Energy and AI 5 (2021) 100093

Contents lists available at ScienceDirect

Energy and AI

journal homepage: www.sciencedirect.com/journal/energy-and-ai

Prediction of office building electricity demand using artificial neural network by splitting the time horizon for different occupancy rates

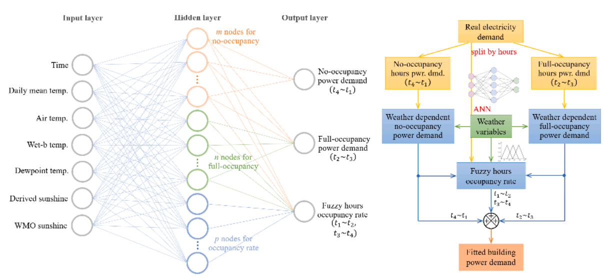
Si Chen [a], Yaxing Ren [a], Daniel Friedrich [b], Zhibin Yu [a,*], James Yu [c]

[a] James Watt School of Engineering, the University of Glasgow, Glasgow G12 8QQ, UK
[b] School of Engineering, The University of Edinburgh, Edinburgh EH9 3FB, UK
[c] SP Distribution PLC, Glasgow G72 0HT, UK

HIGHLIGHTS

- Two approaches proposed to predict the electricity demand of buildings according to time series and weather variables.
- Split the electricity demand data by time horizon for different occupancy rates.
- Use ANN to train the no-occupancy power demand, full-occupancy power demand and occupancy rates.
- Proposed approaches are validated in a case study of predicting the electricity demand of buildings in a university campus.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Building energy
Electricity demand prediction
Statistical modelling
Artificial neural network
Occupancy rate

ABSTRACT

Due to the impact of occupants' activities in buildings, the relationship between electricity demand and ambient temperature will show different trends in the long-term and short-term, which show seasonal variation and hourly variation, respectively. This makes it difficult for conventional data fitting methods to accurately predict the long-term and short-term power demand of buildings at the same time. In order to solve this problem, this paper proposes two approaches for fitting and predicting the electricity demand of office buildings. The first proposed approach splits the electricity demand data into fixed time periods, containing working hours and non-working hours, to reduce the impact of occupants' activities. After finding the most sensitive weather variable to non-working hour electricity demand, the building baseload and occupant activities can be predicted separately. The second proposed approach uses the artificial neural network (ANN) and fuzzy logic techniques to fit the building baseload, peak load, and occupancy rate with multi-variables of weather variables. In this approach, the power demand data is split into a narrower time range as no-occupancy hours, full-occupancy hours, and fuzzy hours between them, in which the occupancy rate is varying depending on the time and weather variables. The proposed approaches are verified by the real data from the University of Glasgow as a case study. The simulation results show that, compared with the traditional ANN method, both proposed approaches have less root-mean-square-error (RMSE) in predicting electricity demand. In addition, the proposed working and non-working hour based regression approach reduces the average RMSE by 35%, while the ANN with fuzzy hours based approach reduces the average RMSE by 42%, comparing with the traditional power demand prediction method. In addition, the second proposed approach can provide more information for building energy management,

* Corresponding author.
E-mail address: Zhibin.Yu@glasgow.ac.uk (Z. Yu).

https://doi.org/10.1016/j.egyai.2021.100093
Received 15 April 2021; Received in revised form 26 May 2021; Accepted 31 May 2021
Available online 5 June 2021

including the predicted baseload, peak load, and occupancy rate, without requiring additional building parameters.

1. Introduction

Due to the environmental degradation and global warming, many countries aim to develop low-carbon technologies to reduce the consumption of fossil fuels and greenhouse gas emissions [1, 2]. In terms of energy consumption, the proportion of building energy consumption in global energy consumption has increased rapidly [3] and the rate reaches approximate 40% of the total energy consumption in Europe [4]. Therefore, the energy consumption prediction for buildings is essential for designing high-efficiency buildings and maintaining low-energy operation and optimal control of buildings. Accurately predicting the energy consumption of buildings can provide benchmarks for energy management of building systems [5] and show the energy-saving potential of buildings [6].

The energy consumption of a building is affected by many factors, including weather variables, especially the dry-bulb temperature, the structure of the building and the thermal properties of physical materials used, occupancy and human behaviour, and secondary components, such as lighting system [4]. In the literature, many studies have used engineering methods [7, 8] or statistical methods [9, 10] to model and predict the weather-related thermal energy requirements of buildings. However, due to the influence of occupants' activities, predicting the electricity demand is more complicated than predicting the heat demand. In the literature, different prediction methods are used for the thermal load and electrical load of various buildings. The heat load model is usually based on regression analysis, and the power load model is usually based on the probability distribution of hourly daytime analysis [11]. Statistical models can be used to study the effects of temperature, and time series models can be used to predict daily power demand [12].

The data of building energy consumption shows that the basic electricity consumption of buildings includes emergency lighting, service and safety electricity, which are basically constant throughout the day [6]; the variable electricity consumption of buildings includes heating/cooling loads, household air-conditioning equipment, hot water, and other power consumption [4], which are obviously affected by weather variables [12]. In most buildings, the HVAC (heating, ventilation, and air conditioning) systems consume the most electric energy, which can provide a sense of comfort for the working space of the building [13]. However, the simple understanding of occupancy in current research has led to a huge performance discrepancy between estimated energy consumption and measured energy consumption [14, 15]. Due to the overall increase in per capita building area, power consumption indicators based on building area are no longer suitable for predicting the energy demand of buildings [6]. In [16], Newsham and Birt especially emphasized the influence of occupancy rate, which obviously can improve the accuracy of the model. In the modelling of occupancy rate, it is difficult to collect information about equipment occupancy and operation and, thus, time indicators are usually selected as input related to the timetable to indicate occupancy and equipment usage [17]. In order to solve the influence of seasonal changes and human activities on the power demand forecast of buildings, some studies have established separate models for different seasons or months to predict the power demand including human activities [18, 19].

The research in [20] aims to study a short-term, real-time energy demand forecasting method to cope with changing loads to effectively operate and manage buildings. In some buildings with complex application functions, such as hotels and shopping centres, the randomness of human activities is relatively high, which can greatly reduce the reliability of data and the accuracy of predicting the building energy consumption [21]. In addition, some researchers use data mining techniques to discover and summarise electricity consumption patterns hidden in the data [22]. Amasyali and El-Gohary reviewed the development of data-driven building energy consumption models in existing research using machine learning algorithms, including support vector machines and artificial neural networks (ANN), decision trees and other statistical algorithms, and highlighted future research directions [23]. Nizami and Al-Garni tried a simple feedforward neural network to correlate power consumption with the number of residents and weather data [24]. In literature [25], Massana et al. proposed a support vector regression model that uses the temperature and occupancy rate of buildings to predict the electrical load of non-residential buildings. Paterakis et al. proposed a framework based on deep learning to predict electricity demand by taking care of long-term historical dependence [26]. Ahamd et al. proposed and evaluated a novel random neural network technology, which can predict the energy consumption of non-residential buildings [27]. Zeng et al. conducted a comparative study on four data-driven methods used in online building energy consumption prediction and proved that the ANN method has better accuracy for energy consumption prediction [5]. Luo et al. proposed a deep feedforward neural network architecture determined by genetic algorithm for the day-ahead hourly and week-ahead daily power consumption of campus buildings in the UK [17]. Rahman proposed a recurrent neural network model, which can predict mid-to-long-term (more than one year) electricity consumption in commercial and residential buildings with a resolution of 1 hour [28]. Wei et al. proposed an occupancy estimation method based on blind system identification (BSI), and estimated the number of occupants based on artificial neural networks and using BSI and developed and reported a power consumption prediction model for air conditioning systems [14].

These methods are suitable for long-term forecasting of average daily electricity demand to reduce the impact of personnel activities on forecast accuracy, or for short-term forecasting of electricity demand to reduce seasonal effects. However, these methods are not suitable for solving some typical cases in specific area, such as office/education buildings heated by electricity in cold areas. In this type of buildings, the heating load is much higher than the cooling load, and the long-term and short-term correlation between electricity demand and temperature are in opposite direction. That is, in the trend of seasonal power consumption, temperature and power demand are in negative correlation, while in the trend of hourly power consumption, temperature and power demand are in positive correlation. This causes difficulties in data fitting between power demand and weather variables. To our best knowledge, no method can solve this problem effectively at present. Therefore, this paper aims to seek a simple and reliable method to accurately predict the electricity demand of buildings in seasonal and hourly simultaneously by considering the impact of occupancy rates.

This paper proposed two approaches to predict the electricity demand of target buildings. In the first prediction approach, the model uses the non-working hours power demand data to predict the building baseload depending on ambient temperature and uses the working hours power demand data to predict the occupants' activities power demand, which is only dependent on the time series in a day. After that, this paper proposed a method that combines ANN and fuzzy logic technologies to estimate the building baseload, peak load, and the real-time occupancy rate from multi-variables weather variables. The proposed prediction method can predict the hourly power demand of target buildings based on the prediction results of baseload and occupants' activities from ANN. This method was applied in the University of Glasgow campus as a case study. The results show that the proposed ANN-based method can significantly reduce the prediction error, improve the prediction accuracy of the electrical demand of the target buildings and the entire campus.

2. Problems in electricity demand data fitting

In the data processing of the electric power and weather variables, it is found that the relationship of power consumption and ambient temperature are in negative and positive correlation in long-term (seasonal variation) and short-term (hourly variation), respectively. The comparison between ambient temperature and electricity demand is given in Fig. 1. The total data of 1.5 years on the left shows that the relationship between ambient temperature and electricity demand is in negative correlation. The lower air temperature (blue curve), the higher electricity demand (red curve). But if the comparison is enlarged into single days, as shown in the figure on the right, their relationship is positive correlation. The electricity demand is increasing and decreasing following the temperature.

The main reason of this situation can be explained as the impact of human activities of occupants in the building. For example, during working time, occupants are using electric devices with the demand in an approximate Gaussian distribution. Thus, the power demand at noon is the highest while the power demand at night is the lowest. This tendency of power demand in a day matches the tendency of temperature, but it cannot say that the power demand is determined by the temperature.

In a long-term period or seasonal variation, the HVAC (heating, ventilation, and air conditioning) system is managed to heat the building continuously. Thus, the power demand is in inverse correlation with the temperature, or the lower temperature causes the higher power consumption for space heating. It shows obvious inverse correlation between electricity demand and temperature.

This is a typical issue in office/education buildings heated by electricity. It is more obvious in cold areas, such as in Scotland, that the heating load is much higher than the cooling load. In the opposite condition where the cooling load is much higher, the inverse correlation between electricity demand and temperature in the long term is not as obvious as the issue shown above. This paper focuses only on fitting the data of electricity demand of buildings where the heating demand is much higher than cooling. Therefore, in a longer time period, the temperature and electricity demand are in inverse correlation, while in a shorter time period, the temperature and electricity demand are in positive correlation. That causes difficulties in data fitting using traditional statistical methods of fitting the power demand with temperature directly.

3. Development of electricity demand prediction approach

3.1. Working hour splitting based regression approach (Approach 1)

The approach is designed to split the building baseload power consumption and occupants' activities by time series. In order to achieve this, it is assumed that the occupants only consume power at particular period of time while building baseload power is consumed 24 h continuously. In the time period when the occupancy rate is low or zero, the power demand is mainly the baseload of the building; and in the time period when the occupancy rate is high, the power demand data includes both the building baseload and occupants' activities. It is difficult to separate them from the recorded data. Therefore, the present approach uses the power demand data of the low occupancy time period to find the dependence on environmental conditions to adapt to the building baseload, which is independent with occupants' activities. After that, the building baseload in the remaining time period can be predicted by the weather variables of the same period and the fitted dependence above. It is assumed that the difference between the actual power demand and the building baseload during this period is caused by the influence of the occupants' activities.

In the case study, the target buildings at university campus are used for teaching and officing. Thus, the normal working hours are between 9:00 and 17:00, depending on the work schedule. However, the actual working hours are flexible for employees, depending on their preference on working time. After comparing the electricity demand data of normal working days with public holidays, it is found that the impact period of occupant's power consumption is several hours wider than the normal working hours. Therefore, in order to further reduce the uncertainty of occupants' activities, a three-hour redundancy period has been added before and after the normal working hours to eliminate possible occupants' activities during non-working hours. The data between 20:00 every day and 6:00 the next day is defined as the non-working hours and used to fit weather variables to electricity consumption without occupants' activities, as shown in Fig. 2.

In the approach, the non-working hours power demand data is used to fit the most significant weather variables as the building baseload power demand. To choose the most significant weather variable for fitting the non-working hours power demand, the sensitivity analysis technique is used to find which weather variable has the highest sensitivity to the power demand. The sensitivity analysis uses the coefficient of determination, R_k^2, of the kth model input variable as the index of showing the quality of each weather variable to the electricity demand as

$$R_k^2 = 1 - \frac{\sum_i (y_i - x_{k,i})^2}{\sum_i \left(x_{k,i} - \bar{x}_k\right)^2} \quad (1)$$

where $x_{k,i}$ and y_i indicate the ith sample points data of the kth model input and model output, respectively; $\bar{x}_k$ indicates the mean of model input.

After the weather variable with highest sensitive to power demand is found, their relationship is found using the simplest statistical method (i.

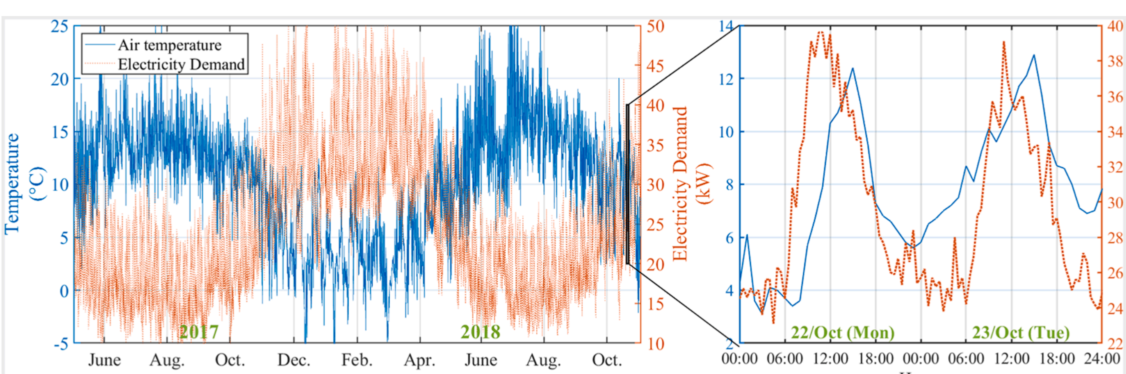


Fig. 1. Half-hourly power demand of different campuses comparing with temperature.

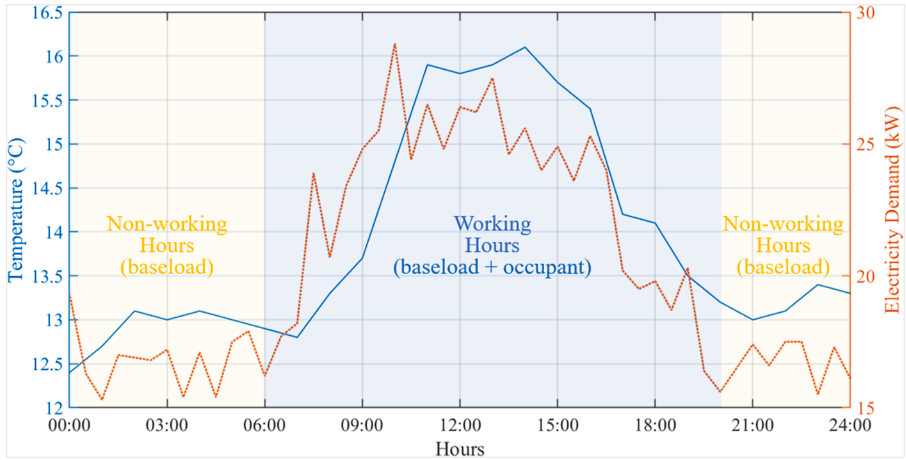


Fig. 2. Splitting the time into working hours and non-working hours of every 24 h (example of a single day).

e., the linear regression method) to predict the building baseload power demand. The difference between the recorded real power demand and fitted baseload power is then known as the power demand determined by occupants' activities. It is assumed that the power demand of occupants' activities in each hour satisfies a normal distribution. Therefore, at the same hour on different dates, the mean value of the difference between the recorded power demand and the fitted building baseload is used to predict the most possible power demand caused by occupants' activities. After getting the average of human activities power demand caused by building occupants, the sum of power demand caused by fitted building baseload and occupant activities are known as the fitted building power demand. The whole process is produced as the flowchart in Fig. 3.

3.2. ANN with fuzzy hours splitting approach (Approach 2)

The design above is the basic method to verify the approach of splitting the data by time of working hours. However, the linear regression method to fit the electricity demand and ambient temperature cannot fully capture the nonlinear dynamics of multi weather variables. In addition, the average occupant power demand ignored a lot of information that causes the power variation of different human activities. In order to improve the power demand prediction approach, an approach is developed with artificial intelligent technologies.

In the original design, the splitting method is based on a fixed time

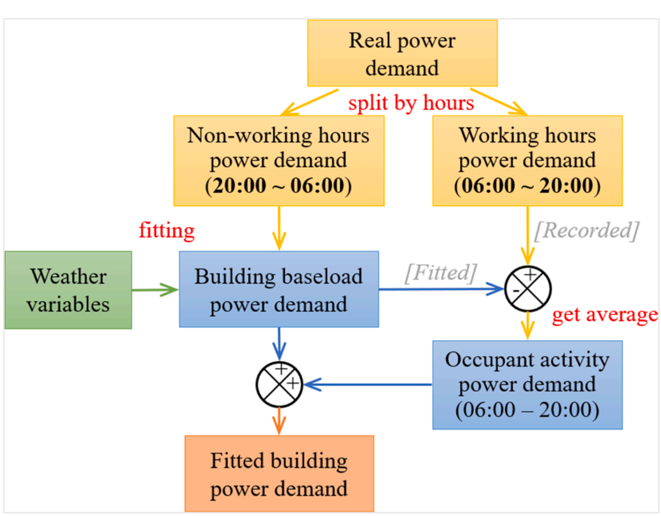


Fig. 3. Flow chart of electricity demand fitting approach.

from normal working hours. However, this splitting approach is too arbitrary and does not consider the possible changes of different conditions. In the new approach, the full-occupancy hours are defined from t_2 to t_3 while the no-occupancy hours are defined from t_4 to t_1. The area between full-occupancy hours and no-occupancy hours are defined as the fuzzy area that is not quite clear to belong to the full-occupancy or no-occupancy, as shown in Fig. 4.

The power demand can be split into three sections by time as

$$P_{\text{building}}(t) = \begin{cases} P_{\text{no}} & t_4 \le t \le t_1 \\ P_{\text{fuzzy}} & t_1 \le t \le t_2 \& t_3 \le t \le t_4 \\ P_{\text{full}} & t_2 \le t \le t_3 \end{cases} \tag{2}$$

Assume the membership function of the fuzzy area is $f(t)$, which indicates the occupancy rate. The power demand of the fuzzy hours $P_{\text{fuzzy}}(t)$ can be obtained from the power demand of no-occupancy power, the full-occupancy power and the membership function as

$$P_{\text{fuzzy}}(t) = P_{\text{no}}(t)\cdot(1 - f(t)) + P_{\text{full}}(t)\cdot f(t) \tag{3}$$

where $P_{\text{no}}(t)$ is the power demand of time during no-occupancy hours, which also indicates the building baseload power demand without the impact of the occupants. $P_{\text{full}}(t)$ is the power demand of time during full-occupancy hours, which includes the power demand at peak load caused by full-occupancy activities.

In the fitting approach between weather variables and power demand, the working hour splitting based regression approach uses linear regression to find the relationship between non-working hours power demand and the most sensitive weather variable. That approach assumes that the demand of building baseload is in linear relation to a single weather variable. The impact of other less sensitive weather

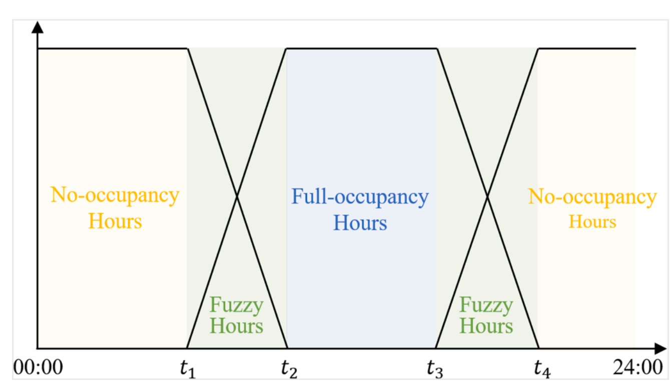


Fig. 4. Splitting the time into full-occupancy hours, no-occupancy hours, and fuzzy hours.

variables and other less dominate nonlinear relationships are ignored. Therefore, in the new approach, the artificial neural network (ANN) method is used to find the nonlinear relationship from multiple weather variables to power demand of no-occupancy hours and full-occupancy hours, respectively. Similar with the regression approach, the sensitivity analysis referring to the coefficient of determination is required to rank the weather variables based on their sensitivity. But different with the regression approach above that only choose one most significant variable, the ANN based approach chooses more weather variables for fitting their nonlinear relationships with power demand. Based on the result of sensitivity analysis, the number of input neurons in ANN can be defined according to the number of weather variables with high sensitivity to the power demand.

The process of ANN can be described in mathematical formulas. Define $x_k(k = 1, 2, \ldots, n)$ as the k-th input attribute value, which is passed along the links to the other layers. The weighted sum of signals, $\sum$, arriving at the input of the next neuron is subjected to a transfer function, which is the most commonly used 'sigmoid' function as

$$f\left(\sum\right) = \frac{1}{1 + e^{-\sum}} \tag{4}$$

The jth hidden neuron $h_j(j = 1, 2, \ldots, p)$ receives the sum of neuron value multiplied by the weights $w_{kj}^{(2)}$ and bias $b_{kj}^{(2)}$ associated with the link as

$$h_j = f\left(\sum_{k=1}^{n} w_{kj}^{(2)} x_k + b_{kj}^{(2)}\right) \tag{5}$$

The output neurons are defined as $y_i(i = 1, 2, \ldots, m)$, which are summed up with their input signals and activation transfer function as

$$y_i = f\left(\sum_{j=1}^{p} w_{ji}^{(1)} f\left(\sum_{k=1}^{n} w_{kj}^{(2)} x_k + b_{kj}^{(2)}\right) + b_{ji}^{(1)}\right) \tag{6}$$

where $f(\cdot)$ is the activation function, the sigmoid function used in the paper; $w_{ji}^{(1)}$, $b_{ji}^{(1)}$, $w_{kj}^{(2)}$ and $b_{kj}^{(2)}$ are the weights and bias linked to the output layer (1) and hidden layer (2), respectively. This is a typical two-layer ANN model with an output layer and one hidden layer.

The training error is used to update the ANN parameters of weights and bias of each neurons in the hidden and output layers. The training based on backpropagation (BP) learning algorithm is adopted to a typical two-layer ANN model to search for the global optimum as

$$\delta_i^{(1)} = y_i(1 - y_i)(t_i - y_i) \tag{7}$$

$$\delta_j^{(2)} = h_j(1 - h_i)\sum_i \delta_i^{(1)} w_{ji} \tag{8}$$

where $\delta_i^{(1)}$ and $\delta_j^{(2)}$ indicate the responsibilities of output-layer neurons and hidden-layer neurons, respectively. Then the weights and bias of links can be updated based on the responsibilities as

$$w_{ji}^{(1)} = w_{ji}^{(1)} + \eta\delta_i^{(1)} h_j \tag{9}$$

$$w_{kj}^{(2)} = w_{kj}^{(2)} + \eta\delta_j^{(2)} x_k \tag{10}$$

$$b_{ji}^{(1)} = b_{ji}^{(1)} + \eta\delta_i^{(1)} \tag{11}$$

$$b_{kj}^{(2)} = b_{kj}^{(2)} + \eta\delta_j^{(2)} \tag{12}$$

where η is the learning rate of the BP based ANN.

The original design assumes that the power demand of occupants' activities only related with the working time and calculates the average occupants' power consumption by subtracting the fitted building baseload from the recorded full power demand. However, the occupant activities can somehow be affected by different weather variables. The new approach uses the ANN to find the occupancy rate in fuzzy hours, in which the number of occupants is varying from no occupants to full occupants. Thus, the occupancy rate of fuzzy hours is described by membership function between no-occupancy and full-occupancy and is fitted with the weather variables in ANN. Assuming the fuzzy hours are between t_1 and t_2, the fitted fuzzy area membership function can be obtained by

$$\bar{f}(t)\Big|_{t_1}^{t_2} = \min\left(\max\left(\frac{P_{\text{fuzzy}}(t) - \overline{P_{\text{no}}}(t)}{P_{\text{full}}(t) - \overline{P_{\text{no}}}(t)}, 0\right) 1\right) \tag{13}$$

where $P_{\text{fuzzy}}(t)$ is the real power demand of fuzzy hours for training, $\overline{P_{\text{no}}}(t)$ indicates the predicted baseload power demand in fuzzy hours using the ANN trained by no-occupancy hours data. $\overline{P_{\text{full}}}(t)$ indicates the predicted full occupants power demand in fuzzy hours. The power demand data of no-occupancy hours and full-occupancy hours is used to train the ANN model. After it is well trained, the model can be used to calculate the no-occupancy power demand and full-occupancy power demand. Therefore, in the fuzzy hours, its no-occupancy power demand and full-occupancy power demand can be predicted and used as the upper and lower limit. Its real value can be obtained using both the predicted limits and the fitted membership function of occupancy rate, $\bar{f}(t)|_{t_1}^{t_2}$, in fuzzy hours between t_1 and t_2. The fitted membership function shows the occupancy rate between 0 and 1. If the membership value is 0, no occupant activity affects the power demand, and the building power demand only includes the baseload power demand. If the membership value is 1, the power demand is impacted by full occupants. The membership value between 0 and 1 indicates the ratio of current occupants to the full occupants. Different with Approach 1 that assumes the occupancy rate is the same in all weekdays or weekends, the fitted occupancy rate in Approach 2 depends on both the weather variables and time horizon. Thus, Approach 2 covers the uncertainties of occupancy rate caused by weather and better predicts the real occupancy rate of target buildings.

In the ANN design, it combines three ANN together. The baseload power demand is fitted using m nodes and trained with the data of no-occupancy hours. The full-occupancy power demand is fitted using additional n_2 nodes and trained with the data of working hours as the full-occupancy power demand is calculated by both the n_1 nodes for baseload and n_2 nodes for occupants. The membership function of occupancy rate in fuzzy hours for training is calculated using Eq. (13) with the real power demand in fuzzy hours and the fitted full-occupancy power demand from $n_1 + n_2$ nodes. The membership function is fitted to time series and weather variables using another n_3 nodes. In summary, the ANN is developed to use the weather variables and time series as inputs and the power demand data split into no-occupancy hours, full-occupancy hours and fuzzy hours, as shown in Fig. 5.

The power demand of the fuzzy area ($t_1 \sim t_2$ and $t_3 \sim t_4$) is calculated according to the trained membership function of occupancy rate. Then the predicted power demand from 00:00 to 24:00 is compared with the recorded result to validate the approach.

The final artificial intelligent based approach is developed as the flowchart shown in Fig. 6. The real power demand data is split the full-occupancy hours ($t_2 \sim t_3$) and no-occupancy hours ($t_4 \sim t_1$) to fit with weather variables, respectively. The power demand data of fuzzy hours ($t_1 \sim t_2$ and $t_3 \sim t_4$) is then used to calculate the occupancy rate using Eq. (13) and fitted with weather variables as well. At last, the fitted building power demand is calculated from the predicted no-occupancy power, full-occupancy power and predicted occupancy rate from 00:00 to 24:00 from the ANN for validation.

4. Simulation result of electricity demand prediction

In the case study of electricity consumption of the University of

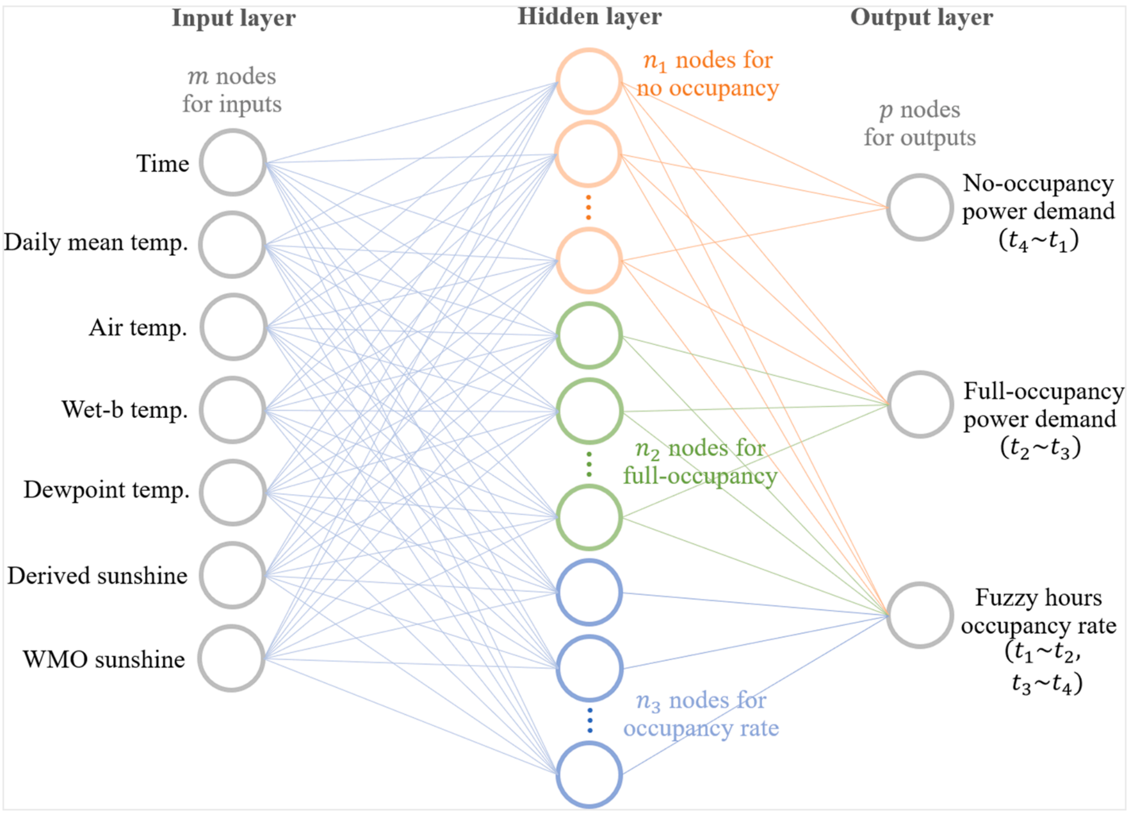


Fig. 5. Artificial neural network for fitting the weather variables and non-working hours and working hours power demand.

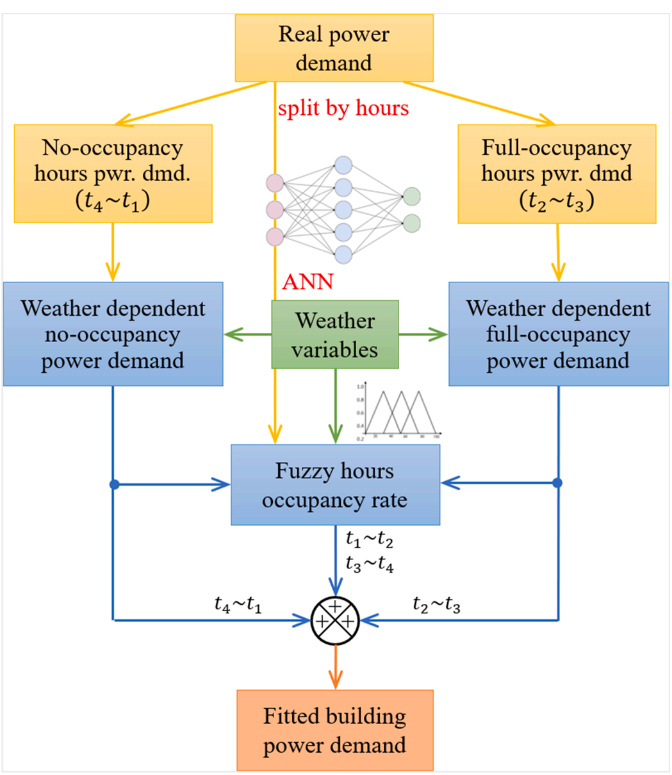


Fig. 6. Embed the developed electricity demand fitting approach with artificial intelligent techniques.

Glasgow, most buildings do not have individual power meters to record their electricity usage. The buildings with individual power metre to record their electricity demand are the Maths & Stats school, the St Andrews building, and the Wolfson medical school. Other buildings are included in the two campuses, the North campus and Main campus, as shown in Fig. 7. The Maths & Stats school building is used as the target building to develop the approach for electricity demand prediction.

4.1. Fitting of electricity demand

The data used in the case study included 13 weather variables recorded by the local weather station and the electricity consumption of whole campus recorded by the university energy centre. The weather variables used for fitting the whole hours power demand include the dry-bulb or air temperature, the wet-bulb temperature, the dew-point temperature, the daily mean temperature, the derived sunshine, the WMO sunshine (measured by World Meteorological Organization), the wind speed, the wind direction, the relative humidity, the station pressure, the mean sea-level pressure, the visibility, and the cloud base height.

In the sensitivity analysis, each weather variable is analysed with the electricity demand with R^2. The data is recorded in every hour for a total of 549 days from 1st May 2017 to 31st October 2018. Therefore, the full hours power demand including the total of 13,176 h data points and the non-working hours power demand including just 4941 h data point are used in the sensitivity analysis. The comparison between analysing full hours power demand and non-working hours power demand is shown in Fig. 8. In the result, the blue bar indicates the sensitivity of each weather variable to non-working hours power demand, where the air temperature is higher than other weather variables. The red bar indicates the sensitivity of each weather variable to 24 h power demand, where the daily mean temperature is obviously higher than other weather variables and the sensitivity of all weather variables is less than 0.5 in R^2. This verifies that the effectiveness of the proposed approach that the

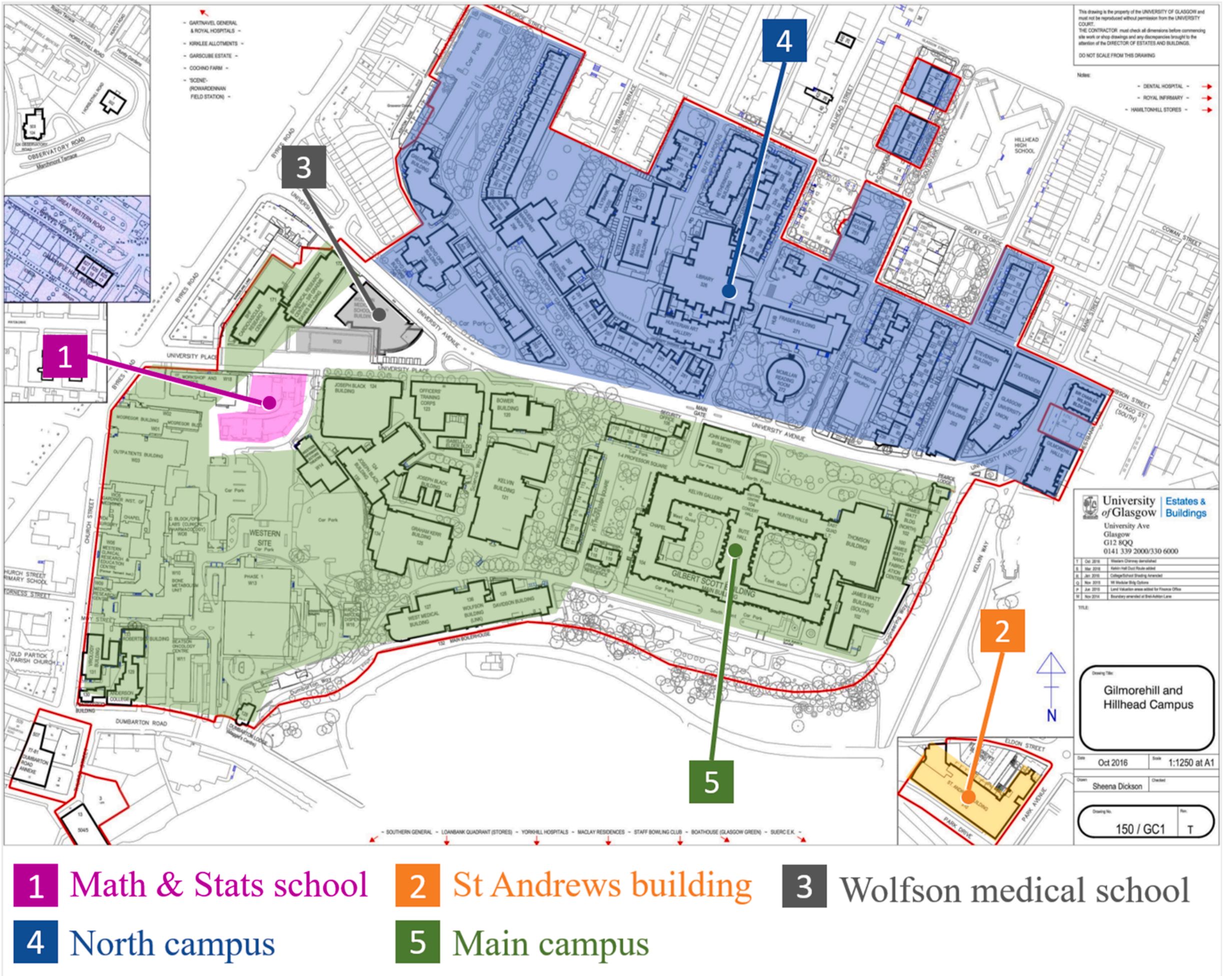


Fig. 7. Campuses of the University of Glasgow with individual electricity demand data.

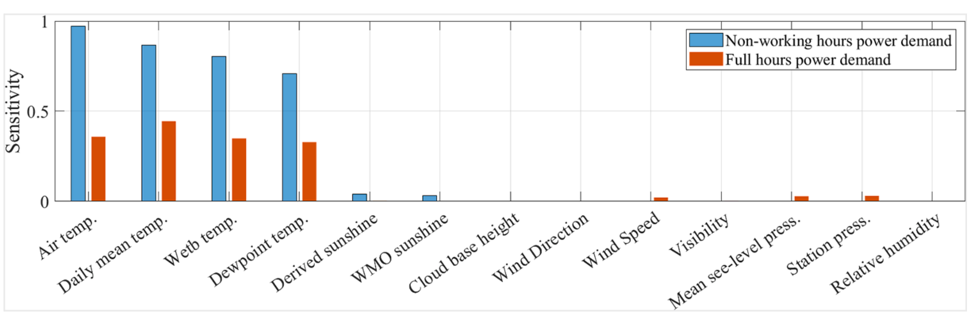


Fig. 8. Sensitivity analysis of each weather variable to non-working hour power demand.

non-working hours power demand has higher sensitivity to the electricity demand and the impact of occupant's activities in working hours can reduce the sensitivity and causes more uncertainties.

In the results, the air temperature has the highest sensitivity to the non-working hours power demand. Thus, the linear regression method is used to find the linear proportion between air temperature and non-working hours power demand. The fitted power demand with linear regression is defined as the building baseload power demand. The difference between real power and fitted baseload power is known as the occupant's activities and its average power demand can be obtained from the boxplot shown in Fig. 9. The occupant's activities are different between workdays and weekends/holidays. Therefore, the fitting of power difference is separated by workdays and weekends/holidays and getting their mean hourly power demand of occupants, respectively.

In the result of mean value, the peak power demand caused by occupants' activities in weekdays is about 15 kW while that on weekends is only approximate 6 kW. In this test, the power demand caused by human activities of occupants are assuming to be the same every workday or weekends. In the box plot, the red central segment indicates the median, the top and bottom edges of each box indicate the 1/4 and 3/4 percentiles, and the '+' symbol indicates the outliers. The uncertainties are obtained as the range between upper and lower limits to the median by ignoring the outliers. In the result, the average power demand caused by human activities of occupants has the uncertainties of up to ±10 kW in both working days and non-working days. For example, at the time of 12:00, the occupants' power demand in weekdays is 15±10 kW while

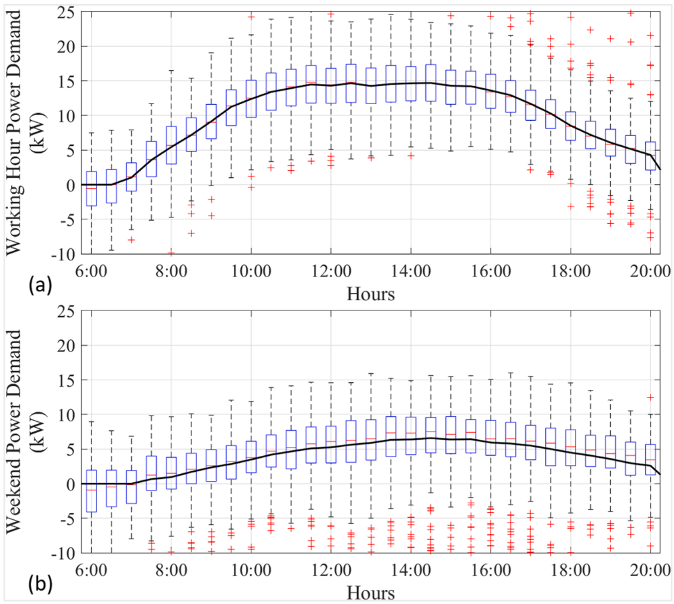


Fig. 9. Statistical mean value of working hours power demand caused by occupant activities in working days and weekends/holidays.

that in weekends is 5±10 kW. The negative power demand just presents the difference between the average fitted power demand and real value. This regression approach mainly considers the variation of occupancy with time, and its relation to weather variables is not considered in the regression for simplification. The variation of occupancy behaviour caused by weather variables is considered as uncertainties in this approach. But this approach has a limitation that the relationship between occupancy behaviour and weather variables is not considered and fitted.

After the fitting of both building baseload power demand and occupants power demand, their sum is known as the total power demand of the target building and it is then compared with the real power demand. Fig. 10 shows the results of the comparison between real power demand (blue solid line) and fitted power demand (red dashed line). Their difference is known as the fitting error as shown below. In the fitting results, the fitted power demand matches well with the recorded real power demand. The average fitting error is around 18%.

The fitted power demand is the sum of weather dependent baseline power and the power caused by the occupants' activities. The example of fitted power demand in one week is shown on the right of Fig. 10. The blue dashed line shows the temperature. The blue shadow shows the fitted baseload power, which is in negative correlation to temperature. The yellow shadow shows the fitted power caused by occupants' activities. The red line shows the final fitted power demand while the black pointed line shows the real power demand for comparison.

However, due to the power of the occupants' activities is fitted using the mean value, it causes more fitting error of the unpredictable human activities. Therefore, using ANN to find the relationship between the power demand caused by occupants and the weather variable in the new design is theoretically one solution to improve the prediction accuracy.

4.2. Fitting result of ANN based approach

In this section, the artificial intelligent based fitting approach is tested in fitting the power demand of target building/campus. As in the design, the ANN method is used to find the relationship between weather variables and power demand of no-occupancy and full-occupancy activities. In the previous case, the statistical results of power demand caused by occupancy show that the probability density function of occupancy within a day is approximately a bell-shaped curve. Therefore, in order to minimize the impact of the occupancy rate variation, an interval where the occupancy rate is as stable as possible should be selected near the peak and trough values of the occupancy rate. This will result in a narrower time interval. However, if the time interval is selected as narrow as possible, it will result in less data available for ANN training, which will reduce the prediction accuracy. Therefore, it is necessary to consider the trade-off between narrower time intervals and more data for ANN training in defining the full-occupancy and no-occupancy time period. In the ANN training, the full-occupancy hours are set as from t_2=11:00 to t_3=15:00. The no-occupancy hours are set as from t_4=00:00 to t_1=04:00. The switching between no-occupancy hours and full-occupancy hours is using the predicted occupancy rate from the ANN model.

The fitted membership function of occupancy rate is related to the weather variables and time horizon using the trained ANN. Fig. 11(a) shows fitted membership function for one week as an example. The result shows that the occupant's membership is the highest at noon and is the lowest at night when it is normally close to zero. The result of predicted power fitted by ANN is shown in Fig. 11(b). The blue shadow on the bottom shows the fitted no-occupancy power demand, which indicates the building baseload. The green dashed line shows the fitted full-occupancy power demand, which indicates the peak load. The yellow shadow in the middle shows the fitted power demand caused by occupants' activities from the predicted occupancy rate in Fig. 11(a).

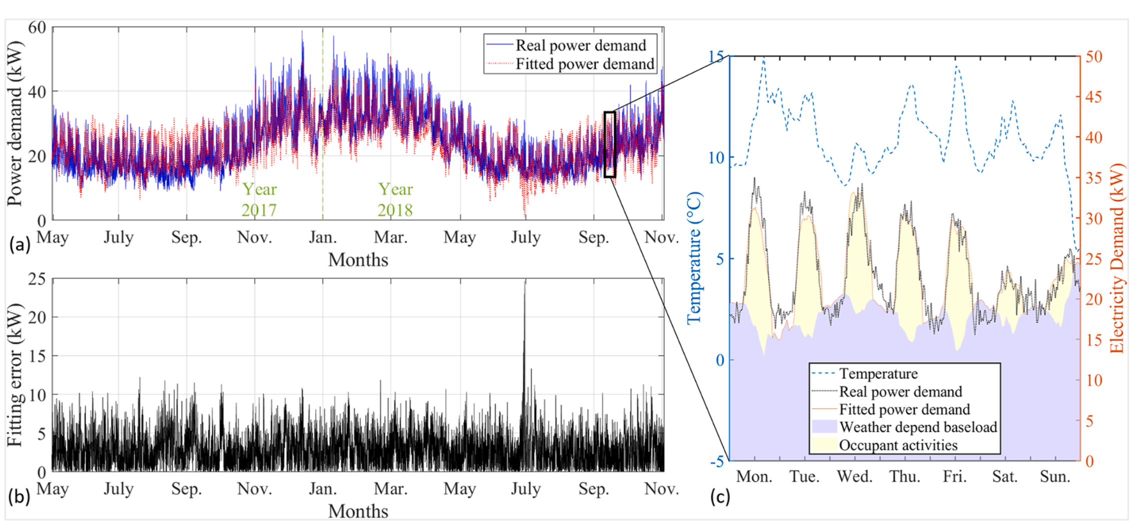


Fig. 10. Prediction result of fitted power demand (approach 1) comparing with real power demand.

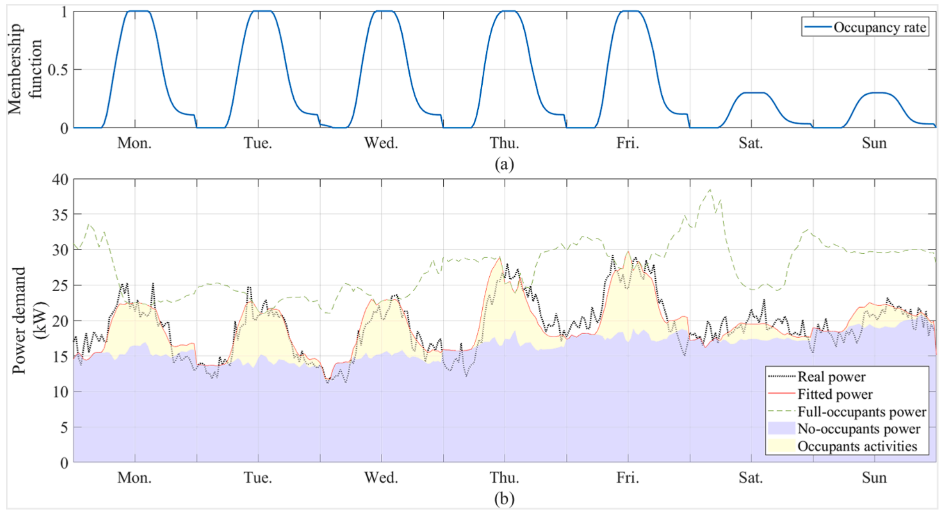


Fig. 11. Prediction result of proposed ANN based approach (approach 2) in one week as an example. (a) Fitted membership function of occupancy rate, (b) fitted power demand.

Combining the baseload power demand, the occupants' activities related power demand and occupancy membership function fitted from ANN using Eq. (13), the final fitted power demand is shown as the solid red line shown in Fig. 11(b). Comparing it with the real power demand shown as the dotted black line, the fitted power demand tracks the real power demand and is able to predict the future electricity demand.

In order to validate the effectiveness of the proposed approaches, the comparison among the linear regression-based approach (Approach 1), the ANN based approach (Approach 2) and the conventional ANN fitting approach in one week is shown in Fig. 12. The result of conventional ANN is shown with the dashed green line. Due to the issues mentioned in Section 2, the conventional ANN cannot fit the data because the long-term and short-term time horizons are negative and positive correlation with weather variables. Thus, the conventional ANN cannot find the best relationship between weather variables and target power demand. With splitting the data by different time periods of no-occupancy hours and full-occupancy hours, the proposed approaches can have better prediction performance and less absolute prediction error.

Using Approach 1, the non-working hour power demand is fitted with the linear regression of air temperature and working hour occupancy power demand is obtained from its average value. This test validates the effectiveness of splitting the data by working and non-working hour time periods. However, the linear regression cannot fully use the information of weather variable. In the ANN with fuzzy hours splitting approach (Approach 2), the no-occupancy hours power demand, the full-occupancy hours power demand and occupancy rate is fitted with weather variables by the ANN approach. The result shows that Approach 2 has less absolute prediction error than the linear regression-based approach and the conventional ANN.

In neural network technology, data is usually classified into training set and test set, usually in the ratio of 70:30, to guarantee the model accuracy. This case study contains the electricity demand data of 18 months in total, from May 2017 to October 2018. Therefore, the data of the first 12 months is used to train the ANN, and the data of the next 6 months is used for testing. In addition, the power demand data from December to February has the highest power demand in a year. As the training data includes this period, the power demand of other months can be guaranteed to be within the boundary of the model.

The power demand of the target building of Maths & Stat School for training and the predicted power demand is shown in Fig. 13(a). In addition to the Maths & Stat School, other university campuses have different dependency of power demand to weather variables and each campus has different capacity of rated power demand and occupant behaviour.

As described in Section 2, the Maths & Stat School, St Andrews Campus and Wolfson Medical School have their own individual power

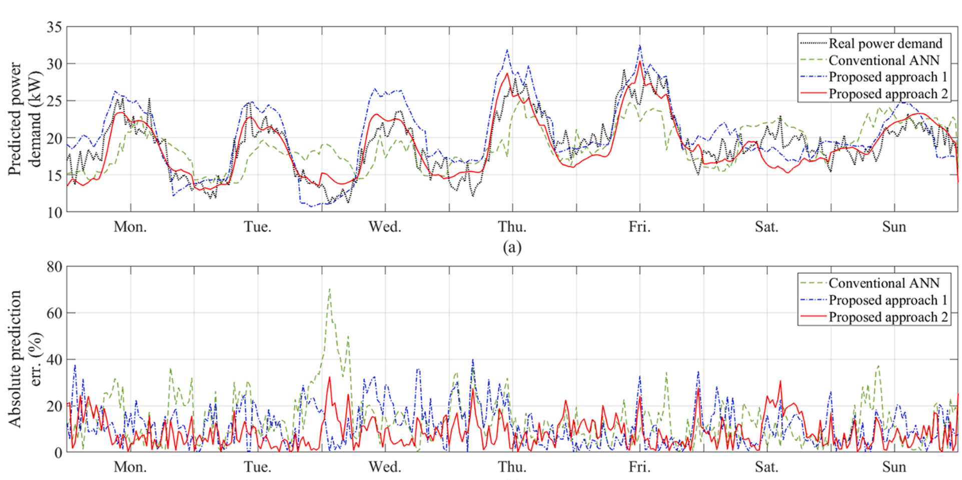


Fig. 12. Fitting result comparison among the ANN to full power, the baseload & human activity approach, and the ANN and fuzzy based approach.

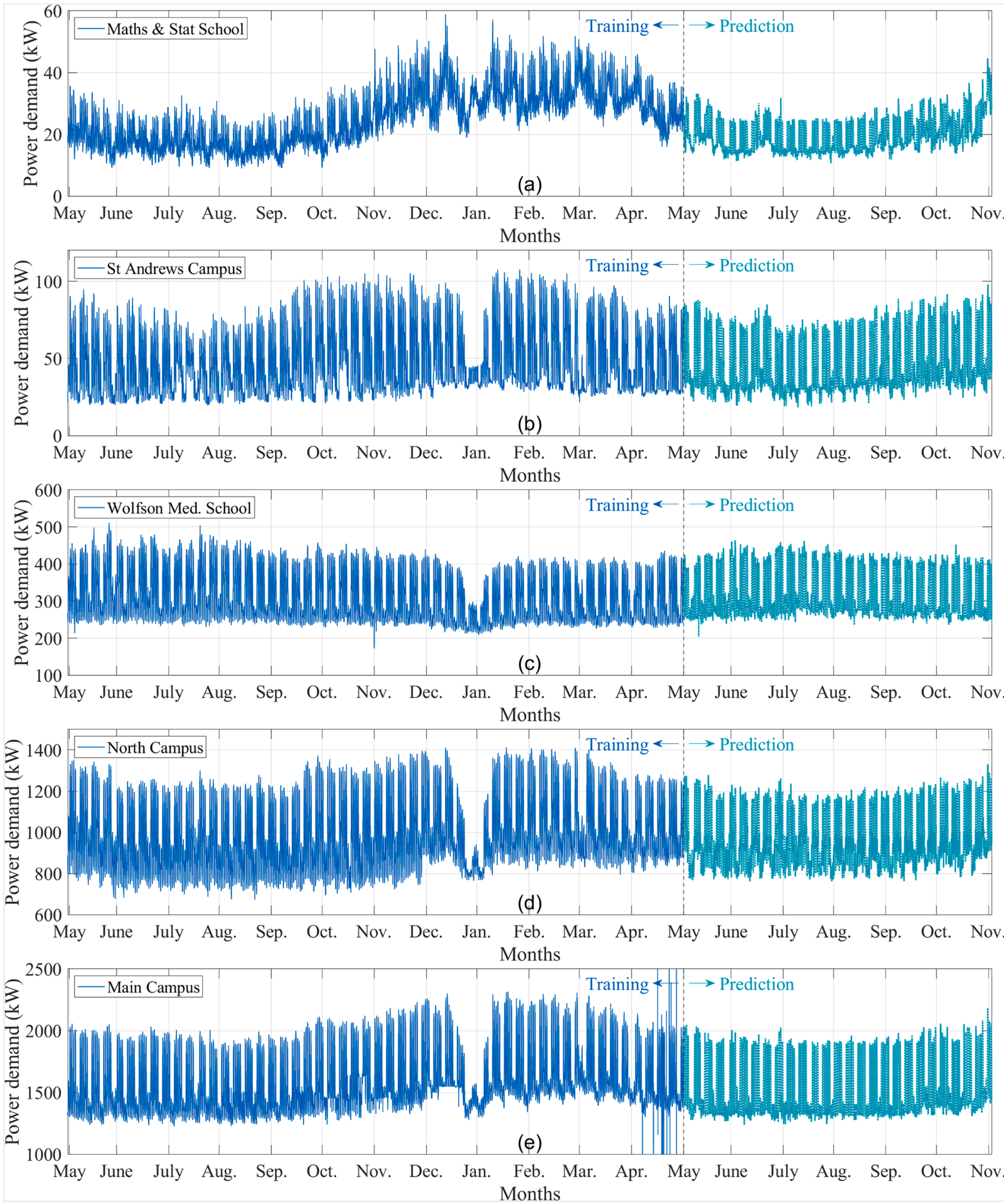


Fig. 13. Prediction of power demand of different campuses using the proposed approach.

metre to record the electricity consumption. The north campus and main campus only recorded the total power consumption of tens of buildings. In addition, the university has its own district heating system for most old buildings. As the Maths & State School is newly built, it is not included into the district heating system. Its space heating is fully supplied by the electrifying heating and, therefore, its electricity demand shows more relating to the weather conditions. Other buildings have different percentage of electrifying heating depending on occupants' behaviour. The weather conditions have less influence on these buildings comparing with the Maths & State School. Therefore, to validate the universality of the ANN based approach and its robustness to different data, the prediction results of other buildings or campuses using the same prediction approach are given in Fig. 13(b)-(e).

In order to compare their prediction performance of all methods numerically, the performance index is choosing the root-mean-square error (RMSE) between predicted power demand and real power demand of the last 6 months. The RMSE of all five campuses predicted by the conventional ANN approach and two proposed approaches are compared in a bar chart as shown in Fig. 14.

From the result, the proposed approach that uses the ANN to fit the

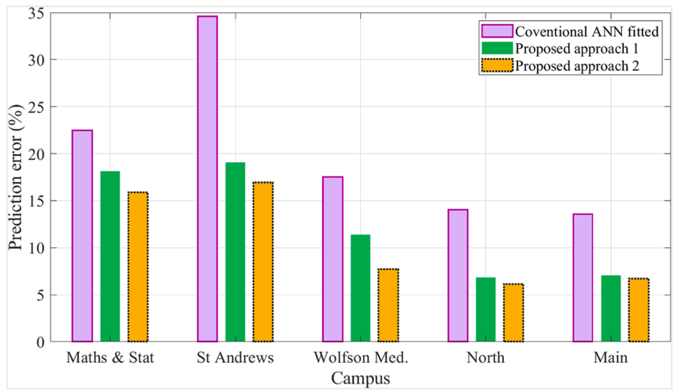


Fig. 14. Prediction error of different fitting approaches to each campus.

power demand data of working and non-working hours separately can get the best prediction performance with less RMSE. The proposed Approach 2 can reduce RMSE by 5% to 30% compared with the proposed Approach 1, and it reduces RMSE by 30% to 55% compared with the conventional ANN. In average, the proposed working hour splitting based regression approach and ANN with fuzzy hours splitting approach can reduce the RMSE prediction error by 35% and 42%, respectively. Thus, the proposed approaches can be used to predict the power demand, whose long-term and short-term data are in negative and positive correlation, respectively. In addition, the proposed approach with ANN and fuzzy technologies can be used as a 'grey-box' to include the knowledge of physical process to explain the effect of occupancy activities.

5. Conclusion

The electricity demand of office buildings seems to be in negative and positive correlation with weather variables in long-term and short-term time horizon, respectively, as a result the conventional ANN approach cannot accurately capture the relationships between them. In this paper, two electricity demand prediction approaches have been proposed to solve this issue. The initial proposed approach splits the power demand data by working hours and non-working hours to avoid the impact of occupants' activities to building power demand. Using this method, the linear regression approach is used to fit the building base-load power to a weather variable using the non-working hours data and find the average occupants power demand using the data of working hours. To fit the power demand with more weather variables, the proposed approach is further developed to use ANN to fit the non-working hours data and working hours data and the membership function of fuzzy hours between them. With the second proposed approach, more weather variables can be considered in the model to predict the power demand more accurately. In the simulation results, both approaches have been validated to show less RMSE value than the conventional ANN approach in predicting the power demand. In addition, the ANN with fuzzy hours splitting approach has the best performance among the three approaches and reduces RMSE by 5% to 30% compared with the working hour splitting based regression approach and reduces RMSE by 30% to 55% compared with the conventional ANN. Therefore, both approaches are able to solve the issue that the input and output fitting data are in negative and positive correlation in long-term and short-term time horizon, respectively. The proposed approaches can achieve good performance with RMS prediction error as low as 6% in building power demand prediction. In future works, the working hour splitting approach and fuzzy hour approach will be applied to other models, such as deep learning and stochastic models.

Declaration of Competing Interest

The authors declare no conflict of interest. The sponsors had no role in the design, execution, interpretation, or writing of the study.

Acknowledgment

The research presented in this article was undertaken as part of a project joint founded by Energy Technology Partnership (ETP), SP Distribution PLC (Scottish Power), grant number 146. The authors would like to thank Gillian Brown for providing valuable information on the campus energy consumption.

References

[1] Heo Y, Choudhary R, Augenbroe G. Calibration of building energy models for retrofit analysis under uncertainty. Energy Build 2012;47:550–60.
[2] Lyons P, Wade N, Jiang T, Taylor P, Hashiesh F, Michel M, Miller D. Design and analysis of electrical energy storage demonstration projects on UK distribution networks. Appl Energy 2015;137:677–91.
[3] Reynolds J, Ahmad MW, Rezgui Y, Hippolyte J-L. Operational supply and demand optimisation of a multi-vector district energy system using artificial neural networks and a genetic algorithm. Appl Energy 2019;235:699–713.
[4] Zhao H-x, Magoulès F. A review on the prediction of building energy consumption. Renew Sustain Energy Rev 2012;16(6):3586–92.
[5] Zeng A, Liu S, Yu Y. Comparative study of data driven methods in building electricity use prediction. Energy Build 2019;194:289–300.
[6] Ding Y, Wang Q, Wang Z, Han S, Zhu N. An occupancy-based model for building electricity consumption prediction: a case study of three campus buildings in Tianjin. Energy Build 2019;202:109412.
[7] De Rosa M, Bianco V, Scarpa F, Tagliafico LA. Heating and cooling building energy demand evaluation; a simplified model and a modified degree days approach. Appl Energy 2014;128:217–29.
[8] Chen S, Friedrich D, Yu Z, Yu J. District heating network demand prediction using a physics-based energy model with a Bayesian approach for parameter calibration. Energies 2019;12(18):3408.
[9] Jang J, Baek J, Leigh S-B. Prediction of optimum heating timing based on artificial neural network by utilizing BEMS data. J Build Eng 2019;22:66–74.
[10] Chen S, Ren Y, Friedrich D, Yu Z, Yu J. Sensitivity analysis to reduce duplicated features in ANN training for district heat demand prediction. Energy AI 2020;2:100028.
[11] Pedersen L, Stang J, Ulseth R. Load prediction method for heat and electricity demand in buildings for the purpose of planning for mixed energy distribution systems. Energy Build 2008;40(7):1124–34.
[12] De Felice M, Alessandri A, Ruti PM. Electricity demand forecasting over Italy: potential benefits using numerical weather prediction models, 104. Electric Power Systems Research; 2013. p. 71–9.
[13] Ahmad A, Hassan M, Abdullah M, Rahman H, Hussin F, Abdul-lah H, Saidur R. A review on applications of ANN and SVM for building electrical energy consumption forecasting. Renew Sustain Energy Rev 2014;33:102–9.
[14] Wei Y, Xia L, Pan S, Wu J, Zhang X, Han M, Zhang W, Xie J, Li Q. Prediction of occupancy level and energy consumption in office building using blind system identification and neural networks. Appl Energy 2019;240:276–94.
[15] Wilke U, Haldi F, Scartezzini J-L, Robinson D. A bottom-up stochastic model to predict building occupants' time-dependent activities. Build Environ 2013;60:254–64.
[16] Newsham GR, Birt BJ. Building-level occupancy data to improve ARIMA-based electricity use forecasts. In: Proceedings of the 2nd ACM workshop on embedded sensing systems for energy-efficiency in building; 2010. p. 13–8.
[17] Luo X, Oyedele LO, Ajayi AO, Akinade OO, Delgado JMD, Owolabi HA, Ahmed A. Genetic algorithm-determined deep feedfor-ward neural network architecture for predicting electricity consumption in real buildings. Energy AI 2020;2:100015.
[18] Kim S, Song Y, Sung Y, Seo D. Development of a consecutive occupancy estimation framework for improving the energy demand prediction performance of building energy modeling tools. Energies 2019;12(3):433.
[19] Yuan J, Farnham C, Azuma C, Emura K. Predictive artificial neural network models to forecast the seasonal hourly electricity consumption for a University campus. Sustain Cities Soc 2018;42:82–92.
[20] Kwak Y, Seo D, Jang C, Huh J-H. Feasibility study on a novel methodology for short-term real-time energy demand prediction using weather forecasting data. Energy Build 2013;57:250–60.
[21] Zeng A, Ho H, Yu Y. Prediction of building electricity usage using Gaussian process regression. J Build Eng 2020;28:101054.
[22] Cai H, Shen S, Lin Q, Li X, Xiao H. Predicting the energy consumption of residential buildings for regional electricity supply-side and demand-side management. IEEE Access 2019;7:30386–97.
[23] Amasyali K, El-Gohary NM. A review of data-driven building energy consumption prediction studies. Renew Sustain Energy Rev 2018;81:1192–205.
[24] Nizami SJ, Al-Garni AZ. Forecasting electric energy consumption using neural networks. Energy Policy 1995;23(12):1097–104.

[25] Massana J, Pous C, Burgas L, Melendez J, Colomer J. Short-term load forecasting for non-residential buildings contrasting artificial occupancy attributes. Energy Build 2016;130:519–31.

[26] Paterakis NG, Mocanu E, Gibescu M, Stappers B, van Alst W. Deep learning versus traditional machine learning methods for aggregated energy demand prediction. In: 2017 IEEE PES innovative smart grid technologies conference Europe (ISGT-Europe). IEEE; 2017. p. 1–6.

[27] Ahmad J, Larijani H, Emmanuel R, Mannion M, Javed A, Phillip-son M. Energy demand prediction through novel random neural network predictor for large non-domestic buildings. In: 2017 annual IEEE international systems conference (SysCon). IEEE; 2017. p. 1–6.

[28] Rahman A, Srikumar V, Smith AD. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Appl Energy 2018;212:372–85.