Renewable and Sustainable Energy Reviews 73 (2017) 1104–1122

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

A review and analysis of regression and machine learning models on commercial building electricity load forecasting

B. Yildiz, J.I. Bilbao*, A.B. Sproul

School of Photovoltaics and Renewable Energy Engineering, University of New South Wales, Sydney, NSW 2052, Australia

ARTICLE INFO

Keywords:
Short term load forecasting for commercial buildings
Review of regression models
Machine learning
Neural Networks
Support Vector Regression
Regression Trees

ABSTRACT

Electricity load forecasting is an important tool which can be utilized to enable effective control of commercial building electricity loads. Accurate forecasts of commercial building electricity loads can bring significant environmental and economic benefits by reducing electricity use and peak demand and the corresponding GHG emissions. This paper presents a review of different electricity load forecasting models with a particular focus on regression models, discussing different applications, most commonly used regression variables and methods to improve the performance and accuracy of the models. A comparison between the models is then presented for forecasting day ahead hourly electricity loads using real building and Campus data obtained from the Kensington Campus and Tyree Energy Technologies Building (TETB) at the University of New South Wales (UNSW). The results reveal that Artificial Neural Networks with Bayesian Regulation Backpropagation have the best overall root mean squared and mean absolute percentage error performance and almost all the models performed better predicting the overall Campus load than the single building load. The models were also tested on forecasting daily peak electricity demand. For each model, the obtained error for daily peak demand forecasts was higher than the average day ahead hourly forecasts. The regression models which were the main focus of the study performed fairly well in comparison to other more advanced machine learning models.

1. Introduction

According to reports from the International Energy Agency (IEA) [1], the commercial building sector accounts for 32% of the final electricity consumption in OECD countries. In particular, this number was reported as 29% for European countries and in the USA, more recent reports showed that commercial buildings accounted for over 35% of end-use electricity consumption [2]. In Australia, commercial buildings accounted for around 30% of the electricity end-use consumption [1] and 10% of the total greenhouse gas emissions in 2013 [3]. Most of these buildings have inefficiencies in energy use due to their physical nature. The Rocky Mountain Institute has stated that there is the potential to reduce commercial building energy use by 20% in the USA and other reports indicate that there is a reduction potential of about 29% [4]. These numbers suggest the importance of focusing efforts on understanding and reducing the energy use and demand of commercial buildings. Furthermore, it's well known that reducing peak electricity demand is a clear pathway to achieve economic and environmental benefits. For example, peak demand is identified as the main driver for the growing investments in network infrastructure which exerts upward pressure on electricity prices [5]. The Energy Supply Association of Australia estimates that 80% of the investment in grid upgrades was required to meet the growing peak demand in Sydney [6]. Hence, the implementation of accurate and robust electricity load forecast methods both at distribution network and end user levels can assist demand management and energy efficiency activities which can be considered as alternative solutions to electricity network augmentation [7].

Commercial buildings equipped with modern monitoring and metering systems along with building management systems are well

Abbreviations: AR, Auto Regressive; MA, Moving Average; ARMA, Auto Regressive Moving Average; ARIMA, Auto Regressive Integrated Moving Average; ANN, Artificial Neural Network; NARX, Nonlinear Autoregressive Network with Exogenous Inputs; SVM, Support Vector Machine; SVR, Support Vector Regression; SLR, Single Linear Regression; MLR, Multivariate Linear Regression; PRISM, The Princeton Scorekeeping Method; R^2, Coefficient of Determination; $R_{adj}{}^2$, Adjusted Coefficient of Determination; CV, Coefficient of Variance; RMSE, Root Mean Squared Error; CV-RMSE, Percentage RMSE by the mean; MAPE, Mean Absolute Percentage Error; MPE, Mean Percentage Error; TMY, Typical Meteorological Year; DDCAV, Dual Duct Under Constant Air Volume; DDVAV, Dual Duct Under Variable Air Volume; TRCAV, Terminal Reheat Under Constant Air Volume; TRVAV, Terminal Reheat Under Variable Air Volume; DBT, Dry Bulb Temperature; T_{dp}, Dew Point Temperature; RH, Relative Humidity; q_{sol}, Solar Heat Gains; q_i, Sensible Heat Gains; I, Indicator Variable; WWR, Window to Wall Ratio; UNSW, University of New South Wales

* Corresponding author.
E-mail address: j.bilbao@unsw.edu.au (J.I. Bilbao).

http://dx.doi.org/10.1016/j.rser.2017.02.023
Received 29 March 2016; Received in revised form 2 December 2016; Accepted 4 February 2017
1364-0321/

suited to implement electricity load reduction activities. Moreover, potential economic benefits brought about by reducing the demand can become more significant for prosumers – customers who produce as well as consume energy. However, energy systems in commercial buildings can be complex systems, particularly in buildings with large heating, ventilation, and air-conditioning (HVAC) systems. This complexity makes the evaluation and forecasting of electricity demand quite challenging. The main cause of the difficulty can be attributed to the variation in the energy consumption profiles within buildings [8]. The problem increases in buildings that have a mix of areas which have different HVAC and lighting requirements such as offices, laboratories, lecture theatres, operating theatres, event rooms, and data centre and manufacturing facilities. In addition, building electricity loads vary with internal factors such as occupancy and scheduling. Last but not least, building loads are also susceptible to the changes in external weather parameters such as temperature, solar radiation and humidity.

Because of the reasons outlined above, numerous attempts have been made to accurately forecast commercial building electricity loads. Different techniques such as thermal models, statistical regression models, time series models and machine learning models have been used in forecasting commercial building electricity loads for various climates and time horizons (short-term, mid-term and long-term). Shor-term forecasts (minutes to a week ahead) can have an immediate impact on a building's operation and scheduling and is a crucial component for building energy management systems. Mid-term (a week to a year ahead) and long-term (more than a year ahead) forecasts have greater importance for longer term planning. Existing review papers have provided a good foundation for classifying the work done in terms of model types, forecast horizon and scale (single building to regional or national level) [9–13], while some articles provide a particular focus on certain methods [11,14,15].

Among the different forecasting methods, regression models are simple to develop, use, and interpret, in comparison to other more complex techniques hence, they have been commonly used for building load forecasting [9,16]. Regression models correlate the energy consumption with external weather and internal building parameters. These models can be developed by using real historical load data [16–20] or simulated load data [21–25]. To the authors' knowledge, there hasn't been a detailed review particularly focused on regression models in commercial building load forecasting, although there have been comprehensive studies where regression models were used in commercial building electricity load forecasting and its performance was compared with other methods [26–28]. Our study therefore presents a thorough review on regression models and aims to inform the reader about the range of different applications where these models can be successfully used. For clarity, regression models are classified under different categories such as their area of application, commonly used input parameters, methods to improve model performance and comparison with other models.

In order to extend our study beyond theory, regression models are implemented in a day ahead hourly electricity load forecast analysis by applying the methodologies discussed in the review section. Studies where regression models were used for commercial building electricity load forecasting and their performances were compared with other methods, were mainly limited to single building level and the results were only analysed for the overall data-set. This study extends the analysis by implementing forecast models for different scales: single building and university Campus level which allows us to observe the impact of load scale on forecast performance. Furthermore, our study also allows the comparison of model performance for different seasons. The analysed models are trained and tested not only for the overall data set but also with seasonal sub-sets. In addition to the day ahead hourly load forecasts, the analysis is broadened to forecast daily peak loads by modifying the models which gave another opportunity to compare model performance on different target loads.

The remainder of the paper is organized as follows. In Section 2, a brief review of models apart from regression models is introduced. In Section 3, a detailed review on the use of regression models is presented. Following the review sections, in Section 4 the implementation of regression models for forecasting day ahead hourly and daily peak electricity loads, for a single university building (Tyree Energy Technologies Building – TETB) and a university Campus with around 50 buildings on a 38 ha site (Kensington Campus – UNSW) is discussed. Following the regression analysis, four other machine learning models are used for the forecast analysis: Artificial Neural Networks (ANN) with Levenberg Marquitd (LM) and Bayesian Regulation (BR) Backpropagation, Nonlinear Autoregressive Network with Exogenous Inputs (NARX) with LM and BR Backpropagation, Regression Trees (RT) and Support Vector Regression (SVR). ANN and SVR are the most commonly used machine learning models in the area [9,14]. Implementation of regression trees is also not uncommon [26,27,29] whereas NARX is a relatively new method and to our knowledge its implementation has been limited, thus we wanted to compare this method with other popular machine learning models. The performance, ease of use, and interpretability of the models are compared in Section 5.

2. Models used in commercial building electricity load forecasting

2.1. Thermal models

Thermal models calculate heat transfers and energy behaviour on a sub or whole building level [9]. Heat transfer calculations are based on the interaction of the building envelope with internal and external environments. Comprehensive thermal models may require a high number of inputs in comparison to simpler models. Historical data is typically not required for these models [30].

Analytical thermal modelling software such as DOE-2, Energy Plus, BLAST, and ESP-r has been developed for evaluating energy consumption and efficiency in buildings. This type of software has been widely used for developing building energy standards and analysing energy consumption and conservation measures in buildings. Although these models are quite powerful, they require detailed data regarding the building envelope, external weather, occupant behaviour and interior equipment performance, which may not be accessible to some users [30]. Section 3.1 gives examples of regression models developed by using thermal model simulation data.

2.2. Auto regressive models

Auto regressive models analyse sets of data points in a time series and correlate the future value of a certain variable with its past values [30]. It is possible to correlate different input variables to the output such that in commercial building load forecasting, load can be correlated with other important weather and building parameters. For example, Espinoza et al. [31] used a periodic auto regression model for 245 substations where each substation had four years of hourly data points. The method yielded satisfactory results for short term forecasting (225 substations out of 245 showed an R^2 higher than 90%).

One of the most commonly used forecasting techniques amongst the time series models is based on the Box-Jenkins methodology, which combines the Auto Regressive (AR) order p and Moving Average (MA) order q of the time series. The model is called ARIMA when an additional differencing order d is integrated into the model in order to remove the possible non-stationarities within the data [30]. A study done by Amjady [32] uses a novel Box Jenkins method for short term and peak load forecasts. This modified ARIMA method uses an initial forecast input and combines it with temperature and load data for the regression analysis. The method can accurately forecast hourly and peak loads and gives better results than the standard ARIMA model (for three operators located in different climatic zones of Iran, the