SciEn
Journal of Engineering Advancements Vol. 04(03) 2023, pp 64-69 https://doi.org/10.38032/jea.2023.03.002

Forecasting Model Selection with Variables Impact to Predict Electricity Demand at Rajshahi City of Bangladesh

Md. Rasel Sarkar, Rifat Afroz Orpe and Lafifa Margia Orpa*

Department of Industrial & Production Engineering, Rajshahi University of Engineering & Technology, Rajshahi, Bangladesh

Received: May 07, 2023, Revised: August 05, 2023, Accepted: August 05, 2023, Available Online: August 18, 2023

ABSTRACT

The purpose of this study is to forecast electricity demand by using the best-selected method which untangles all the factors that affect electricity demand. Three different methods traditional methods (Multiple Regression Model), modified-traditional methods (ARMA), and soft computing method (Fuzzy Linear Regression Model) are selected for prediction. Environmental parameters like temperature, humidity, and wind speed are included as variables as Rajshahi has very impactful weather. The impact of each variable was calculated from their standardized values to know the effect of environmental parameters. The accuracy of the three forecasting models is compared by different statistical measures of errors. Using Mean Absolute Percentage Error (MAPE), the errors of the Multiple Regression Model, ARMA, and Fuzzy Linear Regression (FLR) Model are 6.85%, 22.24%, and 4.45%. The other three measures of error also give the FLR gives the best results. Finally, the electricity demand of Rajshahi City for the next five years is forecasted using the Fuzzy Linear Regression Model.

Keywords: Forecasting, Soft Computing, Fuzzy Linear Regression, Root Mean Square Error, Correlation Coefficient, Forecasting Error.

1 Introduction

In developing countries like Bangladesh, the electricity problem would create lots of hassles and problems. Rajshahi the fourth biggest city in Bangladesh has a population of over 880,026 residents [16]. Rajshahi is called the "education city" of Bangladesh and it has very impactful weather the weather is very hot during the summer season and very cold during the winter season also humidity is high due to the nearby Padma River. Here electricity demand is high because there are a lot of students in this city and the extreme weather affects the demand during the summer season. Here, Northern Electricity Supply Company (NESCO) calculates the demand manually by predicting values depending on previous data. If the demand can be forecasted by using advanced methods, then the supply calculation would be more accurate. Again, sometimes lack of data on any variable can cause problems in forecasting. If the weights of the variables are known, then data can be collected and used according to priority.

This research aims to forecast electricity demand using the best available tool, which unties all the factors that cause demand changes and pinpoints the root causes. This study is a continuation of various previous studies that analyzed different forecasting methods for calculating load forecasting, now focusing on involving weather variables in forecasting and computing the importance of weather variables, and selecting the best model for electricity demand prediction.

Researchers have used a lot of forecasting techniques from different perspectives. Five artificial intelligence forecasting methods were compared to forecast the monthly flow discharge of the Lancangjiang river [1]. Different forecasting techniques like artificial neural networks, support vector machines, genetic programming, adaptive neural-based fuzzy inference system, and autoregressive moving average are used as the methodological approaches for assessment. Correlation coefficient (R), Nash-Sutcliffe efficiency coefficient (E), root mean square error (RMSE), and mean absolute percentage error (MAPE) were used for the computation of error. Different qualitative and quantitative methods of forecasting, their advantages, disadvantages, and applications were described [2]. Some methods of model selection such as forecast error measures, information criterion, cross-validation, stepwise model selection, and residual diagnostic were pointed out. The classification, advantages, and limitations of different forecasting techniques were discussed [3] Authors have categorized forecasting methodologies into three primary classes, including soft computing forecasting techniques, modified traditional forecasting techniques, and traditional forecasting techniques. Soft computing methods are said to be more effective forecasting methods. One method from each of the three groups for load forecasting is chosen for comparison in this research work. For taking uncertainties into account, an integrated approach to electricity consumption in Iran was described. It included fuzzy linear regression (FLR), time series analysis, and principal component analysis (PCA) [4]. To address the uncertainties of meteorological factors and statistical model inaccuracy in electric load forecasting, fuzzy logic was used in conjunction with the min-max algorithm [5]. Forecasting the load for the city of Dhaka was done to test the performance of fuzzy logic, and the results showed daily forecasting of similar days was extremely accurate. Short-term electrical load forecasting was done for a remote area in Bangladesh where there were no historical data on load demand [6]. For demand forecasting, inverse matrix computation and linear regression analysis were utilized.

To predict Turkey's electricity energy usage, regression analysis, neural networks, and least square support vector machines (LS-SVM) were all compared [7]. The population,

*Corresponding Author Email Address: smdrasel10@gmail.com

installed capacity, total subscriber count, and gross electricity generation are all employed as independent variables in these models. By assessing various errors and doing a Receiver Operating Characteristics (ROC) study, LS-SVM prevailed. Using a fuzzy linear regression model, Rajshahi City's electricity demand was predicted by taking into account the number of consumers and temperature [8]. This research also used an autoregression model for forecasting the future values no of consumers and temperature. They worked with one weather variable which is temperature. The importance of wind variables was presented in load forecasting [9]. They used the sample of eight load zone and a total load of ISO New England for forecasting. They included wind variables in a regression analysis framework. The out-of-sample tests proved that variables related to wind velocity improve the models only related to temperature. But this paper has a gap in that it didn't show any complete forecasting model which can forecast the demand by combining the weather variables. A fuzzy logic model is used to forecast short-term (hourly) loads [10]. Using historical load data and the time of day, the method developed a fuzzy rule base to forecast the daily load curve. The future energy consumption of the Indian state of Tamil Nadu was forecast using an artificial neural network (ANN) enhanced by particle swarm optimization (PSO) and genetic algorithms (GA) [11]. The anticipated outcomes from the hybrid ANN-PSO model have been compared to those from the ARIMA, hybrid ANN-GA, ANN-BP, and linear models. A functional vector autoregressive state-space model's historical data on power use was the only source of information used to forecast future electricity demand [12]. The variables that comprise the model were determined using likelihood maximization, spline smoothing, functional principal components analysis, and Kalman filtering.

For a yearlong time frame, the three major Australian states' peak electricity demand was predicted [13]. This research demonstrated how crucial environmental elements are to raising predicting precision. The maximum temperature, minimum temperature, and solar exposure were the weather variables used in this paper's usage of the seasonal autoregressive integrated moving average (SARIMA) model. It had been demonstrated that when factors related to the environment were taken into account, MAPE generally improved by at least 38%. The authors also suggested that other variables like humidity and wind strength, added to the existing ones can give better prediction accuracy. Two more parameters like wind velocity and humidity to increase the accuracy of the models are added in this research work according to the suggestion of this paper. A deep learning-based approach was presented [14] for predicting power usage while considering long-term historical dependence. The study's objective was to investigate the predictive capabilities of models based on functional data analysis, an area of energy research that has received little attention to date [15]. Functional autoregressive (FAR), FAR with exogenous variable (FARX), and traditional univariate AR models are used to simulate and forecast the stochastic component, whereas generalized additive modeling is used to handle the deterministic component. Therefore, long-term (yearly) forecasting is done in our research by calculating the impact of different parameters also comparison among different forecasting methods has been shown.

2 Methodology

Electricity demands depend on a lot of parameters. Researchers have used a lot of techniques and parameters based on their locations. As Rajshahi has impactful weather its parameter was selected carefully. Based on a lot of research papers and expert's opinions we have selected four parameters for our research. Three different methods traditional methods (Multiple Regression Model), modified-traditional methods (ARMA), and soft computing method (Fuzzy Linear Regression Model) are selected for prediction.

2.1 Multiple Linear Regression Model

Multiple linear regression is a model of traditional forecasting methods that permits a statistician to form predictions about one variable using data from another variable. Electricity demand prediction has been done from this model using the parameters considered. The equation for multiple linear regression,

$$Y_i = a_0 + a_1 X_{1i} + a_2 X_{2i} + \dots\dots + a_k X_{kn} \tag{1}$$

where Y_i = i^{th} observation of the output variable, a_j = coefficient of slope for the input variables, X_{ij} = i^{th} observation of the j^{th} input variable.

2.2 ARMA Model

ARMA stands for Auto Regressive Moving Average. This model uses the information obtained from the past data of the variable to forecast its trend which is based on uni variant analysis. For an ARMA model, the formula for forecasting any value (yt) at period t is,

$$y_t = c + \sum_{i=1}^{p} \varphi_i y_{t-i} + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \tag{2}$$

Here, the dependent variable has p lags, and the error term has q lags in this model. y_t and y_{t-i} represent the values in the current period and i period ago respectively.

2.3 Fuzzy Linear Regression Model

Conventional regression model faces some problems when the number of observations is high, or the sample is huge. A normal distribution of error cannot be guaranteed and defining the relationship or vagueness between the dependent and independent variables is difficult. A fuzzy regression model is used to test the functional relationship between input and output variables to solve these problems. In a fuzzy environment, the input or output can be crisp which is converted to fuzzy by the fuzzification process. If the independent variables are denoted by $X_1, X_2, X_3, \dots\dots, X_n$, and the dependent variable is denoted by Y, then the general form of a fuzzy regression model is,

$$Y = A_0 + A_1 X_1 + A_2 X_2 + \cdots + A_n X_n \tag{3}$$

Here, A is a fuzzy number which is a function of middle (p) and spread (c).

In the present study, the probabilistic fuzzy approach is used which tries to reduce the fuzziness of the whole model by minimizing the total spreads of its fuzzy coefficients. The linear programming model to minimize the spread is,

$$Z = minimize \sum_{i=1}^{4}\sum_{j=1}^{13} c_i x_{ij} \quad (4)$$

Subject to,

$$y_j \geq \sum_{i=1}^{4} p_i x_{ij} - (1-h)\sum_{i=1}^{4} c_i x_{ij}$$

$$y_j \leq \sum_{i=1}^{4} p_i x_{ij} + (1-h)\sum_{i=1}^{4} c_i x_{ij}$$

$$p_i \geq 0 \qquad c_i \geq 0$$

where, h is the degree of fit of the estimated fuzzy linear model to the given data, and its value ranging from 0 to 1, will be assumed by the decision maker.

2.4 Forecasting Errors

The accuracy of the three forecasting models is compared by calculating the Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Mean Absolute Deviation (MAD), and Normalized Correlation Coefficient (r). Finally, the best forecasting method was selected. And then the electricity demand of Rajshahi City for the next 5 years using the best model was also forecasted.

Mean Absolute Percentage Error (MAPE):

$$MAPE = \sum_{t=1}^{n} \frac{\left|\frac{A_t - F_t}{A_t}\right|}{n} * 100 \quad (5)$$

Mean Absolute Deviation (MAD):

$$MAD = \sum_{t=1}^{n} \frac{|A_t - F_t|}{n} \quad (6)$$

Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(A_t - F_t)^2}{n}} \quad (7)$$

Normalized Correlation Coefficient:

$$r = \frac{\sum_{i=1}^{n}(A_t * F_t)}{\sqrt{\sum_{i=1}^{n}\{(F_t)*(F_t)\} * \sum_{i=1}^{n}\{(A_t)*(A_t)\}}} \quad (8)$$

3 Data Analysis

Two major sources were used for collecting the necessary data. Required data was collected from 2001 to 2018 from NESCO (Northern Electricity Supply Company Ltd.), Rajshahi, and Bangladesh Meteorological Department (BMD), Dhaka. The data of Annual numbers of consumers and annual Electricity demand of Rajshahi city were collected from NESCO. The meteorological data, which included averages for temperature, humidity, and wind speed was collected on a yearly basis from BMD.

3.1 Empirical Results for Variable Weights

First, the mean, standard deviation, and standardized values (z) of the annual number of consumers, average temperature, average humidity, and average wind speed were determined. From that, the weight of each variable was calculated using MS Excel solver which is shown in Fig. 1.

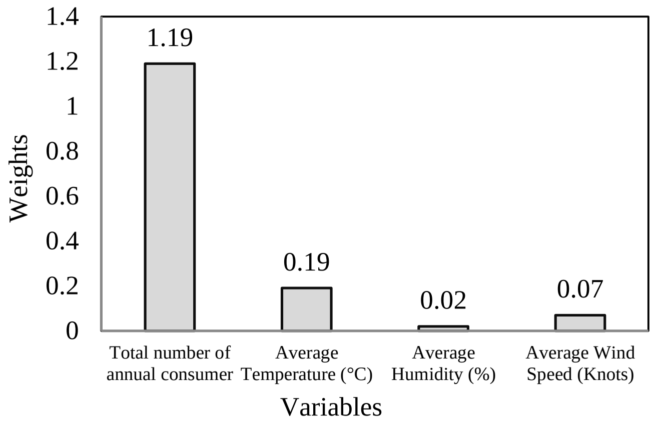

Fig. 1 Weights of the total number of annual consumers, average temperature, average humidity, and average wind speed on electricity demand

Fig. 1 shows that the number of annual consumers has the highest weight of 1.19, so it is the most impactful variable on electricity demand.

3.2 Empirical Results for Multiple Linear Regressions

Using the MS Excel solver function, the intercept and the coefficients of each variable were obtained which are presented in Table 1.

Table 1 Intercept and the coefficients of all variables

Intercept and the variables	Coefficients
Intercept	140.39648
No. of Annual Consumers	0.00044
Average Temperature (°C)	-3.16319
Average Humidity (%)	-0.61482
Average Wind Speed (knots)	0.24850

Then, putting all the coefficients, observed values, and the intercept at the regression equation, forecasted values were obtained. Similarly, electricity demand for 2015 to 2028 is forecasted and the values are presented in Table 2 with the actual electricity demands.

Table 2 Actual electricity demand and forecasted demand

Year	Actual Electricity Demand (MW)	Forecasted Demand (MW)
2014	70	67.78
2015	74	72.74
2016	82	78.39
2017	92	82.85
2018	102	86.68

3.3 Empirical Results for ARMA Model

EViews10 software was used for the identification, estimation, diagnostic, and forecasting of electricity demand by the ARMA model. A correlogram, graph of the model, and unit root test were done. The correlogram and the graph showed no trend. The null hypothesis can be disproved because the p-value is less than 0.05. Thus, the model used for this case study can be described as stationary. Moving average 1 is determined by the autocorrelation function, and autoregression 1 is determined by the partial autocorrelation function. As a result, the ARMA (1,1) model is used in this case study.

The residual diagnostic was carried out after estimating an expression for the ARMA (1,1) model. Because the p-value is higher than 0.05, the null hypothesis cannot be disproved. The residuals are hence white noise. Both the AR and MA roots are located inside the unit circle, which guarantees that the process is invertible and conforms to the covariance stationary condition. So long as the model met all requirements, predicting was feasible.

The electricity demand from 2014 to 2018 in Rajshahi City was forecasted using the forecast function of EViews. The forecasted values and the actual values are given in Table 3.

Table 3 Actual electricity demand and forecasted demand by ARMA Model

Year	Actual Electricity Demand (MW)	Forecasted Demand (MW)
2014	70	65.15
2015	74	64.54
2016	82	63.95
2018	92	63.39
2019	102	62.86

3.4 Empirical Results for Fuzzy Linear Regression

A LINGO code was generated to solve the fuzzy linear programming model and determine the fuzzy parameters. The fuzzy parameter values are obtained in Table 4.

Table 4 Fuzzy parameters obtained from LINGO code

Fuzzy Parameters	Obtained Values	Fuzzy Parameters	Obtained Values
P0	2.78522	C1	2.4404E-05
P1	0.00045	C2	0
P2	0	C3	0
P3	0	C4	0
P4	3.21209	H	0
C0	0		

The values of fuzzy parameters were put into the fuzzy regression functions to determine the upper bound and lower bound of the forecasted electricity demand values. The central value of these two bounds was taken as the forecasted value given in Table 5.

Table 5 Actual electricity demand and forecasted demand by FLR model

Year	Actual Electricity Demand (MW)	Forecasted Lower bound (MW)	Forecasted Upper bound (MW)	Forecasted Central Value (MW)
2014	70	69.09	75.77	72.43
2015	74	73.69	80.84	77.27
2016	82	79.43	87.22	83.33
2018	92	84.12	92.44	88.28
2019	102	88.70	97.61	93.16

The value of forecasted electricity demand by different models along with the actual value is presented in Fig. 3. It is seen that the ARMA model is most deviated from the actual demand line. Deviations of fuzzy linear regression are the least.

3.5 Empirical Results for MAPE, RMSE, r-value, MAD

The outcomes are displayed in Table 6 below. According to experimental findings, the fuzzy linear regression model's MAPE is the lowest of all, at 4.45%. When compared to the autoregressive moving average model, the MAPE for multiple linear regression is similarly incredibly low. The fuzzy linear regression model's RMSE, which is the lowest of all, is 4.6967.

Table 6 MAPE for different forecasting techniques

Forecasting Techniques	MAPE (%)	RMSE	r value	MAD
Multiple Linear Regression	6.85	8.2216	0.9984	6.312
Autoregressive-moving average	22.24	23.6190	0.8111	20.022
Fuzzy Linear Regression	4.45	4.6979	0.9986	3.918

The ARMA model shows the maximum error. In calculation, the Normalized correlation coefficient (r-value) for the FLR model is 0.9986 which is the largest among all. This is more accurate as the r value measures the closeness from the actual value. The r value of MLR indicates that this model is also close to the actual value. The Mean Absolute Deviation (MAD) value for FLR is lower than other methods. The FLR model has the lowest MAD value 3.918 which means the deviation from the actual value of this model is less. The results are compared graphically in Fig. 2.

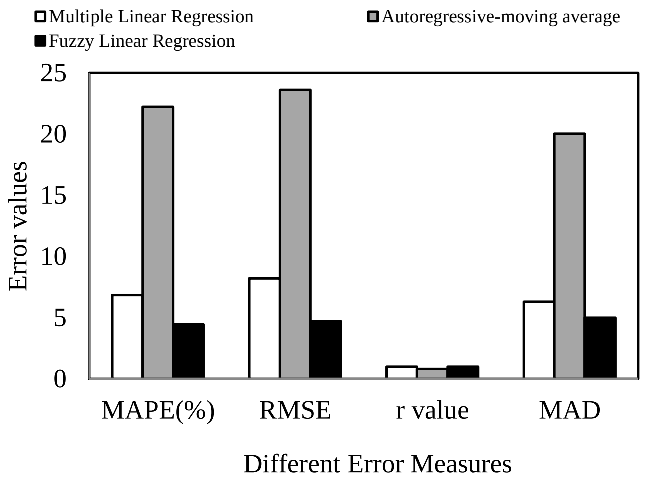

Fig. 2 Error comparison for different forecasting techniques

3.6 Forecast by Fuzzy Linear Regression Model

From Fig. 3 it is obvious that the FLR method's results are the most accurate. The fuzzy linear regression model was also utilized to anticipate Rajshahi's electricity demand for the ensuing five years (2019 to 2023), as it had the lowest error of all the techniques covered in this study.

The values of the four variables from 2019 to 2023 were forecasted using the Autoregression method. These data and values of fuzzy parameters were used in the fuzzy regression equation to obtain the forecasted value. Table 7 gives the predicted electricity demand for the next five years using the fuzzy linear regression model. It is seen that electricity demand will increase gradually in the next five years.

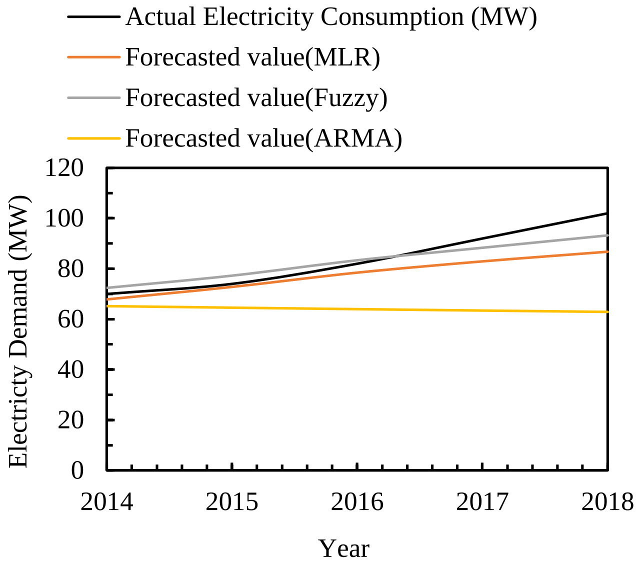

Fig. 3 Forecasted values comparison by different methods

Table 7 The forecasted electricity demand of Rajshahi City from 2019 to 2023

Year	Lower bound (MW)	Upper bound (MW)	Central Value (MW)
2019	94.45	104.01	99.23
2020	100.62	110.88	105.75
2021	107.27	118.29	112.78
2022	114.39	126.21	120.30
2023	122.03	134.72	128.38

The fuzzy linear regression model gives two peak demands – the lowest and highest demands. Central values are calculated from them. Fig. 4 provides a graphical illustration of the central values (forecasted demand) of the fuzzy regression model.

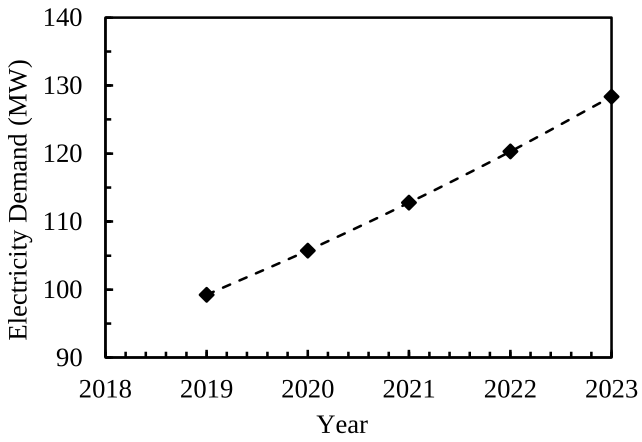

Fig. 4 Forecasting electricity demand for the next five years in Rajshahi City

4 Discussion

The results show some difference between predicted demand and actual demand. This may be due to possible slight data inaccuracies brought on by human data entry, and inherent errors in forecasting models. Only a few previous years' data have been considered, a large database will give greater accuracy. Since the fuzzy linear regression model gives the lowest MAPE, RMSE, and MAD values as well as the highest r value, it was determined to be the best model for predicting electricity demand in Rajshahi. The ARMA model has the biggest forecasting oversights. High errors could happen because of this model's lack of consideration for the variables. The other two models, however, account for the considerations. Consequently, it is true that weather factors play a crucial role in forecasting accuracy.

The first conclusion from this study is that weather factors have a significant influence on projecting electricity demand. The average temperature is also given the most weight among the meteorological factors. Along with the overall number of customers, the mean humidity and average wind speed should also be considered for more precise load forecasting.

Second, accurate forecasting outcomes from comparing various forecasting models from various categories. Different forecasting models have been evaluated using errors. It assessed the forecasting models' accuracy.

5 Conclusions

This study's goal was to explain Rajshahi's electricity demand and how factors affect electricity demand. First, the variable weights were determined. The number of customers, which is given a weight of 1.19, is the variable that affects energy usage the most severely. Next, with rankings of 0.19, 0.07, and 0.02 for average temperature, wind speed, and humidity, respectively. As a result, while average humidity is less relevant, it has a minor impact on projecting electricity demand.

Then, multiple linear regression, autoregressive moving averages, and fuzzy linear regression models were used to forecast the data from previous years. Then, four distinct types of errors were measured to compare the predicted data with the actual data. The fuzzy linear regression model lowered inaccuracy, according to experimental results. The fuzzy linear regression model has the lowest MAPE and MAD values, respectively, at 4.45% and 3.92, respectively. Fuzzy linear regression has a higher normalized correlation coefficient (0.9986) than other normalized correlation coefficients. The fuzzy linear regression model's RMSE is 4.6979. Therefore, among all forecasting models, the fuzzy linear regression model was chosen as the most efficient one.

Using the fuzzy linear regression model electricity demands have been forecasted for the next five years (2019 to 2023). Values of the four parameters from 2019 to 2023 have been forecasted using the Autoregression method. These data and values of fuzzy parameters have been used in the fuzzy regression equation to obtain the forecasted value. And from the fuzzy linear regression model, it is obtained that the electricity demand will increase gradually in the next five years. In the future, to forecast more accurately and especially with weather variables, one may apply a fuzzy linear regression model. Also including more variables will lead to the result being more accurate.

References

[1] Wang, W.C., Chau, K.W., Cheng, C.T. and Qiu, L., 2009. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology*, *374*(3-4), pp.294-306.

[2] Wang, S. and Chaovalitwongse, W.A., 2011. Evaluating and comparing forecasting models. *International Journal of Forecasting*, *14*(1), pp.35-62.

[3] Singh, A.K., Khatoon, S., Muazzam, M. and Chaturvedi, D.K., 2012, December. Load forecasting techniques and methodologies: A review. In *2012 2nd International Conference on Power, Control and Embedded Systems* (pp. 1-10). IEEE.

[4] Azadeh, A., Saberi, M. and Gitiforouz, A., 2013. An integrated fuzzy mathematical model and principal component analysis algorithm for forecasting uncertain trends of electricity consumption. *Quality & quantity*, *47*, pp.2163-2176.

[5] Monne, M.A. and Alam, K.S., 2013. Application of fuzzy logic to electric load forecasting (ELF). *Int. Journal of Science and Advanced Technology*, *3*, pp.47-50.

[6] Islam, A., Hasib, S.R. and Islam, M.S., 2013. Short term electricity demand forecasting for an isolated area using two different approaches. *Journal of Power Technologies*, *93*(4), pp.185-193.

[7] Kaytez, F., Taplamacioglu, M.C., Cam, E. and Hardalac, F., 2015. Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines. *International Journal of Electrical Power & Energy Systems*, *67*, pp.431-438.

[8] Sarkar, M.R., Rabbani, M.G., Khan, A.R. and Hossain, M.M., 2015, May. Electricity demand forecasting of Rajshahi City in Bangladesh using fuzzy linear regression model. In *2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)* (pp. 1-3). IEEE.

[9] Xie, J. and Hong, T., 2017. Wind speed for load forecasting models. *Sustainability*, *9*(5), p.795.

[10] P. Ganguly, A. Kalam, and A. Zayegh, “Short term load forecasting,” in *International Conference on Research in Education and Science (ICRES)*, 2017, pp.355–361.

[11] Anand, A. and Suganthi, L., 2020. Forecasting of electricity demand by hybrid ANN-PSO models. In *Deep learning and neural networks: Concepts, methodologies, tools, and applications* (pp. 865-882). IGI Global.

[12] Nagbe, K., Cugliari, J. and Jacques, J., 2018. Short-term electricity demand forecasting using a functional state space model. *Energies*, *11*(5), p.1120.

[13] Eshragh, A., Ganim, B., Perkins, T. and Bandara, K., 2022. The importance of environmental factors in forecasting australian power demand. *Environmental Modeling & Assessment*, *27*(1), pp.1-11.

[14] Bedi, J. and Toshniwal, D., 2019. Deep learning framework to forecast electricity demand. *Applied energy*, *238*, pp.1312-1326.

[15] Shah, I., Jan, F. and Ali, S., 2022. Functional data approach for short-term electricity demand forecasting. *Mathematical Problems in Engineering*, *2022*.

[16] Rajshahi Population 2023, Available at: https://worldpopulationreview.com/world-cities/rajshahi-population. Date of Access: 10th April 2023.