1-1-2015

A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables

Dao Hoang Vu
University of Wollongong, dhv972@uowmail.edu.au

Kashem M. Muttaqi
University of Wollongong, kashem@uow.edu.au

Ashish Agalgaonkar
University of Wollongong, ashish@uow.edu.au

A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables

Abstract

Selection of appropriate climatic variables for prediction of electricity demand is critical as it affects the accuracy of the prediction. Different climatic variables may have different impacts on the electricity demand due to the varying geographical conditions. This paper uses multicollinearity and backward elimination processes to select the most appropriate variables and develop a multiple regression model for monthly forecasting of electricity demand. The former process is employed to reduce the collinearity between the explanatory variables by excluding the predictor which has highly linear relationship with the other independent variables in the dataset. In the next step, involving backward elimination regression analysis, the variables with coefficients that have a low level of significance are removed. A case study has been reported in this paper by acquiring the data from the state of New South Wales, Australia. The data analyses have revealed that the climatic variables such as temperature, humidity, and rainy days predominantly affect the electricity demand of the state of New South Wales. A regression model for monthly forecasting of the electricity demand is developed using the climatic variables that are dominant. The model has been trained and validated using the time series data. The monthly forecasted demands obtained using the proposed model are found to be closely matched with the actual electricity demands highlighting the fact that the prediction errors are well within the acceptable limits.

Disciplines

Engineering | Science and Technology Studies

Publication Details

D. H. Vu, K. M. Muttaqi & A. P. Agalgaonkar, "A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables," Applied Energy, vol. 140, pp. 385-394, 2015.

A Variance Inflation Factor and Backward Elimination based Robust Regression Model for Forecasting Monthly Electricity Demand using Climatic Variables

D. H. Vu, K. M. Muttaqi and A. P. Agalgaonkar
Australian Power Quality and Reliability Centre, School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong, New South Wales, Australia
(e-mails: dhv972@uowmail.edu.au; kashem@uow.edu.au; ashish@uow.edu.au)

Abstract: Selection of appropriate climatic variables for prediction of electricity demand is critical as it affects the accuracy of the prediction. Different climatic variables may have different impacts on the electricity demand due to the varying geographical conditions. This paper uses multicollinearity and backward elimination processes to select the most appropriate variables and develop a multiple regression model for monthly forecasting of electricity demand. The former process is employed to reduce the collinearity between the explanatory variables by excluding the predictor which has highly linear relationship with the other independent variables in the dataset. In the next step, involving backward elimination regression analysis, the variables with coefficients that have a low level of significance are removed. A case study has been reported in this paper by acquiring the data from the state of New South Wales, Australia. The data analyses have revealed that the climatic variables such as temperature, humidity, and rainy days predominantly affect the electricity demand of the state of New South Wales. A regression model for monthly forecasting of the electricity demand is developed using the climatic variables that are dominant. The model has been trained and validated using the time series data. The monthly forecasted demands obtained using the proposed model are found to be closely matched with the actual electricity demands highlighting the fact that the prediction errors are well within the acceptable limits.

Keywords: Climatic variables, Electricity demand, Electricity forecasting, Multiple regression, Multicollinearity.

1. Introduction

Determining the impact of climate change on electricity demand is one of the challenging aspects in terms of demand forecasting in recent years. Particularly, the slight upward-trend in temperature in Australia [1] can reduce electricity consumption in cold regions due to reduction in the heating demand but may pose more strain on the electricity grid in the other areas due to increase in the cooling requirement. In addition, with the growth of gross domestic product (GDP) and the boost of population, the energy consumption may increase. Consequently, electricity demand in the future is expected to change depending on the life-style and regional influences. Therefore, electricity demand forecasting becomes an essential tool for energy management, maintenance scheduling, and investment decisions in the future energy markets.

An extensive literature on forecasting models and strategies has been reviewed in [2]. The reported forecasting methods are generally classified into two main groups: autonomous models and conditional models [3]. The autonomous models are based on the historical data of the electricity demand for forecasting the future demand while the conditional models build up the relationship between the electricity demand and the other associated variables, and then forecast the future demand based on the changes in the variables. The neural network [4] and Kalman filter application [5], [6] are claimed to be sufficiently efficient in short term forecasting, and the multiple linear regression model is widely used for long term demand forecasting [7], [8], [9], or medium-term forecasting [10], [11].

Since the combination of the traditional models can utilize the advantages of individual models, the combinatorial hybrid model has been used in [12] for electricity demand forecasting. This article has illustrated that the combination of the two main techniques i.e., moving average procedure and adaptive particle swarm optimization algorithm is very effective for forecasting electricity demand. Another way to improve the performance of the forecasting model is to account for uncertainty in load demand [13]. The linear regression has been used to estimate the baseline demand, and then the uncertainty of the model has been estimated and analyzed further to improve the forecasted value of demand. Since demand response is important in modern networks, forecasting the electricity demand at residential level is significantly important. Air conditioning load is one of the most important loads at the residential level and [14] proposes a censored regression model to forecast future air conditioning load.

In [3], multiple linear regression approach is employed to forecast the electricity demand in medium-term period which ranges from several months to several years. In this timeframe, multiple regression model performs comparatively better than the commonly used models such as artificial neural network (ANN), Socioeconomic (S-E), and Box and Jenkins (B&J) models as reported in [15].

For building the multiple regression forecasting model, appropriate variables are needed to be included in the model [3], [16].

The consideration of some variables and the non-consideration of others can obviously affect the precision of the model and influence the accuracy of results. The authors in [17] have stated that the temperature, wind speed, relative humidity, and cloud cover are important to the changes of electricity consumption in Italy. In [18], it is reported that temperature, relative humidity, and wind speed are the key variables for analyzing the sensitivity of electricity and gas consumption in USA. The authors in [19] have restated the importance of these variables by building electricity demand models using five specific variables namely cooling degree days (CDD), heating degree days (HDD), humidity, wind speed and the enthalpy latent days (ELD) for different States of USA. In [15], on the other hand, it has been asserted that the crucial variables for building electricity demand forecasting models for England and Wales are not only temperature, humidity, and wind speed but also the sunshine hours, rainfall, and the GDP. The impacts of energy prices, daylight hours, trend variables, and temperature on electricity demand for residential and commercial sectors in the State of Maryland, USA have been highlighted in [20]. In [21], it has been advocated that the electricity demand in Greece depends not only on the temperature but also on population, GDP, energy intensity, and monthly seasonality of the electricity demand. Different customers have been considered to contribute to different consumption profiles between local areas in Denmark [22]. The authors in [23] have reported that the holiday period is one of the driving factors for forecasting the electricity demand in Japan besides HDD, CDD, and relative humidity. The authors in [24] have claimed that the variables such as GDP, population, import, export, and employment are important for forecasting demand of Turkey. These variables are employed to form different datasets feeding into 4 different models to forecast the demand. The results show that the model, which includes only four variables namely GDP, population, import, and export outperformed the other 3 models. It is noted that the same variables have been used in [25] to forecast the future demand of Turkey. In most of the studies reported in literature, the selection of independent variables has mainly been driven by the choice of the respective researchers and therefore it does not guarantee that the preferred set of variables is the best one. In addition, use of fewer variables makes the model weak while the use of numerous variables can be computationally intensive and may lead to problems associated with multicollinearity [26].

This paper proposes a novel combinatorial method using multicollinearity analysis and backward elimination regression analysis to select the optimized set of variables for an electricity demand forecasting model. First, in the multicollinearity analysis, the redundant explanatory variables will be excluded from the independent dataset. Subsequently, the backward elimination analysis will be adopted to remove the insignificant variables from the model. The developed model including the optimized variable-set includes only the significant variables and eliminates the redundant variables.

This paper is organized as follows: Section II gives the description of the proposed methodology. Section III introduces the mean degree days and adjustment factors. The empirical results and associated discussion is included in Section IV. Section V highlights the model verification and Section VI details the concluding remarks.

2. Proposed Forecasting Model for Electricity Demand

In this paper, an analytical technique has been developed, as depicted diagrammatically in Fig. 1, for building the electricity demand forecasting model. First, the prospective variables which can have significant impacts on the electricity demand are highlighted and the associated data are collated in the dataset 1. Second, the multicollinearity analysis has been conducted to reduce the collinearity by excluding the redundant predictors. The remaining dependent and independent variables including

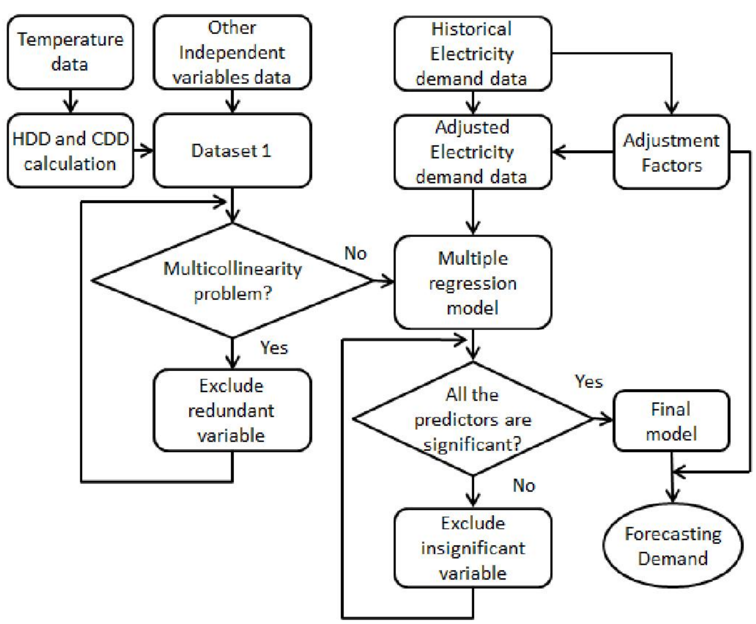

Figure 1: A conceptual diagram for building the electricity forecasting model.

electricity demand form a multiple regression model. Third, this model is examined with the backward elimination regression analysis to remove the insignificant variables. The final model is then modified with the aid of adjustment factors to obtain the forecasted demand.

2.1. Prospective Variables

Both socioeconomic and climatic changes may have considerable effect on the electricity demand. The socioeconomic variables such as population, gross state product (GSP) and electricity price are expected to have strong influence on the electricity demand [8], [10]. Any increase in GSP, indicating the economic growth, can lead to more electricity equipment being used. This leads to the high living standard and also high demand of electricity. In addition, an expansion of population can intuitively cause an increase in total demand. On the other hand, if the price of electricity rises, there could be a reduction in the power consumption. Also, the climatic variables may have significant influence on the electricity demand. Among all the climatic variables, temperature is reported to be the most important variable that can have significant impact on the electricity demand [19], [27]. Additionally, the other climatic variables such as wind speed, humidity, evaporation, rainfall, rainy days, solar exposure, and sunshine hours may have linear relationship with the electricity demand. All the above mentioned climatic and non-climatic variables are considered in this paper as potential predictors and thoroughly investigated.

2.2. Multicollinearity Analysis

Electricity demand can be affected by numerous variables however, it is not necessary to include all these variables in the forecasting model. It has been reported in [28] that linear relationship between the climatic variables and one predictor variable can represent the characteristics of other variables. Consequently, this predictor variable has little or even no new information contributing to the model and it becomes redundant. Furthermore, this redundant predictor variable can affect the precision of the model and lead to unreliable forecasting values due to the multicollinearity phenomenon [26]. As a result, employing the multicollinearity analysis is essential to reveal the relationship between the independent variables.

2.2.1. Analytical Approach

For a multiple regression equation as in (1), the multicollinearity between the predictors can cause large standard error for the coefficients $\beta_1, \beta_2, ..., \beta_m$, and may affect the model precision.

$$y = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + ... + \beta_m * x_m + \varepsilon \tag{1}$$

where, y is the response, $\beta_0, \beta_1, \beta_2, ..., \beta_m$ are the coefficients, $x_1, x_2, ..., x_m$ are the independent variables, *m* is the total number of independent variables, ε is the error term.

For demonstration, it is assumed that the coefficients $\beta_0, \beta_1, \beta_2, ..., \beta_m$ in (1) can be determined from the *n* observation of a dataset given in (2).

$$\begin{cases} y_1 = \beta_0 + \beta_1 * x_{11} + \beta_2 * x_{12} + ... + \beta_m * x_{1m} + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 * x_{21} + \beta_2 * x_{22} + ... + \beta_m * x_{2m} + \varepsilon_2 \\ ... \\ y_n = \beta_0 + \beta_1 * x_{n1} + \beta_2 * x_{n2} + ... + \beta_m * x_{nm} + \varepsilon_n \end{cases} \tag{2}$$

where $y_1, y_2, ..., y_n$ are the *n*-observed values of dependent variable data, x_{ij} *(i=1:n, j=1:m)* is the observation *i* of variable *j*, and $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$ are the observed errors.

The above variables can be written in matrix form as:

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1\, x_{11}\, x_{12} ... x_{1m} \\ 1\, x_{21}\, x_{22} ... x_{2m} \\ ... \\ 1\, x_{n1}\, x_{n2} ... x_{nm} \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ ... \\ \beta_m \end{bmatrix}, \mathbf{E} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ ... \\ \varepsilon_n \end{bmatrix} \tag{3}$$

Then, the model equation for all observations can be expressed as:

$$\mathbf{Y} = \mathbf{X} * \mathbf{B} + \mathbf{E} \tag{4}$$

where, **Y** is the response matrix of the model, **B** is the coefficient matrix, **X** is the independent variables matrix, and **E** is the error matrix.

From (4), with **X’** being the transpose matrix of **X**, one of the least square estimations of **B** can be calculated as $\hat{\mathbf{B}}$ which is presented in (5).

$$\hat{\mathbf{B}} = [(\mathbf{X}'{*}\mathbf{X})^{-1}] * \mathbf{X}'{*}\mathbf{Y} \qquad (5)$$

Since each element ε in (1) is treated as a random error, its expectation and variation is given in (6) and (7) respectively.

$$E(\boldsymbol{\varepsilon}) = 0 \qquad (6)$$

$$Cov(\boldsymbol{\varepsilon}) = \sigma^2 \qquad (7)$$

For the independent random errors $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$ in (2), the expectation and the covariance of the error matrix given in (3) can be rewritten as in (8), and (9) [29].

$$E(\mathbf{E}) = [0], \, or \, E\left(\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \ldots\\ \varepsilon_n\end{bmatrix}\right) = \begin{bmatrix}0\\0\\\ldots\\0\end{bmatrix} \qquad (8)$$

$$Cov(\mathbf{E}) = \sigma^2\mathbf{I}, \, or \, Cov\left(\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \ldots\\ \varepsilon_n\end{bmatrix}\right) = \sigma^2\begin{bmatrix}1\,0\,0\ldots0\\0\,1\,0\ldots0\\\ldots\\0\,0\,0\ldots1\end{bmatrix} \qquad (9)$$

where, each diagonal element of $Cov(\varepsilon)$ is the variance of each individual ε_i and has same value of σ^2. The off-diagonal elements of $Cov(\varepsilon)$ are the covariance between ε_i and ε_j and these off-diagonal elements are zero due to the independence of the errors.

From (4), the matrix $(\mathbf{X} * \mathbf{B})$ is fixed (although **B** is unknown), so the expectation and variation of **Y** can be calculated as in (10) and (11) respectively [29].

$$E(\mathbf{Y}) = E(\mathbf{X}{*}\mathbf{B} + \mathbf{E}) = \mathbf{X}{*}\mathbf{B} + E(\mathbf{E}) \qquad (10)$$

$$Cov(\mathbf{Y}) = Cov(\mathbf{X}{*}\mathbf{B} + \mathbf{E}) = Cov(\mathbf{E}) \qquad (11)$$

Considering (8) and (9), then (10) and (11) will become (12) and (13).

$$E(\mathbf{Y}) = \mathbf{X}{*}\mathbf{B} \qquad (12)$$

$$Cov(\mathbf{Y}) = \sigma^2\mathbf{I} \qquad (13)$$

Now with the estimation of coefficient matrix in (5), the expectation and the variation of $\hat{\mathbf{B}}$ can be expressed as in (14) and (15) respectively.

$$E(\hat{\mathbf{B}}) = E\left(((\mathbf{X}'{*}\mathbf{X})^{-1}) * \mathbf{X}'{*}\mathbf{Y}\right) \qquad (14)$$

$$Cov(\hat{\mathbf{B}}) = Cov\left(((\mathbf{X}'{*}\mathbf{X})^{-1}) * \mathbf{X}'{*}\mathbf{Y}\right) \qquad (15)$$

Applying the properties of expectation and variation calculation to (14) and (15), then we have (16) and (17).

$$E(\hat{\mathbf{B}}) = ((\mathbf{X}'{*}\mathbf{X})^{-1}) * \mathbf{X}'{*}E(\mathbf{Y}) \qquad (16)$$

$$Cov(\hat{\mathbf{B}}) = \left(((\mathbf{X}'{*}\mathbf{X})^{-1}) * \mathbf{X}'\right) * Cov(\mathbf{Y}) * \left(((\mathbf{X}'{*}\mathbf{X})^{-1}) * \mathbf{X}'\right)' \qquad (17)$$

Using the properties of transpose matrix to the right hand side of equation (17) and it will results in (18)

$$Cov(\hat{\mathbf{B}}) = ((\mathbf{X}'{*}\mathbf{X})^{-1}) * \mathbf{X}'{*}Cov(\mathbf{Y}) * \mathbf{X}{*}((\mathbf{X}'{*}\mathbf{X})^{-1}) \qquad (18)$$

By substituting (12) into (16), and (13) into (18) and doing some requisite matrix manipulation, the expectation and the variation of $\hat{\mathbf{B}}$ can be derived as in (19) and (20) respectively.

$$E(\hat{\mathbf{B}}) = ((\mathbf{X}'{*}\mathbf{X})^{-1}) * \mathbf{X}'{*}\mathbf{X}{*}\mathbf{B} = \mathbf{B} \qquad (19)$$

$$Cov(\hat{\mathbf{B}}) = ((\mathbf{X}'{*}\mathbf{X})^{-1}) * \mathbf{X}'{*}(\sigma^2){*}\mathbf{I}{*}\mathbf{X}{*}((\mathbf{X}'{*}\mathbf{X})^{-1}) = (\sigma^2) * ((\mathbf{X}'{*}\mathbf{X})^{-1}) \qquad (20)$$

Equations (19) and (20) show that the expectation of $\hat{\mathbf{B}}$ is exactly the same to $\mathbf{B}$ and the variance of $\hat{\mathbf{B}}$ is proportional to the population variance σ^2 with an amount of $(\mathbf{X}'^*\mathbf{X})^{-1}$. Setting the matrix, $\mathbf{C} = (\mathbf{X}'^*\mathbf{X})^{-1}$, then the variation of $\mathbf{B}$ is given in (21).

$$Cov\left(\hat{\mathbf{B}}\right) = (\sigma^2)^*\mathbf{C} \qquad (21)$$

The off-diagonal elements of matrix $\mathbf{C}$ are related to the covariance between the coefficients, and the diagonal elements are related to the variance of the coefficients in the model as given in (22) [30].

$$Var(\hat{\beta}_j) = (\sigma^2)^* c_{jj} \qquad (22)$$

where,

$$c_{jj} = \frac{1}{(1-R_j^2)^* \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2} \qquad (23)$$

where, R_j^2 is the coefficient of determination of the regression of x_j on all other independent variables in the dataset, $\bar{x}_j$ is the mean value of the observation of x_{ij}, and $\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$ is the denominator of the formula for the variance of the regression coefficient in a simple linear regression.

Substituting (23) into (22), the variance of the coefficient of variable x_j becomes:

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2} * \frac{1}{(1-R_j^2)} \qquad (24)$$

It is noted from (24) that $\sigma^2 / \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$ is the variance of coefficient $\hat{b}_j$ if there is only one variable x_j in the dataset, and it is independent from the relationship between x_j and the other predictor variables. On the other hand, the latter part $1/(1-R_j^2)$ is the factor which depends on the linear relationship between x_j and the other independent variables [x_1, x_2, ... x_{j-1}, x_{j+1}, ..., x_m]. This part will be introduced as the variance inflation factor as in Subsection 2.2.2.

2.2.2. Variance Inflation Factor

In order to determine the multicollinearity problem in a dataset that has *m* independent variables, e.g. [x_1, x_2, ... x_j, ..., x_m], one of the following methods can be used: Pearson's correlation matrix of predictor variables; eigenvalues of the matrix $[\mathbf{X}'^*\mathbf{X}]$; or variance inflation factor (*VIF*) [26]. The Pearson's correlation matrix has a limitation of establishing relationship between only two independent variables at a time. Utilizing the eigenvalues can help determine the linear relationship among more than two variables but it could be computationally intensive, especially with increase in the number of independent variables. *VIF* is an effective approach for multicollinearity assessment since it overcomes the lacunas of the above mentioned methods. In addition, VIF calculations are straightforward and comprehensive; the higher the value of VIF, the higher the collinearity is between the related variables. Accordingly, VIF has been used in the proposed model to identify multicollinearity. VIF_j of one predictor x_j is calculated based on the linear relationship between the predictor x_j and the other independent variables [x_1, x_2, ... x_{j-1}, x_{j+1}, ..., x_m] as in (25) [26].

$$VIF_j = \frac{1}{(1-R_j^2)} \qquad (25)$$

where, R_j^2 is the coefficient of determination of the regression of x_j on all other independent variables in the dataset [x_1, x_2, ... x_{j-1}, x_{j+1}, ..., x_m].

In the case when there is no multicollinearity between the variables in the dataset, the R_j^2 equals to zero, and VIF_j equals to 1. If the multicollinearity exists, the VIF_j progresses to a number that is much greater than 1. In [31], the *VIF* value of 5 is used for examining the multicollinearity phenomenon. It is mentioned that VIF_j is equal to 5, then the value for R_j^2 is found to be 0.8 i.e., eighty percent of the variable x_j can be represented by the other independent variables highlighting the possibility of multicollinearity.

2.3. Backward Regression Analysis

After excluding the redundant variables with the aid of multicollinearity analysis, the generalized regression equation given in (1) can be rewritten as in (26) for electricity demand forecasting.

$$D = c_0 + \sum_{j=1}^{m} (c_j * x_j) + \varepsilon \qquad (26)$$

where D is the electricity demand, c_0 is the constant, x_j are independent variables, c_j are coefficients of variables x_j, ε is the error term, and m is the number of variables included in the model.

In this model, some variables may be insignificant, and the insignificant variables should be eliminated from the model by backward elimination regression analysis. In this process, the p-value, which can be used to estimate the significance of variables of each parameter, is estimated. A p-value with a range between 0 and 1 is used to test the null hypothesis that the coefficient c_j is equal to zero. If the p-value is close to 1, the hypothesis is true and the probability of c_j being zero is very high and the consequent variable x_j becomes insignificant. If the p-value is low, the predictor variable x_j becomes significant in the model. The criterion for the p-value is commonly set as 0.05 [32], which indicates that any variables with a p-value of less than 0.05 should be significant in the model.

3. Average Degree Days and Adjustment Factors

Numerous studies have represented the temperature by using degree days [8], [15], [17], [20], [21], [23], as degree days can represent linear relationship with the electricity demand. In this paper, the main purpose is to forecast the electricity demand so the average degree days are more suitable. In addition, adjustment factors are used to isolate the influence of climatic factors on the electricity demand.

3.1. Balance Point Temperature

As discussed earlier, temperature is considered to be one of the most important variables affecting the electricity demand. The dependency of the demand on the temperature however is not a linear relationship, but is the V-shaped curve for the ideal case [8], [20]. The point at which the electricity demand is at its minimum is called the balance point temperature T_b.

In practice, the relationship between electricity demand and temperature is not perfectly smooth like the ideal case. The balance point temperature T_b, however, can be estimated by using the trend-lines [21], [23]. In Fig. 2, the monthly electricity demand data and temperature data for 12 years from year 1999 to 2010 for the state of New South Wales (NSW), Australia were used to plot the scatter and trend-line to evaluate the balance point temperature. As shown in Fig. 2, the balance point temperature can be determined with the help of trend-line equation which is found to be 19.5°C.

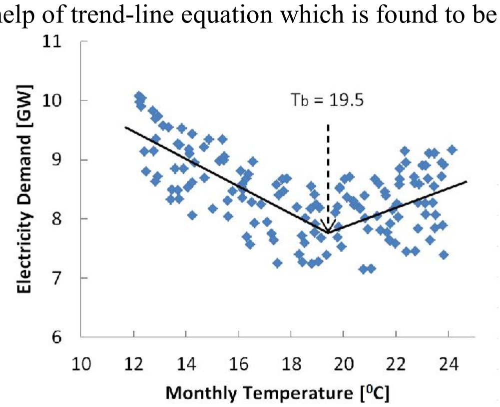

Figure 2: Relationship between electricity demand and temperature in NSW, Australia from year 1999 to 2010.

3.2. Average Degree Days

If the average temperature of a day i is T_i then the cooling degree of that day (CDD_i) is given by:

$$CDD_i = \begin{cases} (T_i - T_b) & \text{if } (T_i > T_b) \\ 0 & \text{if } (T_i < T_b) \end{cases} \qquad (27)$$

From (27), in day i, when T_i is greater than T_b, the CDD_i equals the difference between T_i and T_b. Since T_i is lower than T_b, $CDD_i = 0$ due to no cooling demand required. The average cooling degree days (CDD) in one month can then be calculated by summing up all the degree days in that month and can be expressed the average CDD as in (28).

$$CDD = \frac{1}{N} * \sum_{i=1}^{N} CDD_i \qquad (28)$$

where N is the number of days in one month.

Similarly, the heating degree of one day (HDD_i) and the average heating degree days in one month (HDD) can be identified as in (29) and (30) respectively.

$$HDD_i = \begin{cases} (T_b - T_i) & if\ (T_i < T_b) \\ 0 & if\ (T_i > T_b) \end{cases} \tag{29}$$

$$HDD = \frac{1}{N} * \sum_{i=1}^{N} HDD_i \tag{30}$$

In case of average HDD, the lower the value of temperature T_i and the longer it lasts, the bigger the HDD value. If T_i is greater than T_b, no heating is required.

The balance point temperature is used to calculate the CDD and the HDD for each month between the year 1999 to 2010. The CDD and HDD values are presented in Fig. 3 and Fig. 4 respectively.

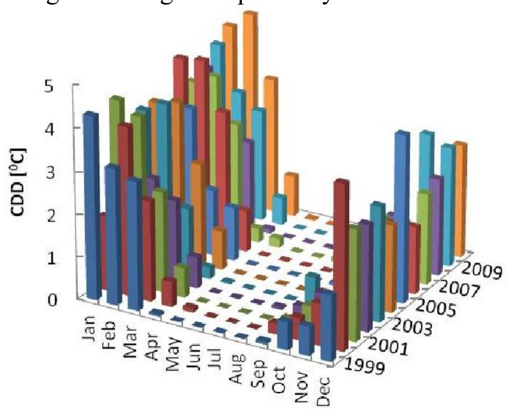

Figure 3: Estimated average cooling degree days in NSW, Australia for each month from year 1999 to 2010.

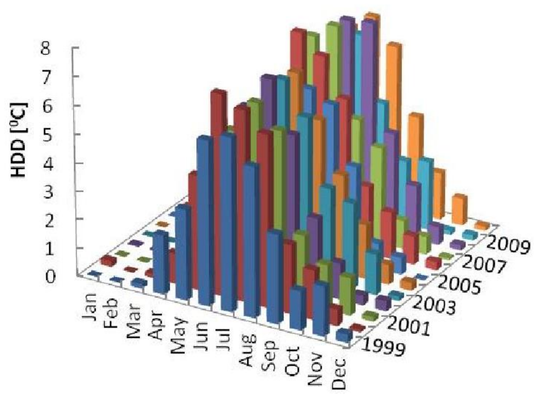

Figure 4: Estimated average heating degree days in NSW, Australia for each month from year 1999 to 2010.

From Figs. 3 and 4, it can be seen that the trend of the variation of the HDD is likely to be opposite to the trend of the variation of the CDD. This can be explained by the repetition of different seasons every year, and the temperature pattern in a particular season could be different to that of the other seasons. The two seasons with the most significant influence on the CDD and the HDD are summer and winter. In the summer time i.e., from December to February in Australia, the CDD reaches to a very high value due to the dominance of hot weather, but the HDD reduces to nearly zero. Contrarily, in the winter time, i.e., from June to August, the CDD is close to zero because of the dominance of cold weather, while the HDD is at its highest. From Figs. 3 and 4, it is noted that the maximum value of the HDD is always higher than that of the CDD. This highlights the fact that the winter has more extreme weather conditions than that of the summer in NSW. Accordingly, it is expected that the electricity demand will depend predominantly on temperature in NSW, Australia.

3.3. Adjustment Factors

Adjustment factors have been used in [15], [18] for building the electricity forecasting model. The main purpose of the adjustment factors is to isolate the influence of the climate factors on the electricity consumption. First, the adjustment factor F_j for each year is calculated using (31), and then the monthly data is adjusted as in (32). This adjusted electricity demand $E_{adj}(i,j)$ will be used to build the forecasting model. This model is then multiplied by the adjustment factor F_j in each year to get the prediction value of electricity demand.

$$F_j = \frac{\sum_{i=1}^{12} E(i,j)}{E_{av}} \tag{31}$$

$$E_{adj}(i,j) = \frac{E(i,j)}{F_j} \tag{32}$$

where F_j is the adjustment factor of year j; E_{av} is the average electricity demand in the study period; $E(i,j)$ is the electricity demand in the month i of year j; and $E_{adj}(i,j)$ is the adjusted electricity demand in month i for year j.

Figs. 5 and 6 depict the relationship between adjusted electricity demand with respect to CDD and HDD respectively. From these two figures, it can be seen that the fit with the electricity demand and HDD ($R^2 = 0.961$) is better than that of CDD ($R^2 = 0.546$), and the dependence of the demand on HDD is stronger due to the greater incline of the trend-line. Accordingly, HDD is expected to have significant impact on the electricity demand of NSW.

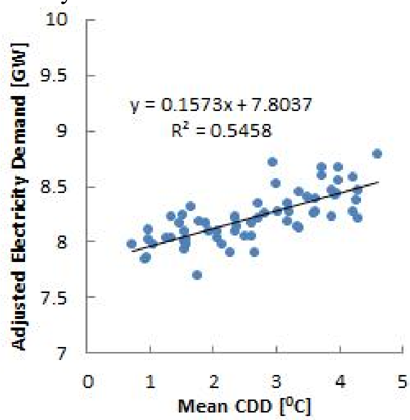

Figure 5: Relationship between monthly electricity demand and CDD from 1999 to 2010.

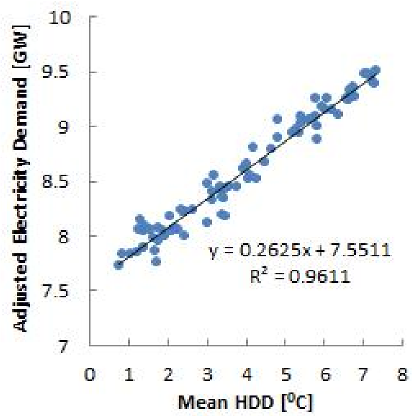

Figure 6: Relationship between monthly electricity demand and HDD from 1999 to 2010.

4. Results and Discussion

A case study has been conducted in the paper with the aid of historical data from the state of NSW, Australia for the year 1999 to 2010. The data associated with electricity demand and electricity price (*Pri*) including industrial, residential and commercial sectors has been collected from the Australian energy market operator [33]. These datasets are available for every half an hour and has been collated on daily and monthly basis for the proposed studies. The annual data of population (*Pop*), and gross state product (*GSP*) are accessible from Australian bureau of statistics [34], and the monthly data during each year has been assumed to be incrementally changing as per the yearly indices. The climatic parameters at Sydney airport station [35] are assumed to be representing the entire state of NSW, as around 75% of population of NSW are in Sydney and the surrounding areas. Consequently, the monthly data of the climatic variables namely average humidity percentage (*Hum*), number of clear days (*CleD*) in one month, number of cloudy days (*CloD*) in one month, mean rainfall (*RaF*) in one month, average wind speed (*Win*), number of rainy days in one month (*RaD*), average sunshine hours (*Sun*), monthly mean solar exposure (*Sol*), average evaporation (*Eva*), mean maximum temperature (*MaT*), and mean minimum temperature (*MiT*) have been acquired at the Sydney airport station for the purpose of the analysis. Furthermore, the SPSS and MATLAB have been employed to develop a statistical tool to perform the data cleansing and the requisite analyses.

4.1. Multicollinearity Analysis

From the calculated values of *CDD* and *HDD* along with the data of the other independent variables, an independent dataset is formed, and it is called as set 1. This dataset is then used in the multicollinearity analysis, and the process is shown as in Table I. In the first step of the analysis, the variable *MiT* has the biggest value of *VIF*, which is 587.4; therefore, it will be removed (remd) from set 1, and then the set 2 is formed. In the second step, the *MaT* with the highest *VIF* of 130.7 is removed from the set 2 to form the set 3. The process continues until set 7 and then stops, as all the remaining variables have the *VIF* values less than 5 which satisfy the multicolinearity examining condition discussed in Section 2.2.

Table I: Results Obtained using Multicollinearity Analysis

Variable name	Variance Inflation Factor (VIF) of the predictors in different datasets						
	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7
CDD	191.7	23.1	4.6	3.6	3.6	3.3	3.0
HDD	490.2	57.9	5.9	4.7	4.7	4.0	3.1
Hum	6.9	5.9	5.3	4.1	4.0	4.0	3.2
RaD	3.5	3.4	3.3	3.3	3.2	3.1	3.1
GSP	24.1	23.0	22.8	22.2	1.2	1.2	1.2
Pri	1.3	1.3	1.3	1.2	1.2	1.2	1.2
RaF	2.2	2.2	2.2	2.1	2.1	2.0	2.0
Win	3.5	3.5	3.2	2.8	2.6	2.1	2.0
CloD	5.1	5.0	5.0	4.9	4.7	4.5	3.1
CleD	4.0	3.8	3.7	3.7	3.7	3.4	3.3
Sun	10.8	9.4	9.3	9.3	8.4	6.0	remd
Sol	25.4	25.4	25.1	10.3	9.2	remd	remd
Pop	22.8	22.3	22.3	22.3	remd	remd	remd
Eva	42.8	40.1	37.3	remd	remd	remd	remd
MaT	222.7	130.7	remd	remd	remd	remd	remd
MiT	587.4	remd	remd	remd	remd	remd	remd

It is noted that the *VIF* values of *MiT* and *HDD* are very high, i.e., 587.4 and 490.2 respectively, in the set 1, but only *MiT* with highest *VIF* is excluded from the dataset. These high *VIF* values are experienced due to the strong linear relationship between *MiT* and *HDD*, which is verified by applying the Pearson's correlation to this pair of variables. The correlation between these variables is found to be 0.952. *MiT*, however, has strong linear relationship with the other variables in contrast to *HDD*. This is the reason why *MiT* should be removed from the data set in the first place. In the second step (set 2), the *VIF* value of *HDD* vastly reduces from 490.2 to 57.9 and even less than the *VIF* value of *MaT*, which is 130.7.

4.2. Backward Elimination Regression Analysis

Backward elimination analysis starts with model 1 (mod 1) which includes all the remaining independent variables after conducting multicollinearity analysis. The process of elimination is illustrated in Table II.

In the first step, the variable *CleD* with the highest p-value of 0.933 is removed from the mod 1, and mod 2 is formed based on the remaining variables. In the second step, the *CloD* is excluded because of the highest p-value of 0.709, and so on. The process continues until the seventh step (mod 7), where all the p-values are found to be less than 0.05. The variables which retain their place till the end are *CDD*, *HDD*, *Hum*, and *RaD*. These could be classified as the most significant variables and will be used to forecast the electricity demand.

Table II: Results obtained using Backward Regression Analysis

Variable name	Significant level of the independent variables (p-value of the coefficients) in different models						
	mod 1	mod 2	mod 3	mod 4	mod 5	mod 6	mod 7
CDD	0.000	0.000	0.000	0.000	0.000	0.000	0.000
HDD	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Hum	0.004	0.003	0.001	0.000	0.000	0.000	0.000
RaD	0.100	0.084	0.086	0.051	0.017	0.015	0.012
GSP	0.134	0.120	0.110	0.119	0.125	0.072	remd
Pri	0.484	0.474	0.431	0.423	0.409	remd	remd
RaF	0.611	0.613	0.593	0.515	remd	remd	remd
Win	0.697	0.681	0.695	remd	remd	remd	remd
CloD	0.710	0.709	remd	remd	remd	remd	remd
CleD	0.933	remd	remd	remd	remd	remd	remd

The values of regression term (*R*), coefficient of determination (R^2) and adjusted coefficient of determination (R^2_{adj}) of the model 1 and model 7 are examined in Table III. The R and R^2 values show the fitness of the modeled curve to the actual demand data, but the R_{adj}^2 indicates the fitness of the model associated with the freedom of the model (or the number of variables in the model). From Table III, it can be seen that before processing the backward elimination regression (in mod 1), the values of R and R^2 are greater than those at the final stage of the analysis (in mod 7) because there is less number of variables in mod 7 than that in mod 1. The R_{adj}^2, on the other hand, has been improved through the backward elimination regression process from 0.941 (in mod 1) to 0.942 (in mod 7). Moreover, the difference between R_{adj}^2 and R^2 in model 7 is smaller than that in model 1. This confirms that the backward regression analysis performs well even with the inclusion of less number of variables.

Table III: Coefficient of Determination of Model 1 and Model 7

Model	R	R^2	R_{adj}^2
Model 1	0.974	0.948	0.941
Model 7	0.972	0.944	0.942

4.3. Final Forecasting Model

The model 7 in Table III is employed as final model for forecasting electricity demand. The coefficient, standard error and t-statistic (t-ratio) values of each variable in this model are given in Table IV.

Table IV: Variables in the Final Model

Variables	Coefficient	Standard error	t-ratio
(Constant)	6.892	0.179	38.6
CDD	0.211	0.014	15.2
HDD	0.268	0.008	31.6
Hum	0.011	0.003	3.9
RaD	-0.011	0.004	-2.6

1) Coefficient of Variables:

Coefficients given in Table IV are the partial coefficient of each variable in the model. From these values, and based on (26), the final model can be established as in (33) and the forecasting value can be determined as in (34).

$$D_M = 6.892 + 0.211 * \mathrm{CDD} + 0.268 * \mathrm{HDD} + 0.011 * Hum + 0.011 * RaD \quad (33)$$

$$E_F = D_M * F_j \quad (34)$$

where D_M is the monthly electricity demand before incorporating adjustment, E_F is the forecasted demand, F_j is the adjustment factor.

2) Standard Error:

The standard error indicates the interval confidence of the coefficients. Assuming that the distribution of the constant associated with CDD follow normal distribution, at the level of 95% confidence, the percentage points of the t distribution are estimated to be 1.99. Thus, with 95% confidence, the coefficient of CDD in Table IV lies between (0.211 - 1.99*0.014 to 0.211 + 1.99*0.014) = 0.183 to 0.239. It indicates that the electricity demand may increase from 0.183 to 0.239 GW when CDD increases by one degree with the assumption that other variables keep constant.

3) t-ratio:

The t-ratio in this study is equal to the coefficients divided by the standard error [32]. The absolute value of these t-ratio values thus, should be greater than 2 to ensure the goodness of the coefficients. As can be seen in the Table IV, all the t-ratios are greater than 2 or less than -2, confirming the goodness of the coefficients. With reference to Table II, it is noted that the p-values of the CDD, HDD and Hum are too small. However, based on the t-ratio indicators in Table IV, it can be concluded that, HDD is the most significant variable in the model with the highest t-ratio.

5. Model Validation

In this Section, the modeled values and the historical data are plotted in the same graph for the total time period to conduct a comparative study. Furthermore, the percentage error is plotted and mean absolute percentage error (MAPE) is calculated to confirm the accuracy of the model. Different divisions of available historical data into training and testing dataset can be formed for verification, and the results would be similar due to very high value of R_{adj}^2 of the model as presented in Table III. This paper verifies the model with a training period from the year 1999 to 2005, and prediction period from the year 2006 to 2010.

5.1. Validation of the Training Period

The comparison of predicted data and historical data for the training period from year 1999 to year 2005 is depicted in Fig. 7. It can be seen that the predicted values are very close to the historical data. Especially, in the winter season, the deviation between the two values is relatively small due to the strong relationship between the electricity demand and the HDD as shown in Fig. 6. The forecasted values are underestimated for the summer season of the year 2004 and 2005 due to the sudden increase in the actual demand in these time intervals.

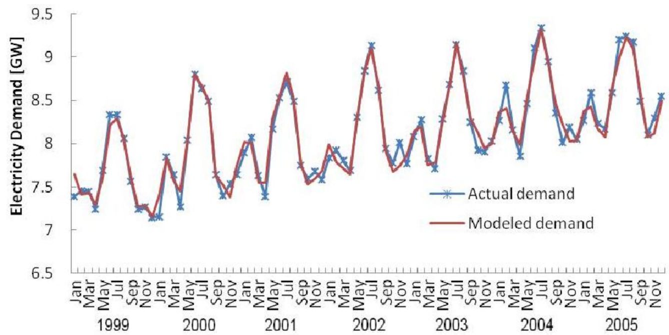

Figure 7: Comparison between modeled and actual electricity demand for the period 1999 to 2005.

It can be seen from Fig. 7 that there is a small decrease in the predicted value of the demand as compared to the actual value of the demand in the month of December for each year. The month of December is the beginning of the summer season with predominantly hot weather (i.e., soaring temperatures), and the demand is expected to be high due to the associated cooling requirement. Therefore, the reduction of actual demand in summer can only be experienced due to some external events such as the holiday period. The summer holidays may lead to sudden decrement in the demand and badly affect the forecasting.

For the training period (1999-2005), the variation of the percentage error is shown in Fig. 8. It can be seen that the error between the modeled values and the actual demand is relatively small, and the maximum error is less than 4%. The Durbin-Watson statistic for the model is calculated and found to be 2.01 highlighting that there is no autocorrelation for the proposed forecasting model in the training period. Furthermore, the MAPE of the model is estimated to be 1.02% indicating that the modeled demand fits very well with the historical data.

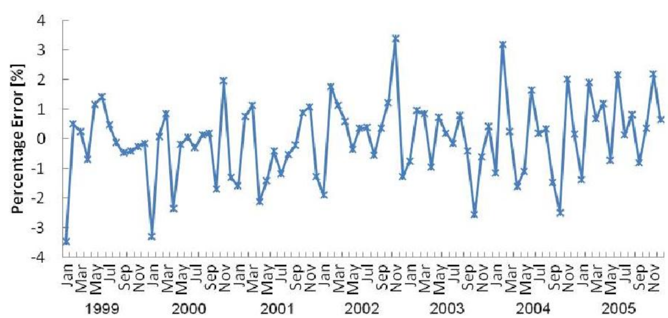

Figure 8: Variation of the percentage errors between modeled and actual electricity demand for the period 1999 to 2005.

5.2. Validation of the Prediction Period

The capability of the model in forecasting the electricity demand is evaluated by applying the model to predict the demand for the year 2006 to 2010. The comparison between the modeled values and the actual demand is shown in Fig. 9. It can be seen that the peak demand in the winter season fits very well with the forecasted values. The lower peaks demand in year 2009 and 2010 are expected due to the warmer winter in recent years. With the warmer winter, the heating requirement in NSW is declined thereby resulting into the decrement of the peak electricity demand.

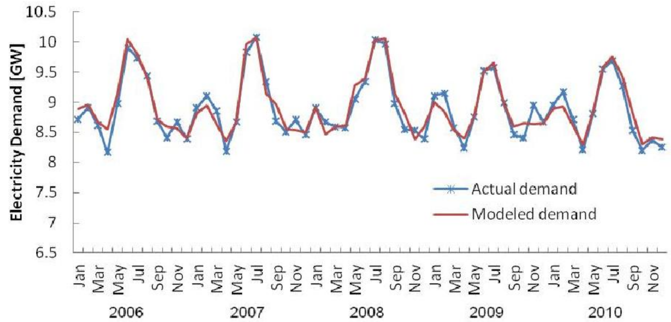

Figure 9: Comparison between modeled and actual electricity demand for the period 2006 to 2010.

Fig. 10 introduces additional details associated with the variation of the percentage error in the prediction period. The MAPE value of the model is found to be 1.35%, and the value of the Durbin-Watson test in this case is obtained as 1.75. As a result, the autocorrelation may exist due to the substantial variation of the demand in the summer time in recent years.

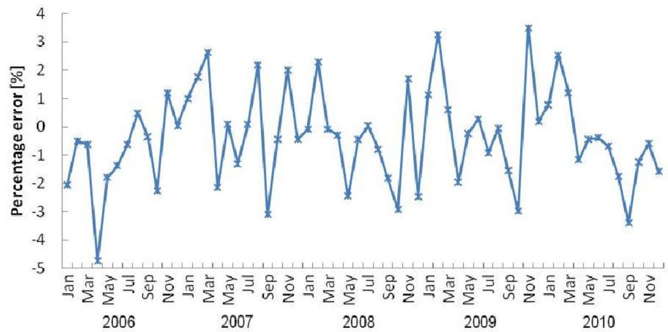

Figure 10: Variation of the percentage errors between modeled and actual electricity demand for the period 2006 to 2010.

The MAPE values for each month in both training and prediction periods are given in Table V. It can be seen from Table V that the MAPE values are lower in June and July as compared to the other months. This may be due to the stronger dependence of electricity demand on temperature.

Table V: MAPE Values in Different Months

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Training period	1.926	1.306	0.734	1.355	0.809	0.897	0.397	0.466	0.561	1.370	1.641	0.743
Prediction period	1.009	2.071	1.022	2.060	1.000	0.761	0.465	1.054	2.040	1.967	1.795	0.952

5.3. Model Comparison

This Section discusses the goodness of the proposed model by comparing it to the other 3 models.

5.3.1. C-D Model

The variables *CDD* and *HDD* are expected to have strong impacts on electricity demand since they are temperature dependent. Besides the V-shape relationship mentioned in Section 3.1, which is widely used in the literature, the U-shape can also be used as another effective way to represent the relationship between demand and temperature [15], [17]. U-shape relationship considers a comfort band in which electricity demand is independent of temperature. In this Subsection, the U-shape relationship is used to derive the CDD and HDD, and then the obtained values are used to test in the proposed model. The U-shape representing the relationship of demand and temperature in NSW, Australia is shown as in Fig.11.

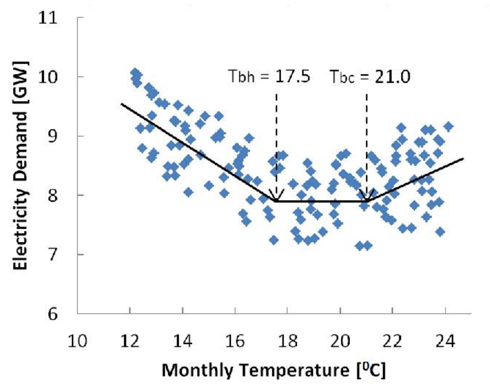

Figure 11: U-shape representing the relationship between electricity demand and temperature in NSW, Australia from year 1999 to 2010.

In order to calculate the average degree days, T_{bh} , T_{bc} are introduced as the threshold for calculating CDD_i and HDD_i, respectively. For the data acquired from State of NSW Australia, T_{bh} , T_{bc} are selected as 17.5°C and 21.0°C, respectively as shown in Fig. 11. The process of *CDD*, *HDD* calculation is similar to that mentioned in Section 3.2 and can be represented using (35) and (36) respectively.

$$CDD_i = \begin{cases} (T_i - T_{bc}) & if\ (T_i > T_{bc}) \\ 0 & if\ (T_i < T_{bc}) \end{cases} \qquad (35)$$

$$HDD_i = \begin{cases} (T_{bh} - T_i) & if\ (T_i < T_{bh}) \\ 0 & if\ (T_i > T_{bh}) \end{cases} \qquad (36)$$

The calculated *CDD* and *HDD* along with the other independent variables are then used in backward elimination regression analysis after eliminating multicollinearity between the variables. The relevant results are included in Table VI. It can be seen that the variables included in C-D model are the same as that of the proposed model (given in Table IV). The parameters such as coefficient, standard error, and t-ratio of the two models are different due to the changes in *CDD* and *HDD*.

Table VI: Variables in the C-D Model

Variables	Coefficient	Standard error	t-ratio
(Constant)	7.286	0.176	41.4
CDD	0.232	0.021	11.0
HDD	0.321	0.010	33.0
Hum	0.010	0.003	3.5
RaD	-0.009	0.004	-2.2

5.3.2. B-R Model

In order to emphasize the importance of multicollinearity analysis, another model (named B-R model) was built only based on the backward regression analysis until four most important variables are remained. The parameters of B-R model are given in Table VII. The variables included in the B-R model are *CDD*, *Eva*, *MaT*, *MiT*. It is noted that there is only one common variable between this model and the proposed model which is *CDD*; the remaining variables are different from that of the proposed model.

Table VII: Variables in The B-R model

Variables	Coefficient	Standard error	t-ratio
(Constant)	12.237	0.259	47.321
CDD	0.502	0.024	21.009
Eva	-0.002	0.001	-3.246
MaT	-0.088	0.020	-4.421
MiT	-0.147	0.013	-11.216

5.3.3. C-L Model

For further comparison, C-L model (model 3 proposed in [15]) is used to compare with the other 3 models namely proposed model, C-D model, and B-R model built in this paper. There are 7 input variables for C-L model, which are *CDD, HDD, Hum, Win, Sol, RaF, GSP*. The significant level (i.e., p-value) and t-ratio of each variable in the model is given in the Table VIII.

Table VIII: Significant Level and t-ratio of Each Variable in C-L model

Vairable	Constant	CDD	HDD	Hum	Win	Sol	RaF	GSP
p-value	0.000	0.000	0.000	0.442	0.169	0.204	0.824	0.023
t-ratio	13.802	10.528	30.695	0.772	-1.388	-1.281	-0.223	2.320

5.3.4. Comparative Analysis

The comparative analysis of all the 4 models in relation to demand prediction is given in Table IX. It can be seen that the proposed model outperforms the other models in term of R_{adj}^2 and MAPE values. In addition, the average residual of the proposed model is relatively small confirming the zero mean of the residuals as in (6).

Table IX: Comparative Analysis of Different Models for Demand Prediction

	Proposed model	C-D model	B-R model	C-L model
R_{adj}^2	0.909	0.895	0.869	0.875
MAPE	1.350	1.521	1.601	2.066
Sum of residuals	-1.940	-2.196	-1.493	-7.823
Average residual	$-3.23*10^{-2}$	$-3.66*10^{-2}$	$-1.49*10^{-2}$	$-2.30*10^{-1}$
Residual sum of square	1.357	1.602	1.892	2.705
Durbin-Watson statistic	1.749	1.617	1.347	0.875

6. Conclusion

In this paper, a robust regression model for forecasting the electricity demand is developed based on multicollinearity and backward elimination processes. The multicollinearity analysis helps to eliminate the variables which are highly related to the other independent variables from the dataset, and the backward elimination regression analysis excludes the insignificant variables from the model. Use of these processes makes the regression model robust and effective for forecasting the electricity demand from climatic variables. The proposed method is tested and validated, and the performance is evaluated in the Australian context. The results show that the electricity demand predominantly depends on the CDD, HDD, humidity and the number of rainy days. The robustness of the model is tested by assessing the impact of climatic variables on forecasting electricity demands for different months of the prediction period. Results have proved that the proposed model can predict the electricity demand with very low prediction error. Moreover, the other 3 models namely C-D model, B-R model and C-L model are built to compare their performance with the proposed model for validation purposes. Based on the obtained results, it is noted that the proposed model outperforms the other 3 models in terms of predicting the future electricity demand.

Acknowledgements

This work is supported by Hong Duc, Thanh Hoa – UOW research scholarship program.

References

[1] Pearce K, Holper P, Hopkins M *et al.* Climate change in Australia: Technical Report. In: CSIRO and the Australian Bureau of Meteorology; 2007.
[2] Suganthi L, Samuel AA. Energy models for demand forecasting—A review. Renewable and Sustainable Energy Reviews 2012; 16:1223-1240.
[3] Al-Alawi SM, Islaw SM. Principles of electricity demand forecasting. I. Methodologies. Power Engineering Journal 1996; 10:139-143.
[4] Rui Z, Zhao Yang D, Yan X *et al.* Short-term load forecasting of Australian National Electricity Market by an ensemble model of extreme learning machine. Generation, Transmission & Distribution, IET 2013; 7:391-397.
[5] Al-Hamadi HM, Soliman SA. Fuzzy short-term electric load forecasting using Kalman filter. Generation, Transmission and Distribution, IEE Proceedings- 2006; 153:217-227.
[6] Han XS, Han L, Gooi HB, Pan ZY. Ultra-short-term multi-node load forecasting - a composite approach. Generation, Transmission & Distribution, IET 2012; 6:436-444.
[7] Imtiaz AK, Mariun NB, Amran MMR *et al.* Evaluation and Forecasting of Long Term Electricity Consumption Demand for Malaysia by Statistical Analysis. In: Power and Energy Conference, 2006. PECon '06. IEEE International. 2006. pp. 257-261.
[8] Ahmed T, Muttaqi KM, Agalgaonkar AP. Climate change impacts on electricity demand in the State of New South Wales, Australia. Applied Energy 2012; 98:376-383.

[9] Parkpoom SJ, Harrison GP. Analyzing the Impact of Climate Change on Future Electricity Demand in Thailand. Power Systems, IEEE Transactions on 2008; 23:1441-1448.

[10] Lam JC. Climatic and economic influences on residential electricity consumption. Energy Conversion and Management 1998; 39:623-629.

[11] Howden SM, Crimp S. Effect of climate and climate change on electricity demand in Australia. 2001.

[12] Zhu S, Wang J, Zhao W, Wang J. A seasonal hybrid procedure for electricity demand forecasting in China. Applied Energy 2011; 88:3807-3815.

[13] Walter T, Price PN, Sohn MD. Uncertainty estimation improves energy measurement and verification procedures. Applied Energy 2014; 130:230-236.

[14] Horowitz S, Mauch B, Sowell F. Forecasting residential air conditioning loads. Applied Energy 2014; 132:47-55.

[15] Hor C-L, Watson SJ, Majithia S. Analyzing the impact of weather variables on monthly electricity demand. Power Systems, IEEE Transactions on 2005; 20:2078-2085.

[16] Islam SM, Al-Alawi SM. Principles of electricity demand forecasting. II. Applications. Power Engineering Journal 1997; 11:91-95.

[17] Apadula F, Bassini A, Elli A, Scapin S. Relationships between meteorological variables and monthly electricity demand. Applied Energy 2012; 98:346-356.

[18] Sailor DJ, Muñoz JR. Sensitivity of electricity and natural gas consumption to climate in the U.S.A.—Methodology and results for eight states. Energy 1997; 22:987-998.

[19] Sailor DJ. Relating residential and commercial sector electricity loads to climate—evaluating state level sensitivities and vulnerabilities. Energy 2001; 26:645-657.

[20] Ruth M, Lin A-C. Regional energy demand and adaptations to climate change: Methodology and application to the state of Maryland, USA. Energy Policy 2006; 34:2820-2833.

[21] Mirasgedis S, Sarafidis Y, Georgopoulou E *et al.* Modeling framework for estimating impacts of climate change on electricity demand at regional level: Case of Greece. Energy Conversion and Management 2007; 48:1737-1750.

[22] Andersen FM, Larsen HV, Gaardestrup RB. Long term forecasting of hourly electricity consumption in local areas in Denmark. Applied Energy 2013; 110:147-162.

[23] Akil YS, Miyauchi H. Elasticity coefficient of climatic conditions for electricity consumption analysis. In: Power System Technology (POWERCON), 2010 International Conference on. 2010. pp. 1-6.

[24] Kankal M, Akpınar A, Kömürcü Mİ, Özşahin TŞ. Modeling and forecasting of Turkey's energy consumption using socio-economic and demographic variables. Applied Energy 2011; 88:1927-1939.

[25] Kavaklioglu K. Modeling and prediction of Turkey's electricity consumption using Support Vector Regression. Applied Energy 2011; 88:368-375.

[26] Alin A. Multicollinearity. Wiley Interdisciplinary Reviews: Computational Statistics 2010; 2:370-374.

[27] Oliveira MO, Marzec DP, Bordin G *et al.* Climate change effect on very short-term electric load forecasting. In: PowerTech, 2011 IEEE Trondheim. 2011. pp. 1-7.

[28] Guan L, Yang J, Bell JM. Cross-correlations between weather variables in Australia. Building and Environment 2007; 42:1054-1070.

[29] Christensen R. Analysis of Variance, Design and Regression: Applied statistical methods. 2-6 Boundary Row, London SE1 8HN, UK: Chapman & Hall; 1996.

[30] Freund RJ, Wilson WJ, Sa P. Regression Analysis : Statistical Modeling of a Response Variable (2nd Edition). Burlington, MA, USA Academic Press 03/2006

[31] Rogerson P. Statistical Methods for Geography. London, GBR: SAGE Publications Inc. (US); 2001.

[32] Montgomery DC, Runger GC. Applied statistics and probability for engineers. United States of America: John Wiley & Sons, Inc; 2003.

[33] Australian Energy Market Operator. In: <http://www.aemo.com.au/data/price_demand.html>. p. [cited 04.12.12].

[34] Australian Bureau of Statistics. In: <http://www.abs.gov.au/>. p. [cited 04.12.12].

[35] Bureau of Meteology. In: Edited by: <http://www.bom.gov.au/climate/change/>. p. [cited 04.01.13].