



Forecasting peak energy demand for smart buildings

Mona A. Alduailij¹ · Ioan Petri²  · Omer Rana³ · Mai A. Alduailij¹ · Abdulrahman S. Aldawood⁴

Accepted: 24 November 2020 / Published online: 8 December 2020
© The Author(s) 2020

Abstract

Predicting energy consumption in buildings plays an important part in the process of digital transformation of the built environment, and for understanding the potential for energy savings. This also contributes to reducing the impact of climate change, where buildings need to increase their adaptability and resilience while reducing energy consumption and maintain user comfort. The use of Internet of Things devices for monitoring and control of energy consumption in buildings can take into account user preferences, event monitoring and building optimization. Detecting peak energy demand from historical building data can enable users to manage their energy use more efficiently, while also enabling real-time response strategies (including control and actuation) to known or future scenarios. Several statistical, time series, and machine learning techniques are proposed in this work to predict electricity consumption for five different building types, by using peak demand forecasting to achieve energy efficiency. We have used several indigenous and exogenous variables with a view to test different energy forecasting scenarios. The suggested techniques are evaluated for creating predictive models, including linear Regression, dynamic regression, ARIMA time series, exponential smoothing time series, artificial neural network, and deep neural network. We conduct the analysis on an energy consumption dataset of five buildings from 2014 until 2019. Our results show that for a day ahead prediction, the ARIMA model outperforms the other approaches with an accuracy of 98.91% when executed over a 168 h (1 week) of uninterrupted data for five government buildings.

Keywords Energy forecasting · Time series · ARIMA · Peak demand · ANN · Smart buildings

✉ Ioan Petri
petrii@cardiff.ac.uk

Extended author information available on the last page of the article

1 Introduction

Climate change strategies were introduced in 2010 by the European Commission with clear objectives to reduce energy consumption and CO₂ emissions by 20%, noting that in Europe 40% of total energy is consumed by buildings (Directive 2010/31/EU) [1]. With the introduction of Smart Building Readiness Level, buildings are expected to “minimize the grid power usage and maximize services efficiency” identifying components such as sensors, renewable energy sources, and energy management system (EMS) [2].

Smart built environments have gone through a continuous transformation over the years, becoming more autonomous and reactive ecosystems that have the ability to balance energy consumption and user comfort, whilst also achieving higher order of safety for users [3]. Minimizing the energy consumption of buildings also has a cost dimension, as energy prices are fluctuating, which gives energy consumers and providers the ability to monetize energy especially when energy peaks can be predicted *a priori* with a certain level of accuracy [4]. The complexity of a building ecosystem requires a holistic analysis, as buildings have a large number of variables, are sensitive to changing conditions which lead to energy variability and dynamism within the building itself [5]. To address such complexity, building information modelling (BIM) can give a digital representation of the building, and support monitoring the performance of the building by facilitating integration of different information sources [6].

Energy consumption data can be interpreted from different perspectives with a view to find the best predictive model that can be used to forecast the use of energy for the next day, week or month. However, trying to find the best technique or algorithm for forecasting is a challenging problem. Some researchers prefer the use of statistical models, such as regression or time series, others adopt machine learning methods, like artificial neural network (ANN) and support vector machine (SVM). As sensors and energy meters have increased in capability and can transmit real-time consumption data, energy forecasting needs to respond to this dynamically produced data. To develop accurate models, using monthly data for predicting electricity consumption is a more common practice instead of prediction on a daily basis, since monthly data is more peremptory, especially when the variables related to users and the indoor environment are fluctuating [7]. In such scenarios, electricity consumption can be forecasted using artificial neural networks (ANN) with exogenous inputs [i.e. a nonlinear autoregressive network (NARX)].

An efficient method for predicting electricity consumption in buildings is the use of “soft computing” techniques which can support the optimization of energy flows in buildings [8]. Such methods make use of data measured by sensors installed in industrial buildings that can enable the implementation of different optimized decisions and actions to save energy. Energy forecasting has been investigated using several techniques such as multiple regression analysis, decision tree and neural networks [9]. These techniques provide satisfactory results for longer seasonal data sets but the results can be significantly influenced by

building type, physical characteristics of the building and operation time of an appliance within the building. Forecasting techniques can also be compared based on accuracy-related metrics, where regression analysis is most widely used due to its simplicity for interpreting model parameters (but at a reduced accuracy). Regression analysis is limited due to lack of mechanisms for assessing the causal dependencies between different input and output parameters. Similarly, neural networks in comparison with regression analysis cannot offer significance testing, e.g. p values, to test the importance of estimated parameters requiring an initial step to select features before learning.

We adopt a mixed-approach combining statistical, time series and machine learning models to forecast electricity consumption 24 h ahead for five different building types with an associated accuracy comparison. Furthermore, a peak detection algorithm is applied on the forecasted results, to determine energy peak and intervals between peaks for buildings. As a result, an integrated predictive model is proposed to assist facilities and building managers to reduce energy bills by predicting the daily peak hours of energy usage within a building. The same approach can also be used to determine peak hours for energy generation (e.g. through the use of photovoltaic panels installed within a building). The methodology and work proposed in this study considers the following:

- Identifying five buildings as representative examples of large-scale government buildings in a capital city. These buildings have a number of different functions, and include participants ranging from government employees, members of the public and specialist contractors.
- The first part of the study involves understanding the general usage of these buildings—including identification of general trends of electricity consumption, showing seasonality and weekday vs. weekend behaviours.
- The second part of the study involves the development of predictive (one day ahead) models derived using a number of different approaches, focusing on forecasting peak energy consumption. These comparative approaches illustrate the most appropriate (in terms of relevance or error rate) on the recorded time series data. The model construction makes use of real-world data recorded over a number of years.

A key focus of this work is a comparison of data analysis techniques (combining statistical analysis and machine learning) to support energy usage prediction for built environments. The outcomes can be used to support both reduced cost of energy and reduction in carbon emissions for smart cities (where buildings are seen as an important contributor). The rest of this paper is organized as follows: related work is reviewed in Sect. 2; followed by a description of the types of buildings, we consider in Sect. 3. The overall research methodology is presented in Sect. 4, with experimental results in Sect. 5. Finally, conclusions are provided in Sect. 6.

2 Related Work

Prediction of monthly energy consumption of buildings using temperature was investigated in [10] to obtain accurate forecasting based on heating and cooling and temperature. Other authors addressed energy consumption forecasting using linear regression for large public buildings [11]. Regression models with different granularities (1 day, 1 week and 3 months) were developed, with prediction error of energy consumption reaching 100%, 30%, and 6%, respectively, which suggests that the regression model is influenced by the length of measurements [12]. The work reported in [12] also demonstrates that day ahead prediction is more difficult compared to a 3 month prediction—primarily due to the potential variability that can be observed over a shorter time interval.

An online consumption prediction of energy for the next day using ARIMA model was implemented using historical data, with the next-day prediction supported through energy load profiles [13]. Other work included analysis of external inputs with ARIMA (ARIMAX) model to predict peak electricity consumption for commercial buildings [14, 15].

Electricity consumption forecasting strategies at a national level have been addressed by measuring hourly consumption patterns using time-series analysis [16]. The analysis shows that there was a 1000 MW difference in consumption between working days and weekends, with peak time on working days occurring around lunchtime, but on weekends, the peak occurred in the evening. A prediction of energy consumption and thermal comfort (PMV) of an indoor swimming pool was also investigated [17]. Several parameters were introduced in this study, such as time (minute, hour, day, month), occupancy, relative humidity, pool water temperature, room temperature, air temperature, and supplied air flow rate; to predict electricity consumption, thermal energy consumption, and PMV using an artificial neural network. Furthermore, it was stated that working with hourly energy consumption is better than smaller periods like minutes or seconds, to avoid noise in the data and improve prediction outcome [18]. Classical time-series decomposition was used to analyse electricity consumption of six commercial buildings [19], along with hourly weather data, like outdoor temperature and solar irradiation. The electricity consumption for residential houses in New Zealand were predicted to be 16–50% of the energy consumed by the residential sector in the country, and 30% globally [20].

Very short-term load forecasting (VSTLF) was identified as a useful method to consider, and which gives load forecast of up to one day ahead [21]. The VSTLF was used to analyse observations of minute-by-minute British electricity demand to evaluate different kinds of methods, like autoregressive integrated moving average (ARIMA) models and two exponential smoothing methods [22]. Linear regression to predict the annual energy consumption was constructed by using three different measurements, one day, one week, and three months [12]. It was shown that the accuracy of the predicted model of annual energy consumption of buildings were influenced by length of the measurement period being considered.

The forecasting models of 113 different studies over 41 academic papers were reviewed to determine which model was best suited for a specific context [23].

A number of different criteria were used in this comparison, such as time frame, inputs, outputs, and data sample size. For energy forecasting, a number of models were preferred: multiple linear regression, time series analysis, and artificial neural network. It was suggested that regression models were best suited for long-term prediction, while time series and ANN were best used for short-term predictions, especially when the pattern of the electricity consumption is complex.

The ARIMA model was compared with ANN and support vector machines, and it was observed that the ARIMA model was superior to other methods for developing a day ahead forecast [24]. In Saudi Arabia, one month ahead forecast of peak load of a utility was performed, where an ARIMA model was used to produce the forecasts [25].

A Holt Winters smoothing model or triple exponential Holt Winters model was used to forecast electricity demand. These smoothing models are used widely for seasonal data analysis. Holt Winters exponential smoothing model was used to forecast peak electricity loads for the national grid of England and Wales, to incorporate seasonal cycles of within a day and a week [26]. This model was then compared with ARIMA model and it was found that Holt Winters model provided better results compared to ARIMA, especially when there are trends and seasonality in the time series. When weather data was introduced to the forecast with Holt Winters exponential smoothing model, it gave better forecasts than ARIMA [22].

The ARIMA forecasting model to provide a day ahead forecast was previously utilized with satisfactory results in other studies [27]. Although the accuracy of the prediction is high, the accuracy gets even better for a very short-term forecast of 4 h ahead. Also, seasonal-related adjustments to reduce the electricity demand from peak periods to periods where the demand is low were developed. Panagiotidis et al. [27] also compared different models, e.g. ANN, ARIMA, and regression models to find the best prediction procedure for energy forecasting. Furthermore, they demonstrated that the ANN model provided equivalent accuracy to the ARIMA model, but that the ANN models were more difficult to generate and maintain. The advantage of the ARIMA model for supporting prediction is that it delivers a clear explanation of the influence of each variable to the overall prediction result [28]. This explanation capability could not be obtained easily for an ANN model, primarily due to the significant number of additional parameters used to specify the model. Therefore, the ARIMA model was used as the best model in this work. Conversely, other researchers recommend using an ANN model. For instance, a real-time energy monitoring system to reduce peak demand for a large government building in the USA was proposed in [29]. The developed ANN model was compared with other forecasting models, such as a simple moving average (SMA), linear regression, and multivariate adaptive regression splines (MARSplines).

Different statistical and machine learning algorithms to build a forecasting approach for predicting the peak electricity load for specific days of the month are identified [30]. The suggested model predicted 74 peak days for a one-year period, 40 of these peaks were true positives. This review also suggested that ARIMA and ANN models are the most frequently used techniques to forecast short term electricity demand. It was also suggested that the most important external variables are outdoor temperature and humidity. Lastly, the most forecasted period used by

researchers is between 2 to 4 weeks. Existing electricity demand prediction techniques are dependent on the geographic location and the condition of the building itself [31].

A stochastic model to predict a “triad” peak on a daily and half-hourly basis on building electricity demand data from Manchester was undertaken in [32]. A “triad” in this context refers to the three peaks that occur between November and February (winter months in the UK) when electricity usage is the highest. Predicting these peaks also required additional data for rescheduling of building operations or use of alternative sources to reduce the peak. Weather data was included in this model to increase accuracy of ANN forecasting model. The accuracy of the model reached 97.6%. To find suitable ranges for hyper parameters for ANN training, the authors performed a parametric study for each building. These parameters include the number of hidden layers, the number of neurons in each layer, learning rate and momentum. The results of the study shows that the best value of the hidden layer is when it matches the number of additional attributes; and the best value of the number of neurons is when it matches the total number of attributes, with learning rate 0.3 and momentum equal to 0.2. This work suggested that ANN models were comparable with other traditional techniques such as, linear regression, support vector machine, instance-based learning and decision trees.

ANN models are of different types, deep neural network (DNN) is currently the most widely used approach, and has been shown to provide highly accurate prediction over time series especially for sequential data [33]. Deep learning is a technique that can be used for predicting and forecasting energy for complex data and is superior to other machine learning and statistical methods [34].

A long-term forecast of annual electricity load that depends on weather parameters, using DNN for European countries is described in Butt et al. [35]. Historic data for Germany from 2006 to 2015 is used as training data, and the DNN were designed with five hidden layers and 1024 hidden neurons per layer. Rahman et al. [36] propose recurrent neural network (RNN) models to predict medium to long term electricity consumption of commercial and residential buildings in Utah and Texas (US) on an hourly basis. Their models have some limitations, especially when weather patterns differ from those at the time of collecting the data. Also, the accuracy of the model decreases when the structure of the building is changed. Nugalayadde et al. [37] forecasted electricity consumption for short-term, mid-term and long-term using previous electricity consumption only. The forecasting is performed using RNN and long short-term memory (LSTM). These two approaches were compared with popular predictions models such as ARIMA, ANN, and DNN.

Phyo [38] used DNN and RNN together with long short-term memory (LSTM) to forecast short-term load forecasting for nonlinear data in an attempt to enhance the accuracy of the results. The data represents 30 min load over March 2009 to December 2013 from the Electricity Generating Authority of Thailand. The experimental results suggest that the recommended model of DNN outperforms ANN and SVM models.

Muzaffar and Afhsari [39] used LSTM to forecast electricity load data combining with other variables such as temperature, humidity and wind speed. The forecast is used for short to medium term (24 h, 48 h, 7 days and 30 days).

Comparison of suggested model with other traditional methods was undertaken using accuracy measures such as RSME and MAPE. The comparison suggested that LSTM is better than other models with the challenge of improving the forecast accuracy.

In this paper, a predictive model is proposed that aims to assist (government) building managers to reduce energy costs by predicting daily peak hours and the energy demand during those hours. The suggested model consists of three parts. In the first part, statistical, time series and machine learning models are developed to forecast the next 24 h electricity consumption. In the second part, the models will be compared based on their accuracy. In the third part, an analysis algorithm is applied to determine a building's peak hours of energy usage. The model is evaluated using real weather and building energy consumption data from governmental buildings in Cardiff (UK).

3 Description of the buildings

Energy analysis is carried out on five government buildings, with data collected from utility electricity meters (kW) taken at a frequency of 30 min intervals over a period of 1–6 years—as summarized in Table 1 [40]. Electricity used for heating and cooling is an important characteristic since only electricity (i.e. not natural gas) data is analysed here.

Hourly weather data was collected within the proximity of buildings—the climate can vary during the warm summer months with a great chance of rainfall [41]. According to Köppen and Geiger, this climate and weather variables are important elements in forecasting models [42].

Electricity consumption is the key factor being considered in this study, as this is the variable to be predicted. The consumption patterns of three buildings are plotted in Fig. 1, where building no.1 (The Hall) had the highest usage among other buildings over the measured period. In the box plot of the five buildings, the first three buildings include data from 2014 until mid 2019, while for the fourth building (The Library), the electricity consumption is for approximately 6 months from the end of 2018 until the middle of 2019 (Fig. 2). Furthermore, Building no.5 (The School) has data for 6 months in 2019.

Building no.1 (The Hall) has the largest electricity consumption of over 100 kW, while the other buildings have a consumption of between 20 and 80 kW. Also, we can observe that there is a similarity between years; but when we focus on monthly periods (Figs. 3, 4), and daily period (Figs. 5, 6), the difference between the buildings appears clearly.

The maximum usage varied between buildings, Building no.1 (The Hall) showed that weekdays (Monday to Friday) were mostly similar in consumption, while weekends have a lower consumption (Fig. 5). Therefore, weekend and weekday data were separated to better demonstrate trends in consumption. On the other hand, Building no.3 (Library) showed that working day trends are from (Monday to Saturday) and weekend only on Sunday (day no.1) in this case (Fig. 6).

Table 1 Buildings properties

No	Building name	Period of data	Year built	No. of floors	Total usable floor area (m ²)	Heating type	No. of occupants
1	County hall	1-1-2014-8-7-2019	Post 1976	5	25,087	Gas	1122
2	New library	1-1-2014-8-7-2019	Post 1976	6	5665	Gas	600
3	Channel view	1-1-2014-8-7-2019	1967-1976	2	2887	Gas	-
4	Old library	26-12-2018-8-7-2019	Pre 1919	3	3889	Gas	-
5	School	22-1-2019-8-7-2019	Pre 1919	3	3462	Gas	362

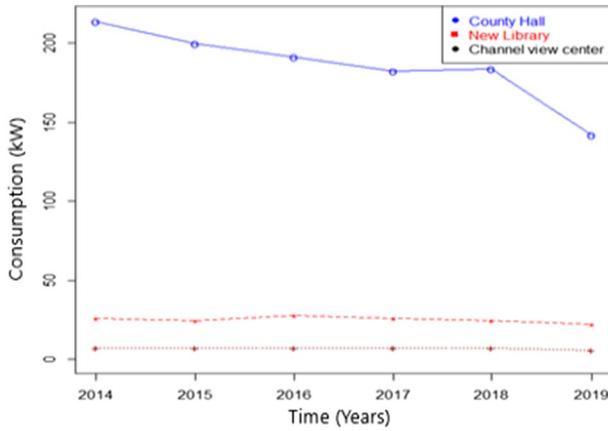


Fig. 1 Electrical consumption for three buildings for the years (2014–2019)

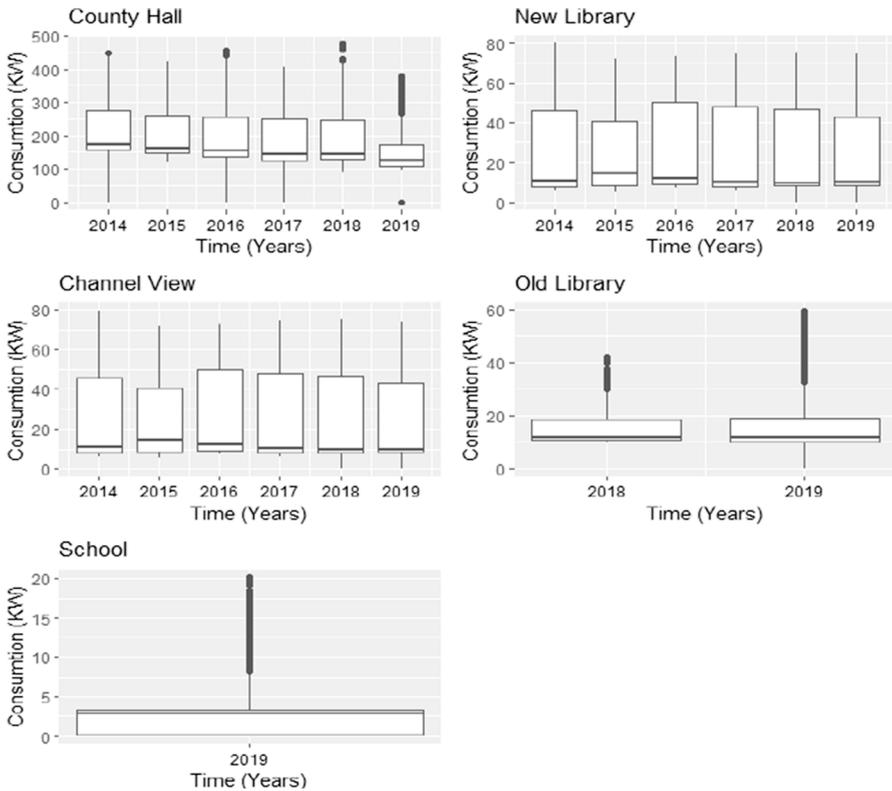


Fig. 2 Box-plot of yearly electricity consumption of the five buildings over 2014–2019

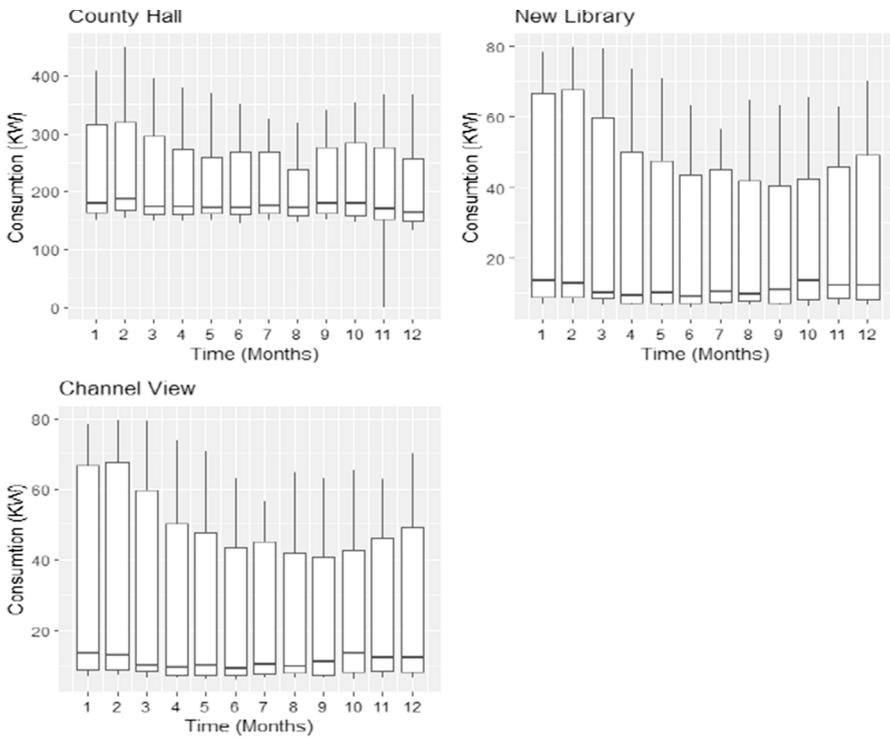


Fig. 3 Box-plot of monthly electricity consumption of three buildings for the year 2014

The prediction methodology relies on data for Building no.1 (The Hall), with 96,686 data points, representing half hour electricity consumption data for each day over the period 2014–2019.

4 Predictive Model

Given the general profile of building electricity consumption provided in the previous section, we develop a forecast of electricity consumption in kW for the next 24 h, based on a number of different parameter values as illustrated in Fig. 8. All statistical analysis was carried out using the R program with significance (p) value of 0.05 [43]. A statistical analysis was conducted on the data sets collected for all variables in order to see if there were any trends or seasonal affects. The variables that were used in the prediction process included:

Usage (Electricity consumption in kW per half hour interval): the electricity consumption of government building under study (Building no.1—The Hall) was measured using smart meters.

Day Type: represented as a number between 1 and 7, where 1 represents Sunday and 7 represents Saturday. Input data were classified according to day type.

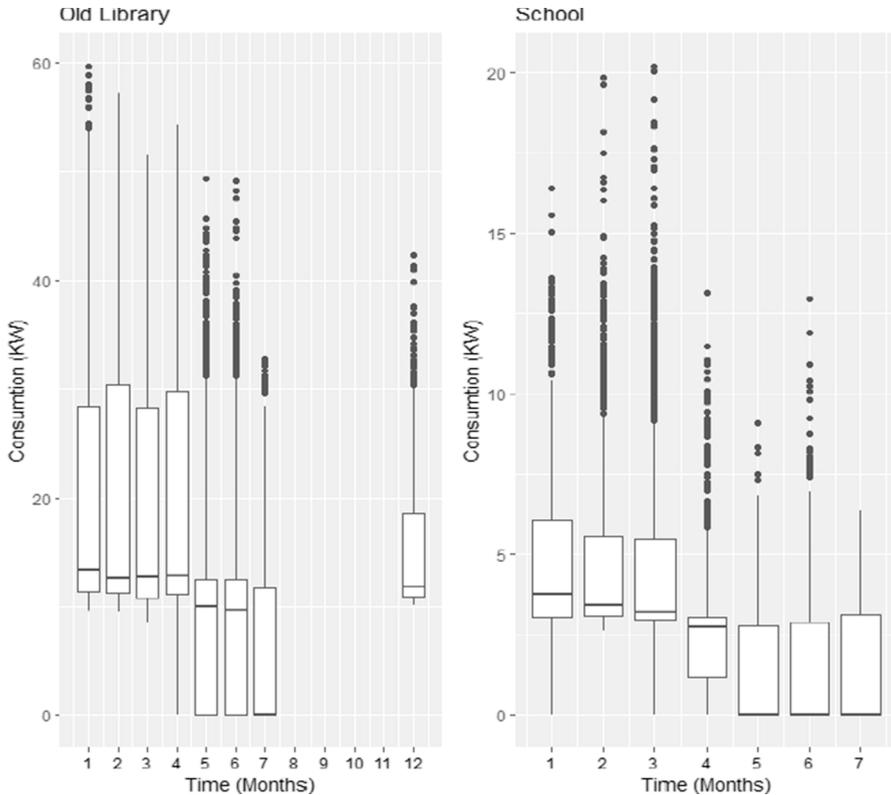


Fig. 4 Box-plot of monthly electricity consumption of two buildings for the years 2018–2019

Time of Day: Since electricity consumption of a building was different throughout the day, the box-plot of hourly electricity consumption was plotted (in Fig. 7) showing the distribution to be normal. High consumption was shown at the middle of the day from 10am until 12 pm, where it reached 450 kW. A lower electricity consumption was observed at early morning and late night, with a maximum of 200 kW. Additionally, even though the maximum electrical consumption was at midday, it can also reach a minimum value of zero, as weekends are included in this plot.

Temperature: There are many variables that can be related to weather, which could be used as indicators in the predicted models. Some of weather conditions are temperature, humidity, wind speed, wind direction, rain, barometric pressure and solar average. Temperature is one of the most important weather factors, as it directly influences electricity consumption. Exterior temperature therefore provides a useful proxy variable to capture the effects of weather.

Humidity: Humidity was used as a variable in the predicted model (over the period 2016–2017). This yearly data was used to represent a general trend to capture variation in humidity over the year.

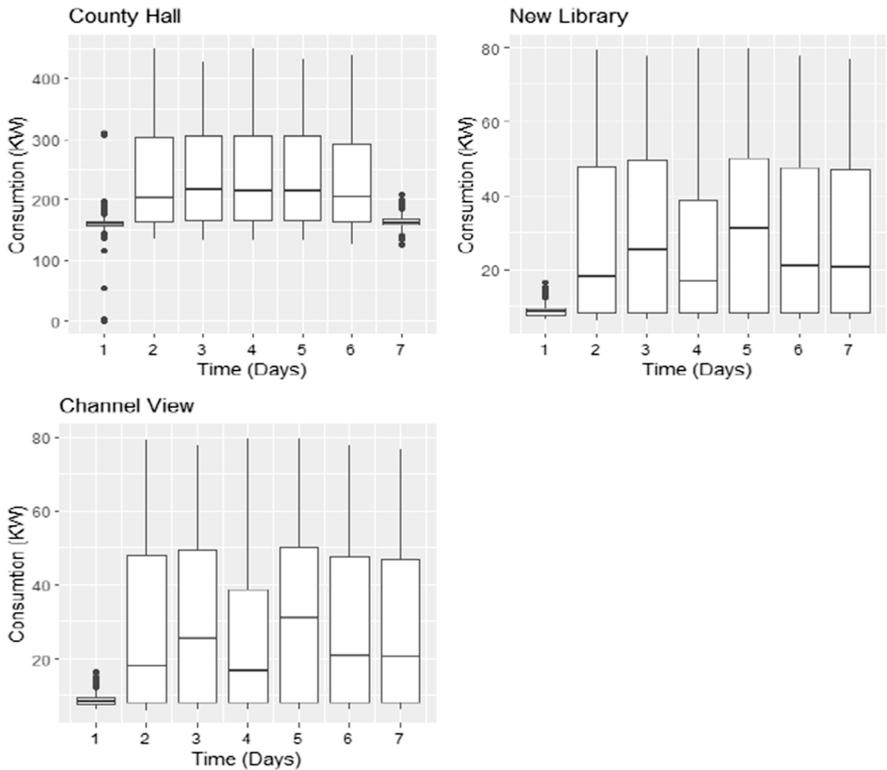


Fig. 5 Box-plot of electricity consumption over weekdays for three buildings for the year 2014

The models that have been used to predict electricity consumption include univariate time series, which depends on a historical perspective on electricity consumption and the period of the year being considered, while in the regression model (linear and dynamic) and machine learning models, other variables are included in the prediction (Fig. 8) [44].

4.1 ARIMA model

The first suggested model is the autoregressive moving average (ARIMA) time series. More specifically, this method involves considering the d th difference $W_t = \Delta^d Y_t$ as a stationary ARMA process. If $\{W_t\}$ follows an ARMA (p, q) model, $\{Y_t\}$ ARIMA (p, d, q) process can be called. The formula of an ARIMA $(p, 1, q)$ process, where $W_t = Y_t - Y_{t-1}$ [45] can be represented as:

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

where p is the number of autoregressive terms, d is the number of non-seasonal differences needed for stationary time series representation, q is the number of lagged forecast errors in the prediction equation.

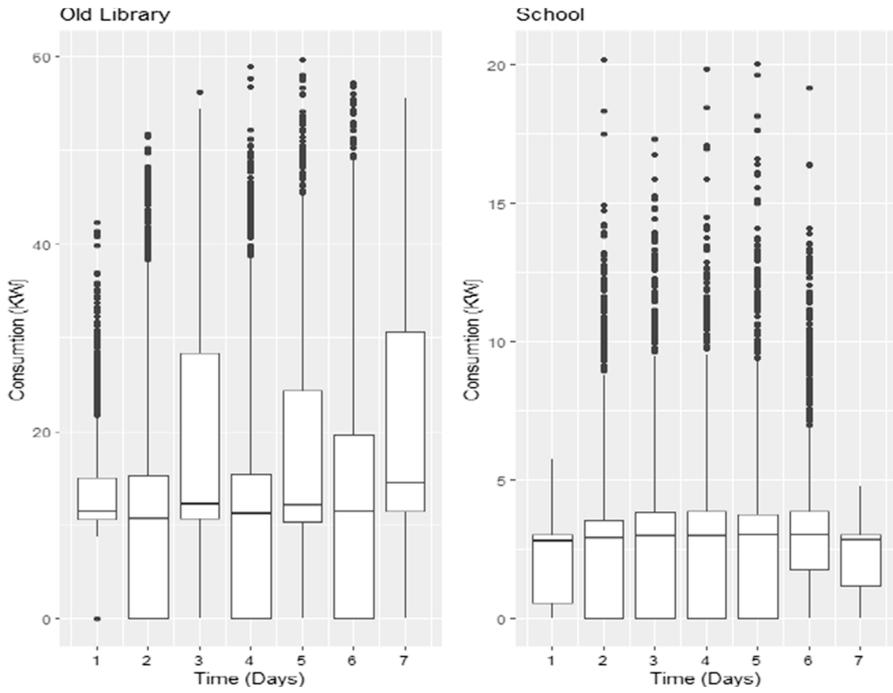


Fig. 6 Box-plot of electricity consumption during week days of two buildings for the years 2018–2019

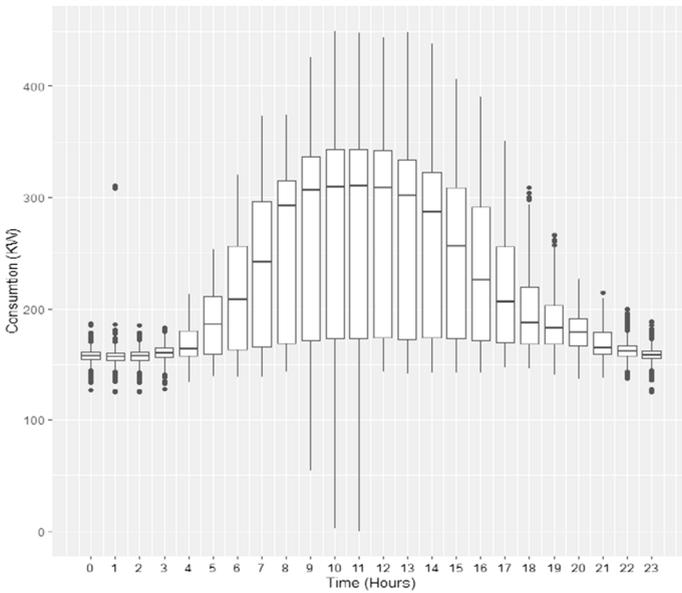


Fig. 7 Box-plot of the hourly consumption during 2014 for Building no. 1 (County Hall)

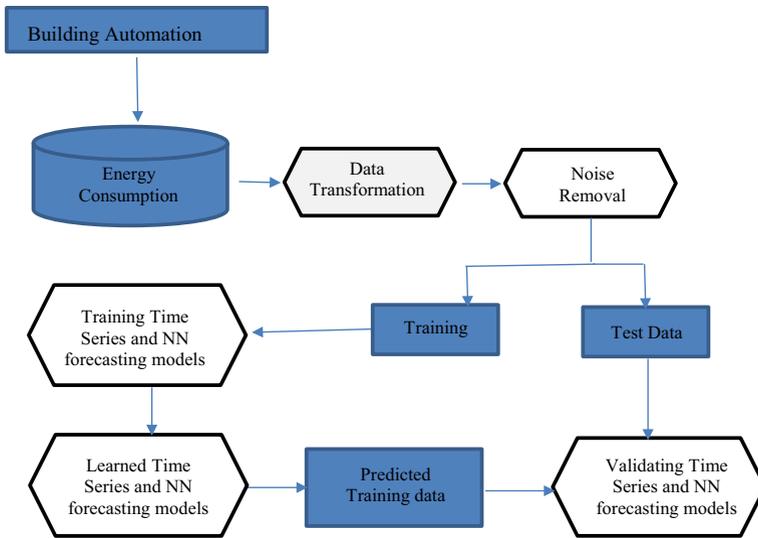


Fig. 8 Proposed forecasting workflow

Instead of deciding the value of parameters p , d , and q , the “R” program has a function called Auto ARIMA, that is used to return the *best* ARIMA model according to an information theoretic model, e.g. the Akaike Information Criteria (AIC). The function conducts a search over possible models within the set of solution based on predefined constraints [46]. Hence, AIC is an estimator of prediction error on a test set and is used to identify the quality of statistical models for a given set of data. Given a collection of models derived from the same data set, AIC estimates the relative quality of each model.

4.2 TBATS model

The second suggested model for forecasting building electricity consumption [47] is a modified version of the exponential smoothing time series with additional features, e.g. TBATS. This model is used to forecast complex seasonal time series data, such as those with multiple seasonal periods or where a high variation can be observed across seasons. The TBATS model incorporates Trigonometric functions, Box–Cox transformations, Fourier representations with time-varying coefficients and ARMA error correction. The TBATS model is as follows:

$$\begin{aligned}
 y_t^{(\omega)} &= \ell_{t-1} + \emptyset b_{t-1} + \sum_{i=1}^T s_{t-m_i}^{(i)} + d_t \\
 \ell_t &= \ell_{t-1} + \emptyset b_{t-1} + \alpha d_t \\
 b_t &= b_{t-1} + \beta d_t \\
 s_t^{(i)} &= s_{t-m_i}^{(i)} + \gamma_i d_t \\
 d_t &= \sum_{i=1}^p \varphi_i d_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t
 \end{aligned}$$

where: m_1, \dots, m_T denote the seasonal periods, ℓ_t is the local level in period t , b is the long-run trend, b_t is the short-run trend in period t , $s_t^{(i)}$ represents the i th seasonal component at time t , d_t denotes an ARMA(p, q) process, and ε_t is a Gaussian white-noise process with zero mean and constant variance σ^2 .

The smoothing parameters are given by α, β , and γ_i for $i = 1, \dots, T$.

4.3 Artificial neural network

The third model used to predict building electricity consumption is artificial neural network (ANN). The ANN is used to extract nonlinear relationships between response and predictor through learning from historical data [48] and [29].

Deep neural network (DNN) is a popular ANN [33] and used for both medium-term and long-term predictions. Recurrent neural network (RNN) and long short-term memory (LSTM) network are the most used deep neural network (DNN), especially networks that adapt feedback loop from past inputs [49]. RNN and LSTM have surpassed other DNN models that do not employ feedback loops [50].

For benchmark comparison across models, linear regression was applied to the one week data set. Dynamic linear models (DLMs) and time-series regression (dynlm) function in R were implemented and also applied to the one-week data set. The DLMs are a linear regression model, in which the parameters are treated as time-varying rather than static [51]. In these models, the coefficients can vary in time. A dynamic linear model can handle non-stationary processes, missing values and non-uniform sampling as well as observations with varying accuracies [52].

5 Results

The implementation of the suggested model for forecasting the next 24 h of electricity consumption of Building no.1—the Hall is discussed in this section. A one-week (168 h) uninterrupted data collection, at 30 min intervals was obtained using smart meters. Due to the pre-scheduled nature of daily building operations, a peak demand was only observed during week days (specifically Monday through Friday). Therefore, only weekday dataset was included in the model.

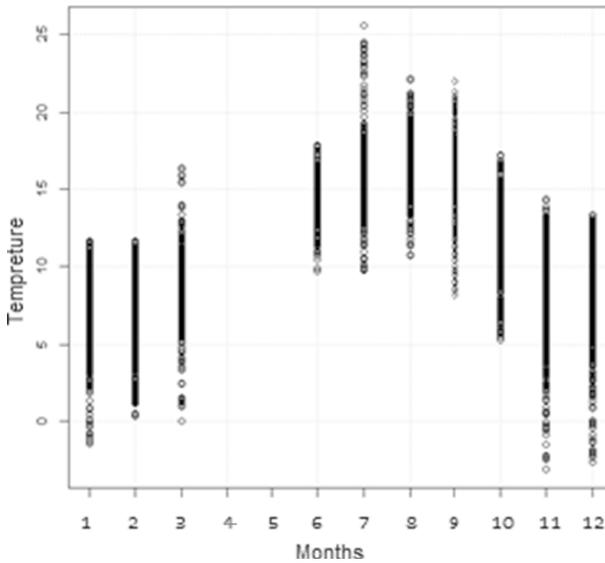
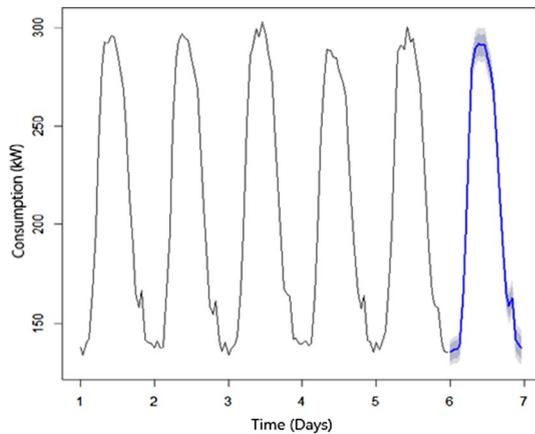


Fig. 9 Outside temperature during 1 year for 10 months

Fig. 10 Predicted electricity consumption using ARIMA (1, 0, 2) (1, 1, 1) model, with forecast for next 24 h



5.1 Forecasting analysis

The electricity consumption was plotted over the period 2014–2019 (Fig. 2), and it was observed that the electricity consumption of the building varies from 0 to 476.2 kW, reaching a maximum of 450 kW in 2018, 2016 and 2014, and a minimum of 0 in 2014, 2016, and 2017. The median values (horizontal line) are less than the average for all years, which equals to 189.35 kW.

Temperature is one of the most important weather factors, and available from June 2016 until March 2017. During this year, the lowest value occurred during

Fig. 11 Predicted electricity consumption using LSTM model, with forecast for next 24 h

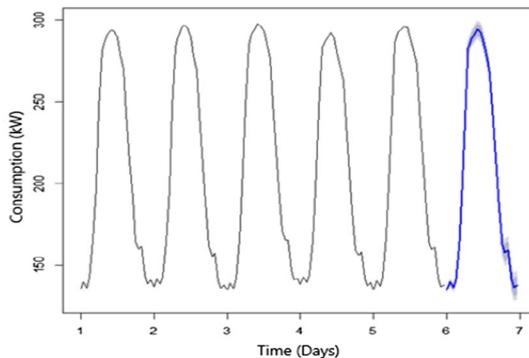


Table 2 One week mean absolute forecast errors (%) for electrical consumption of building no.1

Model	MAPE	Accuracy (%)
Linear regression	26.0246	74.98
Dynamic regression	5.0091	94.99
ARIMA	1.0855	98.91
TBATS	1.2172	98.78
ANN	1.6892	98.31
DNN (LSTM)	1.3804	98.62

November with the most variability, and the highest value in July, while the least temperature variability was in June (Fig. 9).

The ARIMA model was applied on samples of one week electricity consumption of the selected building. The model that gave the best forecast with minimum error had the following parameters, ARIMA (1, 0, 2) (1, 1, 1). The predicted electricity consumption model using ARIMA model and with a forecast for next 24 h (blue line) is shown (Fig. 10).

The TBATS model was also applied on samples of one week's electricity consumption. The model that gave a good forecast with low error had the following parameters: TBATS (0, 1, 1, 1, 1, {4, 3}, {168, 24}).

To reach the most suitable ANN topology (i.e. the number of hidden layers and number of nodes per layer), a parametric search was carried out using a trial and error method, to find the most suitable configuration for the ANN. The ANN is trained with one layer and number of hidden neurons is varied from 10 to 50. The models were trained on 70% of all data, tested on 15% and validated on the remaining 15%. The ANN is trained using the MATLAB 2019 Neural net fitting Toolbox with the following parameters: No. of Inputs = 5; No. of Outputs = 1; No. of Hidden layers = 1; No. of Hidden neurons = 10; Training Function: Levenberg–Marquardt backpropagation.

The models were trained on 90% of all data and tested on 10% of all the data. The LSTM is trained with the four layers with the following specifications: No. of Inputs = 1 (representing the time series input); No. of Outputs = 1; No. of Hidden layers = 4 (Sequence input layer, LSTM layer, fully connected layer, and regression

layer); No. of Hidden LSTM Units = 200; Solver = adam; No. of Training Iterations (MaxEpochs) = 250; Initial Learning Rate: 0.005; Learn Rate Schedule: piecewise; Gradient Threshold: 1; using the Deep Learning toolbox in MATLAB 2019 is used to train the network. (Fig. 11).

The 24-h forecasting mean absolute percentage error (MAPE) for the five models is shown in Table 2, where all errors were transformed to their absolute values. Similar error characteristics can be observed for all models except the linear regression model (which has the largest errors). The MAPE for all models are also plotted in Fig. 12. We observe that MAPE error for ARIMA model is 1.08% (Table 2), which is the lowest among all models; the 24-h forecast for TBATS is 1.21%, and with a similar accuracy to the ARIMA model.

The resulting R dynlm function is: Consumption kW = $15.83 - (0.04 \times \text{dayw}) - (0.31 \times \text{time}) + (1.6 \times L(\text{usage}, 1)) - (0.71 \times L(\text{usage}, 2))$, where $L(x, k)$ is $\text{lag}(x, \text{lag}) = -k$. The model has produced high values of Multiple R^2 with 97%, and Adjusted R^2 of 97%. This function requires that their arguments are time-series objects. The MAPE forecast error for dynamic regression model was 5%, which implies that this model is also suitable for predicting electricity consumption.

Applying linear regression to the one week data set resulted in a significance model with $p < 0$. The resulting regression equation is: Consumption kW = $333.71 - (11.6 \times \text{dayw}) - (1.2 \times \text{time}) + (5.3 \times \text{Temperature}) - (0.23 \times \text{humidity})$. The forecast MAPE of linear regression model was 26.02% (Table 1), which is the highest among all models. The predicted ANN model is illustrated in Fig. 12—the black line, with MAPE of 1.68%. The predicted curve of the five models; ARIMA, TBATS, ANN, dynamic regression, and linear regression are plotted in Fig. 13. The ARIMA, TBATS, ANN and dynamic regression are predicting the original data better than the linear regression (purple line).

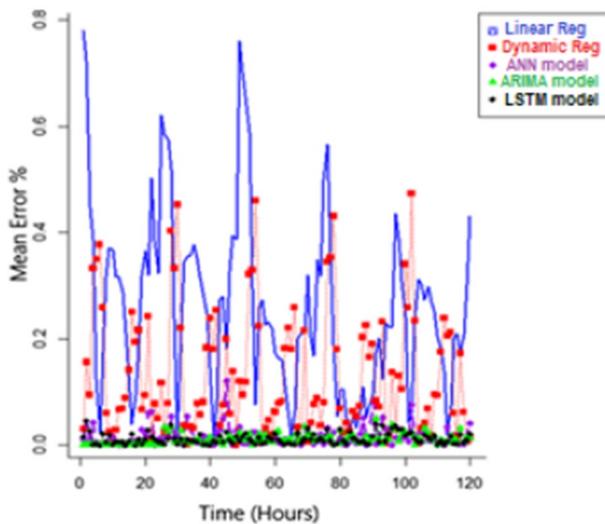
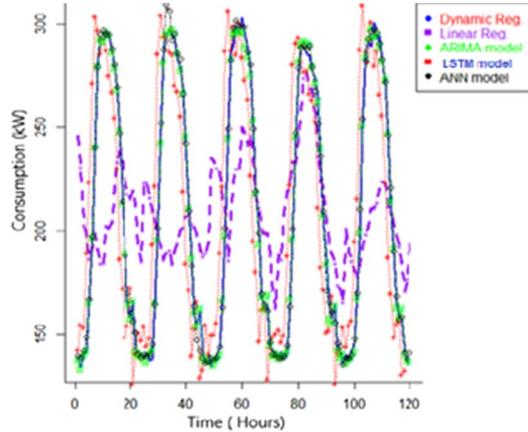


Fig. 12 Mean absolute error (%) for one-week electrical consumption of the five models

Fig. 13 Electrical consumption for one week of building no. 1 compared with the five suggested models



5.2 Peak forecasting for buildings

We apply the same models to the other four buildings with different input data according to the working days of the building involved. It was observed that each building had its own working days (Figs. 5, 6). The weekdays and weekends for each building are shown in Table 3.

From the error metrics and the forecast plot, it was observed that ARIMA, TBATS and ANN models provide the lowest error among the other predictive models. Therefore, ARIMA will be used to find the peak demand for all the five buildings; since it gives the least MAPE for all buildings.

Table 3 Building weekdays and weekends

Building	Weekdays	Weekends
County hall	Monday–Friday 5 days	Saturday and Sunday
New library	Monday–Saturday 6 days	Sunday
Channel view	Monday–Friday 5 days	Saturday and Sunday
Old library	Monday–Saturday 6 days	Sunday
School	Monday–Friday 5 days	Saturday and Sunday

Table 4 Peak and valley hours for the next 24 h with the predicted electrical consumption for the five buildings

Building	Peak hour	Prediction consumption	Valley hour	Prediction consumption
County hall	11	294.322	3 and 23	136.491
New library	10–13	50.391–55.142	3 and 19–22	7.964–8.215
Channel view	11–13 and 15–18	9.729–11.697	3–6 and 23	1.129–1.196
Old library	13–15	44.130–45.605	5–8	12.013–12.387
School	12	10.254	2–6 and 22	3.123–3.397

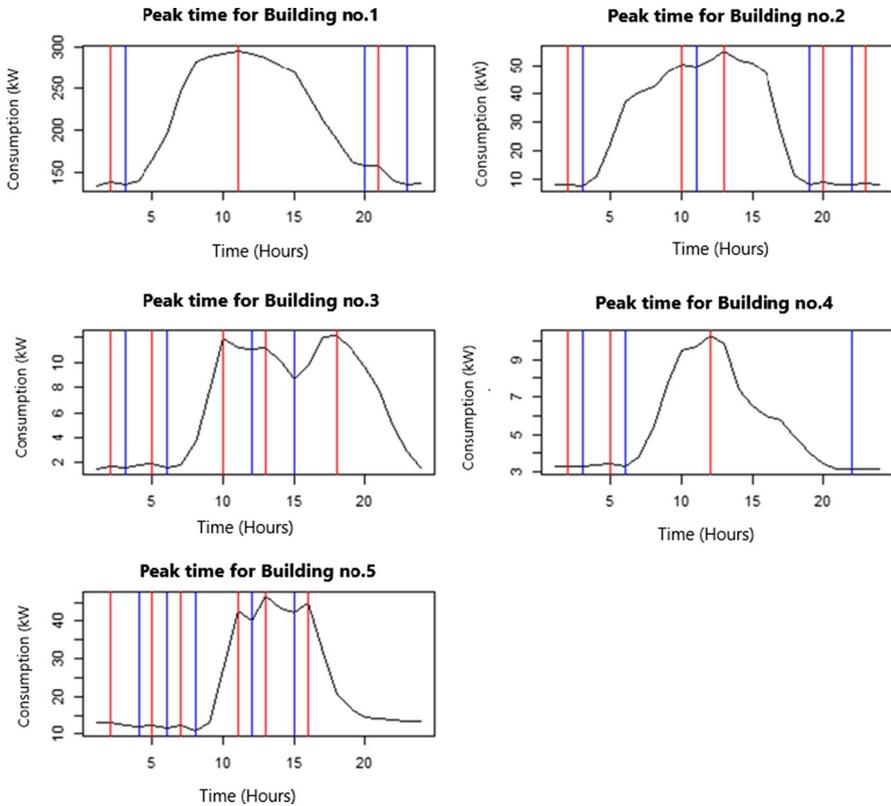
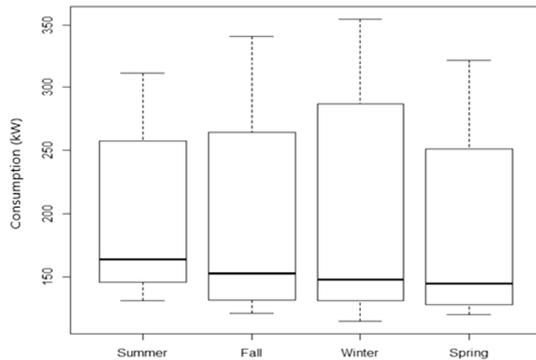


Fig. 14 Peak hours and valley hours for the next 24 h for the five Buildings

Fig. 15 Box-plot of the electrical consumption during the four seasons



The ARIMA model was then used to find the highest (peak) and lowest (valley) hours of electricity consumption for the next 24 h for the five buildings. The peak and valley hours are described in Table 4 and Fig. 14. To ignore the small peaks and large values, the maximum of all peaks and the minimum of all valleys will be computed. There is a large peak in the midday hours (10-1) for all buildings in Table 4.

Also, some buildings (e.g. Building no. 5—The channel) can have two peaks (11–1 and 15–18). Similarly, we can see two valleys, one in the afternoon (from 3 to 6 pm) and the other is at night (around 11 pm). Furthermore, electricity consumptions of the four seasons (summer, autumn, winter, and spring) were compared in Fig. 15 for Building no.1—The Hall. It can be observed that medians of these seasons are approximately the same, except for the difference in the maximum values and the deviation of seasons. For example, the winter had the largest and lowest values of the electrical consumption. It was also found that the mean value for electricity consumption observed across the four seasons had a small variation.

6 Discussion

This paper proposes a mixed predictive approach to forecast the peak energy demand for five large government buildings. Time-series models such as ARIMA and TBATS provide the lowest error in this instance, which is the short term predictions (24 h). Prediction of electricity consumption using ANN and LSTM networks are not too far from the time-series models with 98% and 98.5%, respectively. These results aligns with those reported in [38], as their comparison showed that ARIMA performed very well for short term. However, when the time interval of prediction increases, ARIMA does not exhibit good performance compared to RNN and LSTM. Overall, DNN outperformed other models with average root-mean-square RMSE reaching 0.1, especially for mid-term and long-term predictions.

The results of accuracy of all models especially ARIMA, ANN, and LSTM are competitive with other benchmarks. For example, in [32] the accuracy of their ANN model reach 97.6%, while the accuracy of our ANN is reaching 98.31%. Furthermore, MAPE for LSTM of [5] is about 1.522, while the MAPE of our LSTM is 1.3804, and the MAPE of their ARIMA models is 5.42 in average, while the MAPE of our ARIMA is 1.0855. [29] as we mentioned before compared ANN with other forecasting models, they got the best MAPE for ANN of 3.9 while other models like, simple moving average, linear regression, and multivariate adaptive regression their MAPE was 26.2%, 45.1%, and 22.5%, respectively.

The five buildings that have been chosen for this work are representative of other similar buildings in Cardiff. Understanding the electricity usage of these will provide a useful template for other types of similar built assets. These outcome of the analysis can be used in a number of ways:

- Understanding peaks will enable building managers to understand when additional sources (e.g. battery storage) can be integrated into the building;
- Understand how peak tariffs will influence the overall cost of operational management of the building—as reported in other literature (e.g. for predicting Triad peaks in building) [32].
- Understand how user behaviour can be influenced any reporting on peak usage—and therefore enabling users to become more active “consumer” of energy in the building.

- Other building types—e.g. Community Library in a city environment vs. a School has very different energy consumption patterns. The choice of the buildings in our study is also intended to reflect this diversity in building usage.

7 Conclusion

Smart meter data is used to undertake peak energy forecasting for a group of government buildings in Cardiff, UK. The proposed models are used to predict peak electricity power (kW) for the next 24 h, in order to give building and facilities managers the ability to minimize the peak demand for the next day (and to utilize alternative sources of energy to reduce tariffs – such as energy storage or renewables). Suitable strategies for developing models in this instance include: linear regression, dynamic regression, ARIMA, exponential time series (TBATS), ANN, and LSTM as a kind of deep neural network. The time-series models (ARIMA and TBATS) showed a very high accuracy, approximately 99%; followed by LSTM and ANN with an accuracy of 98.31% and 98.62%, while dynamic regression showed an accuracy reaching 94.99%. Linear regression was the worst performing, with an accuracy of 73.98%.

To predict the peak demand for the next 24 h, the ARIMA model was executed over a 168 h (one week) of uninterrupted data for the five buildings. An initial analysis was carried out on this data to find the peak and valleys hours during the next 24 h for these buildings, which was found to vary according to working hours—i.e. weekdays versus weekends.

The time series, ANN and LSTM models are very suitable for use in these kinds of buildings to predict peak electricity demand. Our future work will involve developing a recommendation system, to offer to the end-user, a forecast of the day ahead demand load as a mean to estimate peak consumption for the next 24 h, facilitating the shift of the high loads from peak periods to periods where the load is low.

Acknowledgements Omer Rana would like to extend appreciation to the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University for supporting his work through the visiting scholar program.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Directive 2010/31/EU (2010) Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings—(recast). Off J Eur Union L153: 13–35
2. Flax B (1991) Intelligent buildings. IEEE Commun Mag 29:24–27

3. Shah Salam A, Nasir H, Fayaz M, Lajis A (2019) A review on energy consumption optimization techniques in IoT based smart building environments. *Information* 10:108. <https://doi.org/10.3390/info10030108>
4. Chen H, Cong TN, Yang W, Tan C, Li Y, Ding Y (2009) Progress in electrical energy storage system: a critical review. *Prog Nat Sci* 19(3):291–312
5. Zhao H-X, Magoulès FF (2012) A review on the prediction of building energy consumption. *Renew Sustain Energy Rev* 16(6):3586–3592. <https://doi.org/10.1016/j.rser.2012.02.049>
6. Petri I, Li H, Rezguy Y, Chunfeng Y, Yuce B, Bejay J (2014) A modular optimisation model for reducing energy consumption in large scale building facilities. *Renew Sustain Energy Rev* 38:990–1002. <https://doi.org/10.1016/j.rser.2014.07.044>
7. Chen S, Ren T, Wu Z (2018) Research on neural network optimization algorithm for building energy consumption prediction. *J Comput Methods SciEng* 18:695–707
8. Moreno MV, Dufour L, Skarmeta AF, Jara AJ, Genoud D, Ladevie B, Beziau J-J (2016) Big data: the key to energy efficiency in smart buildings. *Soft Comput* 20(5):1749–1762
9. Geoffrey KF, Kelvin T, Yau KW (2007) Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks. *Energy* 32(9):1761–1768. <https://doi.org/10.1016/j.energy.2006.11.010>
10. White JA, Reichmuth R (1996) Simplified method for predicting building energy consumption using average monthly temperatures. In: *Proceedings of the 31st Intersociety Energy Conversion Engineering Conference, United States, vol 3*. pp 1834–1839
11. Ma Y, Yu JQ, Yang CY, Wang L (2010) Study on power energy consumption model for large-scale public building. In: *Proceedings of the 2nd international workshop on intelligent systems and applications. (ISA)*. Wuhan, China. pp 1–4
12. Cho SH, Kim WT, Tae CS, Zaheeruddin M (2004) Effect of length of measurement period on accuracy of predicted annual heating energy consumption of buildings. *Energy Convers Manag* 45(18–19):2867–2878
13. Kimbara A, Kurosu S, Endo R, Kamimura K, Matsuba T, Yamada A (1995) On-line prediction for load profile of an air-conditioning system. *ASHRAE Trans* 101(2):198–207
14. Hoffman AJ (1998) Peak demand control in commercial buildings with target peak adjustment based on load forecasting. In: *Proceedings of the 1998 IEEE International Conference on Control Applications, vol 2*. pp 1292–1296
15. Newsham GR, Birt BJ (2010) 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, BuildSys '10*. ACM, New York, pp 13–18
16. Majer, V (2011) Preparing and analysis of electricity consumption data for short term prediction. Intensive Programme “Renewable Energy Sources”, Železná Ruda-Špičák, University of West Bohemia, Czech Republic. pp 134–137
17. Yuce B, Li H, Rezguy Y, Petri I, Jayan B, Yang C (2014) Utilizing artificial neural network to predict energy consumption and thermal comfort level: an indoor swimming pool case study. *Energy Build* 80:45–56. <https://doi.org/10.1016/j.enbuild.2014.04.052>
18. Eisses J (2014) Anomaly detection in electricity consumption data. Thesis, University of Amsterdam, Faculty of Science, Amsterdam, pp 20
19. Pickering EM, Hossain MA, French RH, Abramson AR (2018) Building electricity consumption data analytics of building operations with classical time series decomposition and case based subseting. *Energy Build* 177:184–196
20. Ahmad A, Anderson TN, Rehman SU (2018) Prediction of electricity consumption for residential houses in New Zealand. In: Chong P, Seet BC, Chai M, Rehman S (eds) *Smart grid and innovative frontiers in telecommunications. SmartGIFT 2018. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering; vol 245*. Springer, Cham
21. Luo J, Hong T, Yue MJ (2018) Real-time anomaly detection for very short-term load forecasting. *J Modern Power Syst Clean Energy* 6:235–243. <https://doi.org/10.1007/s40565-017-0351-7>
22. Taylor JW (2008) An evaluation of methods for very short-term load forecasting using minute-by-minute British data. *Int J Forecast* 24(4):645–658
23. Kuster C, Rezguy Y, Mourshed M (2017) Electrical load forecasting models: a critical systematic review. *Sustain Cities Soc* 35:257–270. <https://doi.org/10.1016/j.scs.2017.08.009>
24. Fernandez I, Borges CE, Peña YK (2011) Efficient building load forecasting. *ETFA2011*. pp 1–8
25. Barakat EH, Al-Qassim JM, Al Rashed SA (1992) New model for peak demand forecasting applied to highly complex load characteristics of a fast developing area. *IEE Proc C* 139:136–149

26. Taylor JW (2003) Short-term electricity demand forecasting using double seasonal exponential smoothing. *J Oper Res Soc* 54(8):799–805
27. Panagiotidis P, Effraimis A, Xydis GA (2018) An R-based forecasting approach for efficient demand response strategies in autonomous micro-grids. *Energy Environ* 30(1):63–80. <https://doi.org/10.1177/0958305X18787259>
28. Adebisi AA, Adewumi AO, Ayo CK (2014) Comparison of ARIMA and artificial neural networks models for stock price prediction. *J Appl Math* 2014:1–7
29. Grant J, Eltoukhy M, Asfour S (2014) Short-term electrical peak demand forecasting in a large government building using artificial neural networks. *Energies* 7(4):1935–1953
30. Saxena H (2017) Forecasting strategies for predicting peak electric load days. Thesis, Rochester Institute of Technology. Accessed from <https://scholarworks.rit.edu/theses/9693>
31. Butt AA, Rahim MH, Khan M, Zahra A, Tariq M, Ahmad T, Javaid N (2018) Energy efficiency using genetic and crow search algorithms in smart grid. In: Xhafa F, Caballé S, Barolli L (eds) *Advances on P2P parallel, grid, cloud and internet computing*. Springer, Cham, pp 63–75
32. Marmaras C, Javed A, Cipcigan L, Rana O (2017) Predicting the energy demand of buildings during triad peaks in GB. *Energy Build* 141:262–273
33. Yuan J, Wang Y, Wang (2018) KLSTM based prediction and time-temperature varying rate fusion for hydropower plant anomaly detection: a case study. In: *Proceedings of international workshop of advanced manufacturing and automation*. Springer, pp 86–94
34. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444
35. Beh C, Nolting L, Praktiknjo A (2020) How to model European electricity load profiles using artificial neural networks. *Appl Energy* 277:115564
36. Rahman A, Srikumar V, Smith AD (2018) Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. *Appl Energy* 212:372–385
37. Nugaliyadde A, Somaratne UV, Wong KW (2019) Predicting electricity consumption using deep recurrent neural networks. *Arxiv*, volume=1909.08182
38. Phyo PP (2020) Deep learning for short term electricity load forecasting. Thesis, Ref. code: 25605822043898AVW, Thammasat University, Thailand. http://ethesisarchive.library.tu.ac.th/theses/2017/TU_2017_5822043898_7582_5819.pdf
39. Muzaffar S, Afhsari A (2019) Short-term load forecasts using LSTM networks. In: 10th International Conference on Applied Energy (ICAE2018), 22–25 August 2018, Hong Kong, China and Energy Procedia 158: 2922–2927
40. CarbonCulture. 2020/7/10. Cardiff Council. <https://platform.carbonculture.net/communities/cardiff-council/19/>
41. Merkel A (2020) AM online projects—Oedheim. <https://en.climate-data.org/europe/united-kingdom/wales/cardiff-5419/>. Date accessed: 1 Mar 2020
42. Rubel F, Kottek M (2011) Comments on: the thermal zones of the earth by WladimirKppen (1884). *Meteorol Z* 20(3):361–365. <https://doi.org/10.1127/0941-2948/2011/0258>
43. R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
44. Ruiz LGB, Cuellar MP, Calvo-Flores MD, Jimenez MCP (2016) An application of non-linear autoregressive neural networks to predict energyconsumption in public buildings. *Energies* 2016(9):684. <https://doi.org/10.3390/en9090684.58>
45. Cryer J, Cryer D, Chan K-S (2008) *Time series analysis: with applications in R*. Springer, Mathematics
46. Hyndman RJ, Khandakar Y (2008) Automatic time series forecasting: the forecast package for R. *J Stat Softw* 27(3):1–22
47. De Livera A, Hyndman R, Snyder R (2011) Forecasting time series with complex seasonal patterns using exponential smoothing. *J Am Stat Assoc* 106(496):1513–1527. <https://doi.org/10.1198/jasa.2011.tm09771>
48. Russell SJ, Norvig P (1995) *Artificial intelligence: a modern approach*. Prentice-Hall, Upper Saddle River, p 932
49. Nugaliyadde A, Wong KW, Sohel F, Xie H (2019) Language modeling through long term memory network. *arXiv preprint arXiv:1904.08936*.
50. Gers FA, Schmidhuber J, Cummins F (1999) Learning to forget: continual prediction with LSTM". In: 9th International Conference on Artificial Neural Networks: ICANN '99, pp 850–855
51. Holmes EE, Scheuerell MD, Ward EJ (2019) Applied time series analysis for fisheries and environmental data. NOAA fisheries, Northwest Fisheries Science Center, Seattle

52. Laine M (2019) Introduction to dynamic linear models for time series analysis. In: Montillet J-P, Bos M (eds) A chapter submitted to a book with a proposed title: geodetic time series analysis and applications. https://doi.org/10.1007/978-3-030-21718-1_4. Latest version 21 May 2019

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Mona A. Alduailij¹ · Ioan Petri²  · Omer Rana³ · Mai A. Alduailij¹ · Abdulrahman S. Aldawood⁴

¹ College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia

² School of Engineering, Cardiff University, Cardiff, UK

³ School of Computer Science and Informatics, Cardiff University, Cardiff, UK

⁴ College of Food and Agriculture Science, King Saud University, Riyadh, Saudi Arabia