energy manager. This data is being measured via smart meters. This has been transformed into daily and monthly figures. All equipment is used day to day, only pumps heating is not used during summer.

#### 2.2.1. Hourly electricity consumption data

Figure 4 shows average hourly electricity usage of Keyworth Centre for week days and weekends for each month of the year. It is apparent that the building has a consistent hourly demand of nearly 65 kW on the weekends throughout the year. As the building does not operate on weekends, yet it has the base load such as lighting which seems running all the time. With intelligent lighting controls, this unnecessary consumption could be minimized. In contrast the weekday load does vary with respect to time.

Load starts increasing from 5 am when the Building Management System starts the boilers to maintain the buildings space heating system before the staff and students arrive. During period of moderate temperature i.e. spring months (February & March) and fall months (October & November) the load climbs to as high as 200 kW at 9 am and then remains nearly same until 4 pm after which it starts dropping until it drops back to the base load at 10 pm when the staff closes the building. During summer months, the peak load drops to 170 kW. This drop in peak load could be attributed to no space heating demand which means the boilers and pumps would not run. Because the space heating equipment not only include boiler but also heavy duty pumps and when there is no space heating requirements these pumps will not run. December and January have reduced electricity demand. This could be attributed to lower occupancy due to Christmas holidays.

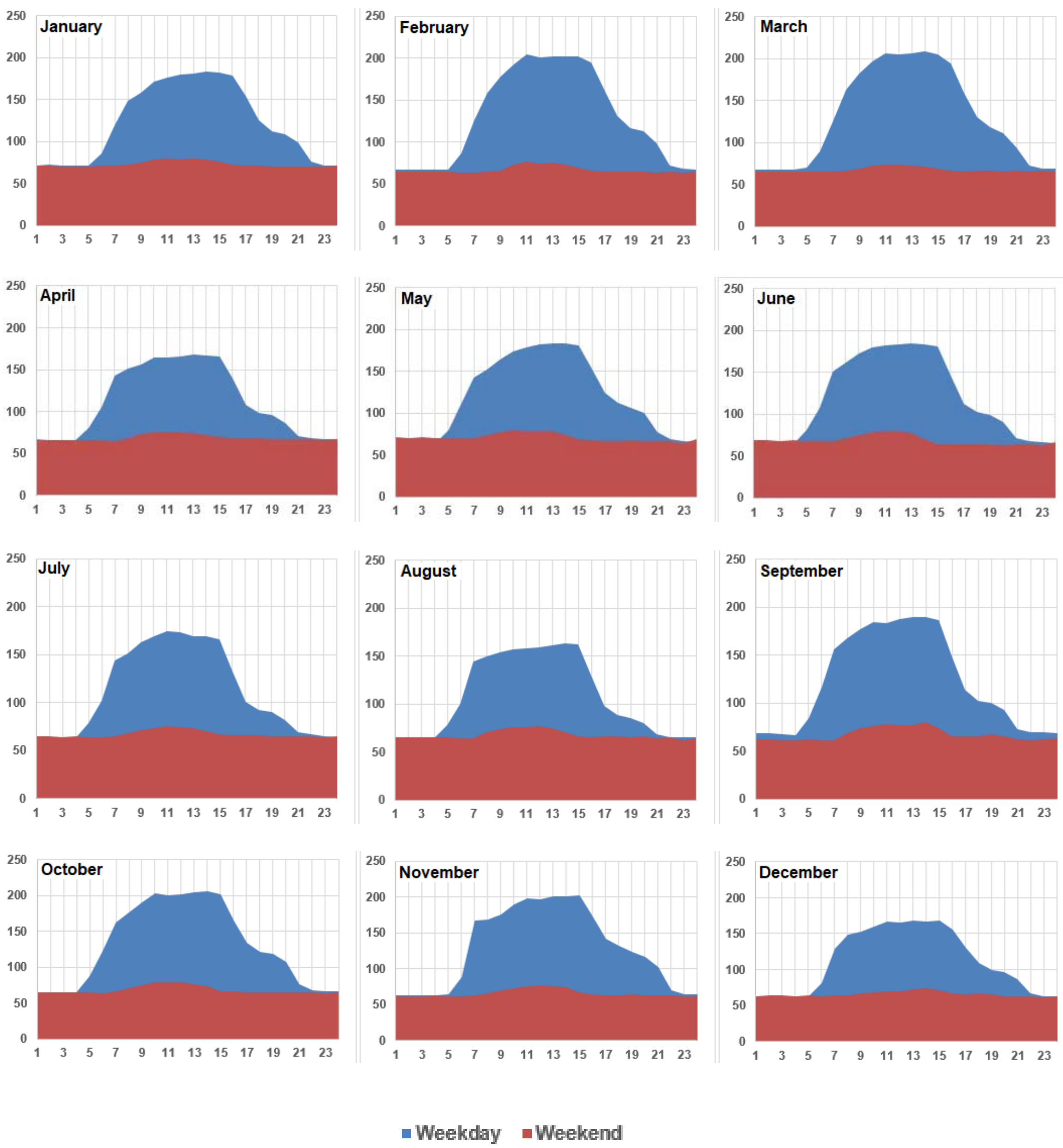


**Figure 4.** Hourly electricity usage of Keyworth Centre building (y-axis shows kW).

### 2.2.2. Daily electricity consumption data

Figure 5 presents daily average electricity use of KWC building on working days (WD) for years 2007 to 2011. It is observed that average daily electricity usage during winter months could be as high as 400 Wh/m$^2$ and drops to 300 Wh/m$^2$ during summer months.

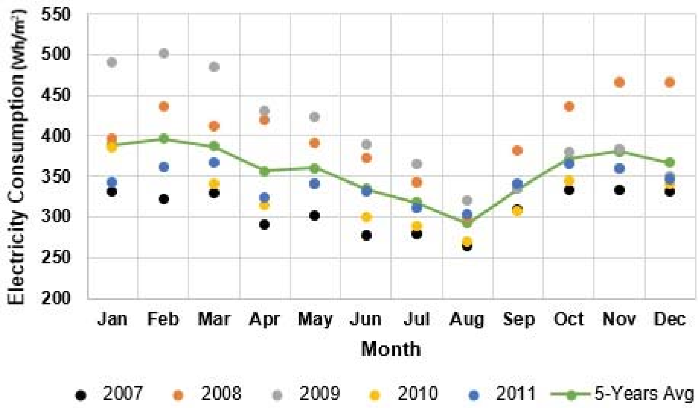


**Figure 5.** Electricity usage on working days.

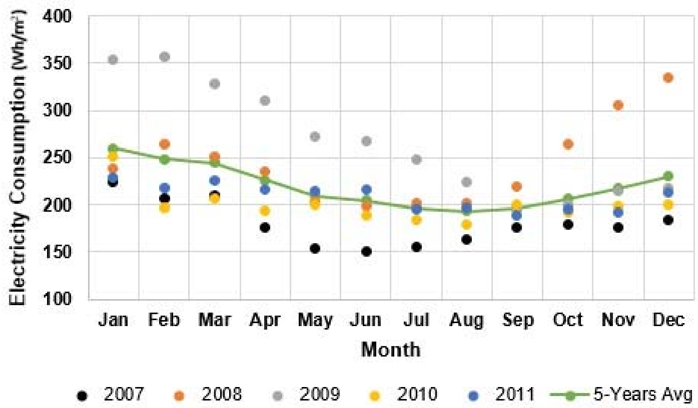


**Figure 6.** Electricity usage on non-working days.

Similarly, Figure 6 presents daily electricity usage on non-working days (NWD) for different months of years 2007 to 2011. Electricity consumption remains in the range of 200 to 250 Wh/m$^2$ throughout the year.

*2.3. Weather parameters*

Weather factors such as temperature, humidity and wind speed directly affect a building's energy consumption [32–34]. For the purpose of this study, data was online available from the website of environmental research group of Kings College London.

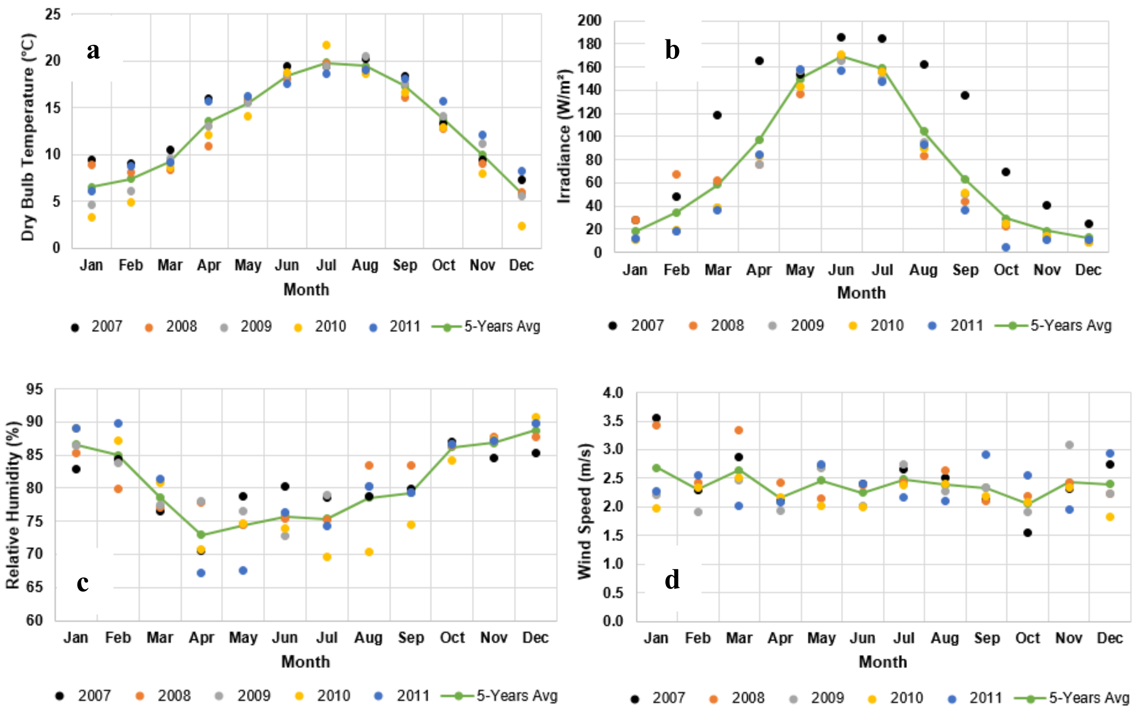


**Figure 7.** Monthly variation of weather parameters: (a) dry bulb temperature, (b) solar irradiance, (c) relative humidity, (d) wind speed.

Figure 7 shows the monthly average profiles of four weather parameters, i.e., wind speed, relative humidity, solar irradiance and temperature for five years, i.e., 2007–2011 for London Region.

It is apparent that average temperature during winter months remains below 10 °C whereas during summer months it remains between 15 °C to 20 °C. Highest temperature is recorded during the month of July and August. Year 2010 was observed colder compared to other years. Solar irradiance is lower during winter months, i.e., less than 100 W/m$^2$ whereas it is higher during summer months, i.e., >100 W/m$^2$ < 200 W/m$^2$. Irradiance was observed higher during 2007 whereas 2011 saw minimum solar irradiance. Humidity is lower during summer months, i.e., <80% whereas during winter months, average humidity values ae above 80%. Wind speed remains fairly constant between 2 m/s to 3 m/s throughout the year and very less variability has been observed in its monthly average values.

*2.4. Building occupancy*

In numerous studies, building occupancy has been observed as one of the significant variables that directly affect a buildings energy consumption [1,35,36]. In this case, no data was available for the building occupancy. However, realizing the significance of this important factor and to investigate its effect on the buildings electricity consumption, 'Weekday Index' as proxy variable is introduced to distinguish high and low occupancy periods. The variable has a value of 1 for WD and 0 for NWD. Non-working days include weekends and holidays. Figure 8 shows the number of non-working days during different months for years 2007 to 2011. It is apparent that average number of NWD are eight which represent the weekends. December has the highest NWD number due to Christmas break.

It is interesting to note that for the establishment of the study only SPSS software and a desktop computer is required.

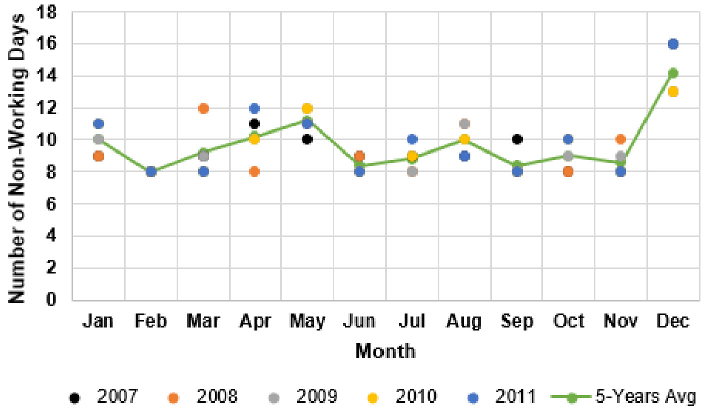


**Figure 8.** Number of non-working days.

## 3. Results and discussion

### *3.1. Results of multiple regression analysis*

Daily electricity consumption in Watt-Hours/Square meters (Wh/m$^2$) is the dependent variable, $'\beta'$ and was regressed on five independent variables;

a) Week Day Index (WDI), $(x_1)$
b) Dry Bulb Temperature, $(x_2)$
c) Solar Irradiance, $(x_3)$
d) Relative Humidity, $(x_4)$
e) Wind Speed, $(x_5)$

In the preliminary analysis, it was observed that there is collinearity among Dry Bulb Temperature, Solar Irradiance and Relative Humidity. Therefore, the latter two variables were eliminated. Wind speed was found insignificant with a t-stat value of less than 1.96 and p-value > 0.05. Only two variables were found significant, i.e., WDI $(x_1)$ and Dry Bulb temperature $(x_2)$. The MR model displays a good predictive power with R = +0.79 and $R^2 = 0.62$.

Equation 3 presents the MR model to predict the electricity usage of Keyworth Centre.

$$\beta = 1572 + 139.75x_1 - 4.72x_2 \tag{3}$$

Using this equation, the daily electricity consumption was predicted for year 2012. Figure 9 shows the distribution of residual plots. Mathematically, standardized residuals are represented as:

$$standardized\ residuals = \frac{observed\ count - expected\ count}{\sqrt{expected count}} \tag{4}$$

It is apparent that the residuals are split into two clusters. These two clusters represent two distinct types of days, i.e., working and non-working days. It is also clear that the distribution of residual is almost symmetric about the reference line, which shows the suitability of MR model. The scatter plot of forecasted and real electricity usage is shown in Figure 10. Again, the two clusters for working and non-working days have appeared here. The MR model predicts daily electricity consumption with a Mean Relative Error (MRE) of 4.59% for working days and 2.81% for non-working days.

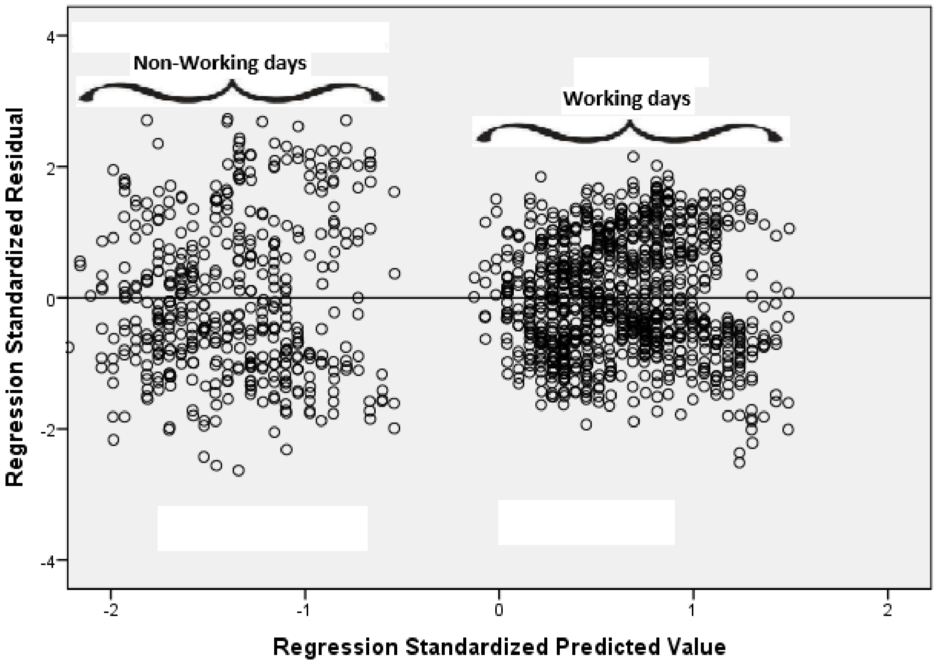


**Figure 9.** Residual plots for regression predicted values.

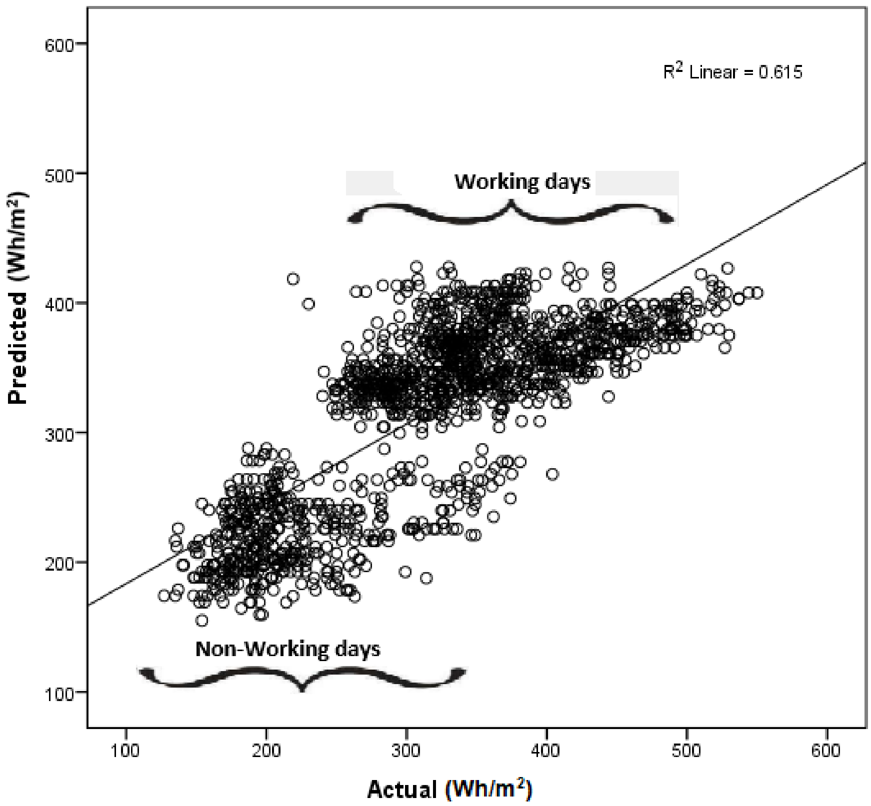


**Figure 10.** Scatter plot of actual and MR model predicted values.

*3.2. Results of artificial neural network*

In this study, 70% data has been used for training the model whereas 30% is used for its testing. A sigmoid activation function has been applied which bounds the output values in the range from 0 to 1. The study also includes a double layered feed-forward architecture along with linear output layers and sigmoid hidden layers. Two hidden layers with 50 neurons have been used as shown in Figure 11.

The plot of residuals is shown in Figure 12. Mathematically, residuals are given as:

$$residuals = observed\ value - predicted\ value \tag{5}$$

Residuals for WD and NWD are plotted against the predicted value. It is apparent that the distribution of residuals is not exactly symmetrical around the reference line. The residuals are within a range of −100 to 100 Wh/$m^2$ with few outliers.

Figure 13 shows a scattered plot between actual and ANN predicted daily electricity consumption. It can be seen that data is split into two clusters. A careful analysis reveals that these two clusters of data represent electricity usage for WD and NWD. Figure 13 also shows that model's predictive power is $R^2 = 0.64$.

Figure 14 shows the importance of five independent variables identified by the ANN model. It is apparent that the model recognizes WDI as the most important variable with temperature as the second most important variable. The least important variable is wind speed. Normalization is normally carried out to remove the variations in data due to different seasons in order to get reference line for comparison.

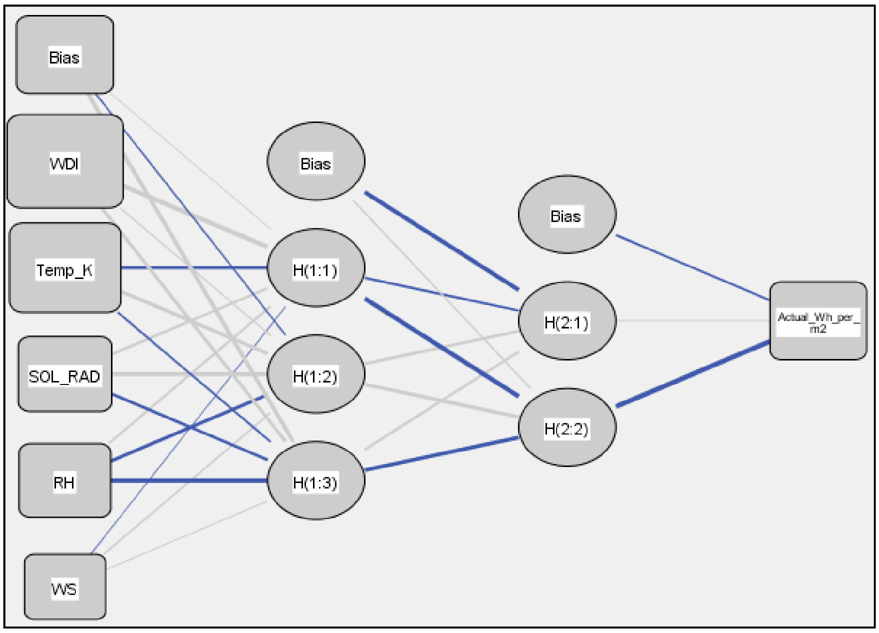


**Figure 11.** Structure of ANN model for KWC building.

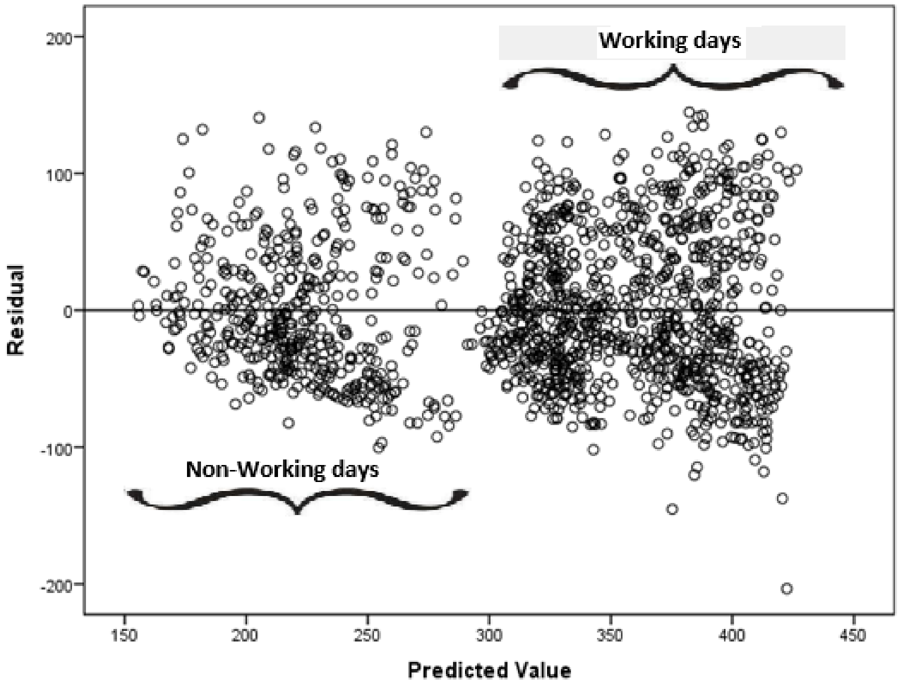


**Figure 12.** Residual plot for ANN predicted daily electricity usage.

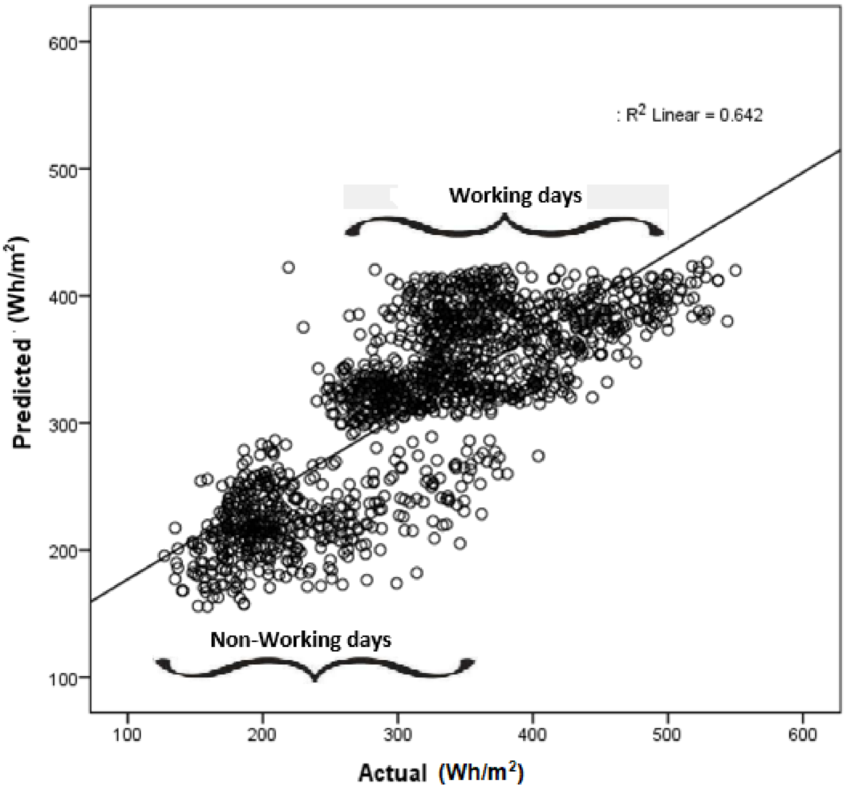


**Figure 13.** Actual Vs ANN predicted values of daily electricity consumption for KWC building.

### 3.3. Error analysis

The error analysis is performed for evaluating the proposed models using following three different error metrics:

(1) Mean Relative Error (MRE)

This error is a measure of difference of observed and predicted values. Mathematically, it is represented as:

$$\mathrm{MRE} = 1/N \sum_{i=1}^{N} \left| \frac{Zi - Z\prime i}{Zi} \right| \tag{6}$$

(2) Mean Absolute Percentage Error (MAPE)

This error is given by Eq 8:

$$MAPE(\%) = \frac{100}{N} \sum_{i=1}^{M} \left| \frac{Zi - Z\prime i}{Zi} \right| \tag{7}$$

where $N$ is sample size and $Zi$ = observed value, $Z'i$ = predicted value.

Predicted electricity consumption (by MR and ANN) have been compared with the actual electricity consumption for WD and NWD as shown in Figures 15 and 16 respectively. It is clear that both models predict electricity usage well during first six months of the year after which ANN predicts little better than MR and follows the profile of actual consumption displaying its higher accuracy during the last six months. Figure 17 displays a scatter plot of real and forecasted daily electricity consumption of Keyworth Centre building for working and non-working days for year 2012.

The performance of two modelling techniques has been evaluated using Mean Relative Error and Mean Absolute Percentage Error. The results are shown in Table 1. It is apparent that for both working and non-working days, ANN performs slightly better than MR with an average MAPE of 4.59%, i.e., about 0.45% less than MR for non-working days, and 2.44% for working days whereas MR displays a MAPE of 3.34%.

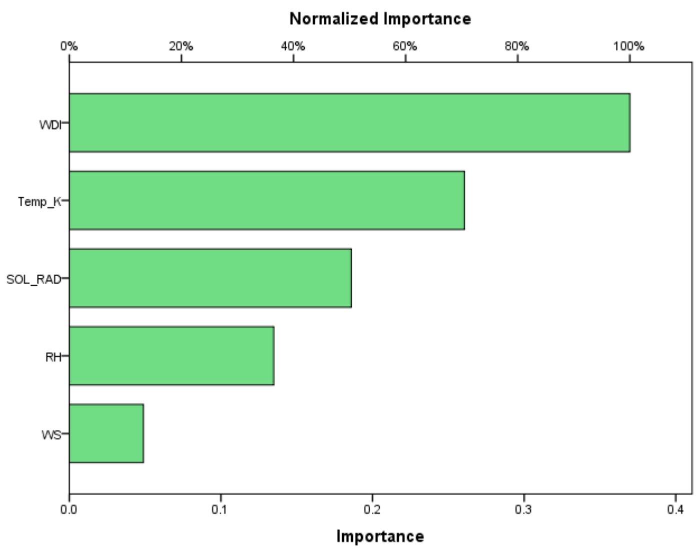


**Figure 14.** Importance level of five different independent variables identified by ANN model.

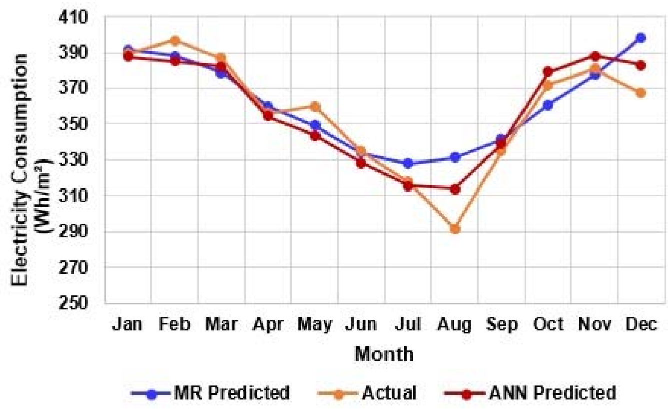


**Figure 15.** Comparison of real Vs forecasted daily electricity usage on WD.

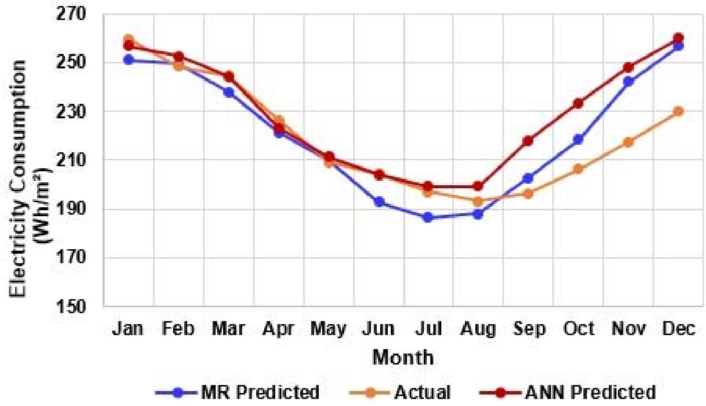


**Figure 16.** Comparison of real Vs forecasted daily electricity usage on NWD.

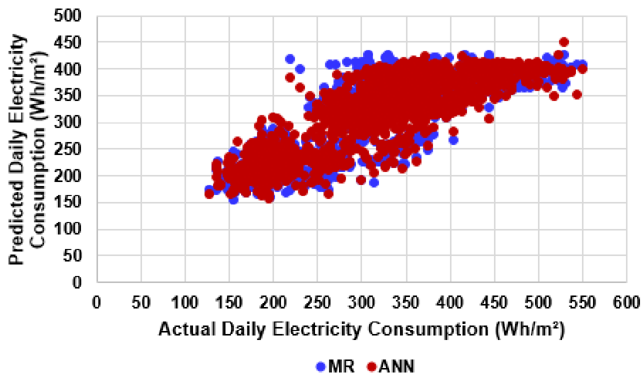


**Figure 17.** Comparison of real Vs forecasted daily electricity usage.

The complete error analysis is summarized in Table 1. It shows that both modelling techniques perform reasonably with ANN outperforming MR by a little margin for both working and non-working days. Moreover, such tools are helpful for policy and decision makers as these models facilitate in setting up their organizations' financial priorities. Energy managers could also use these for the production of their prediction studies.

**Table 1.** Error matrix for the two models.

| | Error matrix | MR (%) | ANN (%) |
|---|---|---|---|
| Non-working days | MRE | 5.04 | 4.59 |
| | MAPE | 5.12 | 4.59 |
| Working days | MRE | 3.09 | 2.81 |
| | MAPE | 3.34 | 2.44 |

### *3.4. Limitations*

Some of the limitations of this study are summarized below:

(1) Use of proxy variable for occupancy determination in the building.
(2) Models are applicable till the operating conditions remain same.
(3) The models lack validation test for similar types of buildings due to unavailability of the data.

## 4. Conclusions

In this study, two data-driven energy forecasting techniques, i.e., ANN and MR have been used to estimate daily electricity usage of an academic building of London South Bank University situated in London, United Kingdom. The historical data of electricity usage and five different climate factors were available for years 2007 to 2011. The study reveals that academic building energy consumption is strongly dependent on Week Day Index which is a proxy variable for the occupancy of the building and on Dry Bulb Temperature. Results also show that ANN performs better than MR with MAPE for working and non-working days 2.44% and 4.59% respectively. In a nutshell, both modelling techniques performed reasonably well with ANN outperforming MR with a little margin. The study lays foundation for prediction studies of other building categories and has the potential to help energy managers to accurately predict the energy usage pattern of the buildings. Secondly, as long as the building operates in the same manner under the same operating schedule the predicted results are free of time horizon.

## Acknowledgements

Authors are cordially thankful to the Energy Manager, Mr. Anuj Saush at LSBU for his cooperation during the data collection stage of this project.

## Conflict of interest

The authors declare no conflict of interest.

## References

1. Gul MS, Patidar S (2015) Understanding the energy consumption and occupancy of a multi-purpose academic building. *Energy Build* 87: 155–165.
2. Amber KP, Aslam MW, Mahmood A, et al. (2017) Energy consumption forecasting for university sector buildings. *Energies* 10: 1579.

3. Amasyali K, El-Gohary NM (2018) A review of data-driven building energy consumption prediction studies. *Renewable Sustainable Energy Rev* 81: 1192–1205.
4. Ghedamsi R, Settou N, Gouareh A, et al. (2016) Modeling and forecasting energy consumption for residential buildings in Algeria using bottom-up approach. *Energy Build* 121: 309–317.
5. Kaytez F, Taplamacioglu MC, Cam E, et al. (2015) Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines. *Int J Electr Power Energy Syst* 67: 431–438.
6. Li C, Ding Z, Zhao D, et al. (2017) Building energy consumption prediction: An extreme deep learning approach. *Energies* 10: 1525.
7. Lü X, Lu T, Kibert CJ, et al. (2015) Modeling and forecasting energy consumption for heterogeneous buildings using a physical–statistical approach. *Appl Energy* 144: 261–275.
8. Zeyu W, Ravi SS (2017) A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. *Renewable Sustainable Energy Rev* 75: 796–808.
9. Zhao HX, Magoulès F (2012) A review on the prediction of building energy consumption. *Renewable Sustainable Energy Rev* 16: 3586–3592.
10. Tardioli G, Kerrigan R, Oates M, et al. (2015) Data driven approaches for prediction of building energy consumption at urban level. *Energy Procedia* 78: 3378–3383.
11. Jin H, Zhang L, Ma HZ, et al. (2017) Machine learning for complex EMI prediction, optimization. *2017 IEEE Electrical Design of Advanced Packaging and Systems Symposium (EDAPS)*, Haining, 1–3.
12. Yedra RM, Díaz FR, Del-Mar-Castilla-Nieto M, et al. (2014). A neural network model for energy consumption prediction of CIESOL bioclimatic building. *In: International Joint Conference SOCO'13-CISIS'13-ICEUTE'13. Advances in Intelligent Systems and Computing*, 239. Springer, Cham.
13. Biswas MR, Robinson MD, Fumo N (2016) Prediction of residential building energy consumption: a neural network approach. *Energy* 117: 84–92.
14. Ahmad T, Chen H, Guo Y, et al. (2018) A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review. *Energy Build* 165: 301–320.
15. Amber KP, Aslam MW, Hussain SK (2015) Electricity consumption forecasting models for administration buildings of the UK higher education sector. *Energy Build* 90: 127–136.
16. Martellotta F, Ayr U, Stefanizzi P, et al. (2017) On the use of artificial neural networks to model household energy consumptions. *Energy Procedia* 126: 250–257.
17. Deb C, Lee SE, Santamouris M (2018) Using artificial neural networks to assess HVAC related energy saving in retrofitted office buildings. *Sol Energy* 163: 32–44.
18. Ekici BB, Aksoy UT (2009) Prediction of building energy consumption by using artificial neural networks. *Adv Eng Software* 40: 356–362.
19. Deb C, Eang LS, Yang J, et al. (2016) Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks. *Energy Build* 121: 284–297.
20. Park SK, Moon HJ, Min KC, et al. (2018) Application of a multiple linear regression and an artificial neural network model for the heating performance analysis and hourly prediction of a large-scale ground source heat pump system. *Energy Build* 165: 206–215.
21. Aranda A, Ferreira G, Mainar-Toledo MD, et al. (2012) Multiple regression models to predict the annual energy consumption in the Spanish banking sector. *Energy Build* 49: 380–387.

22. Bianco V, Oronzio M, Sergio N (2009) Electricity consumption forecasting in Italy using linear regression models. *Energy* 34: 1413–1421.
23. Bingchun L, Chuanchuan F, Arlene B, et al. (2017) Forecasting of chinese primary energy consumption in 2021 with GRU artificial neural network. *Energies* 10: 1453–1467.
24. Joseph CL, Kevin KWW, Dalong L, et al. (2010) Multiple regression models for energy use in air-conditioned office buildings in different climates. *Energy Convers Manage* 51: 2692–2697.
25. Mohammad M, Atefeh M, Shideh SA, et al. (2015) Multi-linear regression models to predict the annual energy consumption of an office building with different shapes. *Procedia Eng* 118: 622–629.
26. Amber KP, Ahmad R, Aslam MW, et al. (2018) Intelligent techniques for forecasting electricity consumption of buildings. *Energy* 157: 886–893.
27. Alireza K, Reza AR, Dominic M (1998) ANNSTLF-Artificial neural network short-term load forecaster-generation three. *IEEE Trans Power Syst* 13: 1413–1422.
28. Ferlito S, Atrigna M, Graditi G, et al. (2015) Predictive models for building's energy consumption: an Artificial Neural Network (ANN) approach. *In XVIII AISEM Annual Conference.*
29. Gamze O, Omer FD, Selim Z (2012) Forecasting electricity consumption with neural networks and support vector regression. *Procedia—Soc Behav Sci* 58: 1576–1585.
30. Hawkins D, Hong SM, Raslan R, et al. (2012) Determinants of energy use in UK higher education buildings using statistical and artificial neural network methods. *Int J Sustainable Built Environ* 1: 50–63.
31. Samani BH, Jafari HH, Zareiforoush H (2017) Artificial neural networks, genetic algorithm and response surface methods: The energy consumption of food and beverage industries in Iran. *J AI Data Min* 5: 79–88.
32. Erba S, Causone F, Armani R (2017) The effect of weather datasets on building energy simulation outputs. *Energy Procedia* 134: 545–554.
33. Fikru MG, Gautier L (2015) The impact of weather variation on energy consumption in residential houses. *Appl Energy* 144: 19–30.
34. Kaufmann RK, Gopal S, Tang X, et al. (2013) Revisiting the weather effect on energy consumption: Implications for the impact of climate change. *Energy Policy* 62: 1377–1384.
35. Kim YS, Srebric J (2017) Impact of occupancy rates on the building electricity consumption in commercial buildings. *Energy Build* 138: 591–600.
36. Ioannidis D, Tropios P, Krinidis S, et al. (2016) Occupancy driven building performance assessment. *J Innovation Digital Ecosyst* 3: 57–69.