Volume 5, Issue 11, November – 2020 International Journal of Innovative Science and Research Technology

ISSN No:-2456-2165

Energy Consumption Forecasting Model for Puerto Princesa Distribution System Using Multiple Linear Regression

Alfred Rey G. Vasquez[1], Michael Ernie F. Rodriguez[2], Roy C. Dayupay[3]
[1]Palawan State University, Philippines
[2]Palawan State University, Philippines
[3]Palawan State University, Philippines

Abstract:- Power system engineers widely consider electric load forecasting because of its vital role in economically optimizing and securing the efficient operation of the power system. A forecast can be utilized by electric utilities to upgrade and improve the existing distribution facilities. Also, through this prediction, future developments could be planned concerning generation and transmission facilities. In this paper, the annual energy consumption of the Puerto Princesa Distribution System for the year 2019-2028 was forecasted using multiple linear regression. The peak demand and the number of consumers were the variables considered for the regression analysis. From the error performance test, the results indicate that multiple linear regression is a useful technique for long-term load forecasting, having a minimum percent error. Based on the regression results, the energy consumption by 2028 is expected to be 566,078,019.1 kWh. The error performance test demonstrates that the mean average percent error of 0.74% which indicates that the multiple linear regression model is a good fit.

Keywords:- *Distribution System, Energy Consumption Forecasting, Long-Term Forecast, Multiple Linear Regression.*

I. INTRODUCTION

Electricity is one of the fundamental needs and an essential resource in sustaining life that people utilize every day. The electricity demand in the world is anticipated to grow due to the dependence on electricity of humanity to perform different tasks. To prepare for the electricity demand growth, load forecasting is conducted to estimate the future demand for electricity.

In [1], there are three types of load forecasting, which is the short-term, medium-term, and long-term forecasts. Short-term forecast is used for hourly and weekly predictions, the medium-term forecast is for monthly predictions, and the long-term forecast is for yearly predictions. Short-term forecast is utilized by Dmitri et al. [2] and Srivastava et al. [3] and Singla et al. [4] while medium-term forecast is utilized in the study of Tay et al. [5]. Long-term forecast is used in the studies on [6]-[10].

Regression analysis is the modeling technique utilized in load forecasting to analyze the relationships of the different variables [6]-[8]. Simple linear regression is used by Khamaira et al. [1] and Ade-Ikuesan et al. [8], while multiple linear regression is used in the studies on [6]-[7], and [11]-[14]. Different variables are considered in performing regression analysis, such as population, gross domestic product (GDP), load demand and electricity cost [6]-[7], [11]-[13].

In this paper, the energy consumption of the Puerto Princesa Distribution System for the year 2019-2028 was forecasted using a multiple linear regression model. This mathematical model considered variables such as peak demand and the number of consumers.

II. METHODOLOGY

The historical data was collected from the utility company to forecast the energy consumption in the next 10 years. The historical data provided by Palawan Electric Cooperative (PALECO) are the number of consumers, peak demand, and energy consumption for the year 2014-2018.

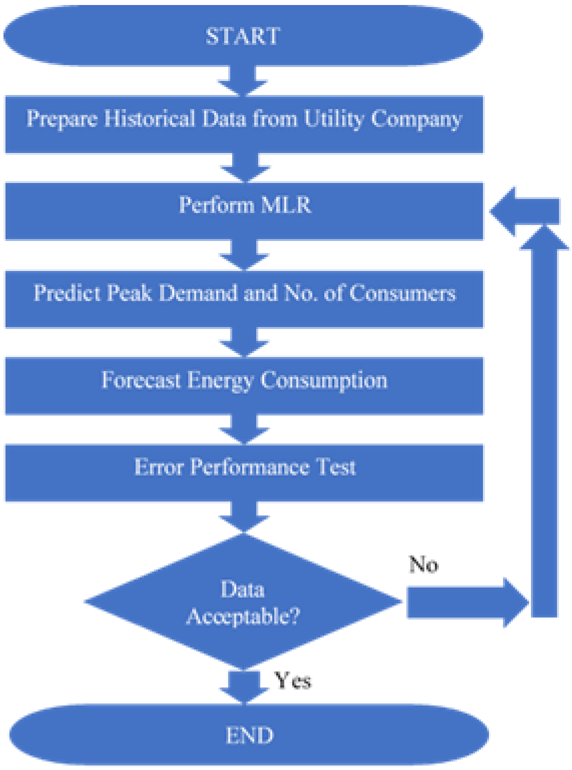

Fig 1: Process Flowchart

As defined earlier, regression is a modeling technique to analyze the relationship between a dependent and one or more independent variables. It aims to identify a function that describes the relationship between these variables as close as possible. Using multiple linear regression (MLR), the energy consumption was found in terms of the independent variables that affect the energy consumption.

2.1 Forecasting Using the MLR Model

The MLR model can be expressed as

$$Y = A + B_1 X_1 + B_2 X_2 + B_3 X_3 + ... + B_n X_n \tag{1}$$

where Y refers to the energy consumption, the A, B_1, B_2, B_3, and B_n are the unknown regression coefficients, and X_1, X_2, X_3, and X_n are the historical variables. The unknown coefficient in Eq. (1) can be calculated using a multiple regression approach by minimizing the sum of the squares of the projected errors. Equation (1) can be expressed for two historical variables, and a matrix is used to determine coefficients:

$$Y = A + B_1 X_1 + B_2 X_2 \tag{2}$$

The Analysis ToolPak of Microsoft Excel® is used in determining the unknown regression coefficients A, B_1 and B_2. To forecast the energy consumption, the peak demand X_1 and the number of consumers X_2 is predicted for 10 years by calculating for the average annual growth rate (AAGR) from the year 2014-2018.

2.2 Error Performance Test

For validation purposes, the error performance test of the forecast model is conducted [6]. Once all the independent variables are correctly identified, the error ε, sum of squares error (SSE), and the total sum of squares (TSS) are calculated as shown in Eqs. (3-5).

$$\varepsilon = yi - \hat{y}i \tag{3}$$

$$SSE = \sum_{i=1}^{n} (yi - \hat{y}i)^2 \tag{4}$$

$$TTS = \sum_{i=1}^{n} (yi - \bar{y}i)^2 \tag{5}$$

where yi is the actual value, $\hat{y}i$ is the predicted value, and $\bar{y}i$ is the average value.

After solving the SSE and TSS, the coefficient of determination (R^2) and the adjusted R^2 ($AdjR^2$) is then solved. The coefficient of determination measures the regression model as a whole, and this determines the acceptability of the model. The closer R^2 to 1, the better the model is, and it tells how well the estimated regression is. The adjusted R^2 is the calculated R^2 from those variables, which is significant to the model only.

$$R^2 = 1 - \frac{SSE}{TTS} \tag{6}$$

$$AdjR^2 = 1 - \left(\frac{SSE}{TTS}\right)\left(\frac{n-1}{n-k-1}\right) \tag{7}$$

where n is the number of data points, and k is the number of independent variables. The values of R^2 and $AdjR^2$ ranges from 0-1, wherein a value closer to one means that the data fits better with the estimated function.

Next, the t-statistic and P-value is determined. T-statistic shows the significance of each explanatory variable in predicting the dependent variable, and it has a generally accepted value of greater than 2 or less than −2 for each variable. P-value is the indicator for the probability if the parameter of population is equal to zero, and if it is equal to 0.1, it indicates a significant regression [10].

Finally, in conducting an error performance test, determining the Mean Absolute Percentage Error (MAPE) is needed. A MAPE with a value of less than 5% indicates excellent accuracy [15].

$$MAPE = \frac{\sum_{i=1}^{n} \frac{y_{actual} - y_{approx}}{y_{actual}}}{n} \tag{8}$$

III. RESULTS AND DISCUSSION

The historical data available are shown in Table 1 were increasing peak load demand was observed. The peak demand and number of consumers were considered as independent variables X1 and X2, respectively, while energy consumption represents the dependent variable Y. Using Equations 1 and 2, regression model coefficients A, B1, and B2 can be calculated, and the forecasting model can be written as:

$$Y = -102364679.6650 + 3634.1012(X_1) + 3732.0936(X_2) \quad (9)$$

Table 1: Historical Data of Energy Consumption, Peak Demand, and Number of Consumers

Year	Peak demand (kW) X_1	No. of Consumers X_2	Energy Consumption (kWh) Y
2014	29,310	38,885	149,258,474
2015	30,440	41,353	161,749,102.1
2016	31,580	44,350	180,486,933.6
2017	34,530	46,710	194,850,933.6
2018	39,120	48,336	221,079,815.6

The average annual growth rate (AAGR) is obtained from the historical data to predict the peak demand and number of consumers. The calculated AAGR was 7.56% and 5.60% for peak demand and the number of consumers, respectively. When the predicted number of consumers and peak demand is found using the calculated AAGR, future energy consumption can be forecasted using Eq. (9). Results obtained are shown in Table 2 and Fig. 2, which imply an increasing energy demand through the succeeding years from 2019 to 2028.

Table 2: Forecasted Energy Consumption for the Year 2019-2028

Year	Predicted Peak Demand (kW)	Predicted No. of Consumers	Forecasted Energy Consumption (kWh)
2019	42,076.93957	51,989.54374	244,577,019.3
2020	45,257.38352	55,919.24565	270,801,089.7
2021	48,678.22576	60,145.97955	299,007,343.0
2022	52,357.63710	64,692.19701	329,345,605.2
2023	56,315.16187	69,582.04663	361,977,027.1
2024	60,571.82164	74,841.50233	397,074,940.3
2025	65,150.22697	80,498.50129	434,825,778.1
2026	70,074.69742	86,583.09239	475,430,065.4
2027	75,371.39081	93,127.59576	519,103,484.3
2028	81,068.44213	100,166.77451	566,078,019.1

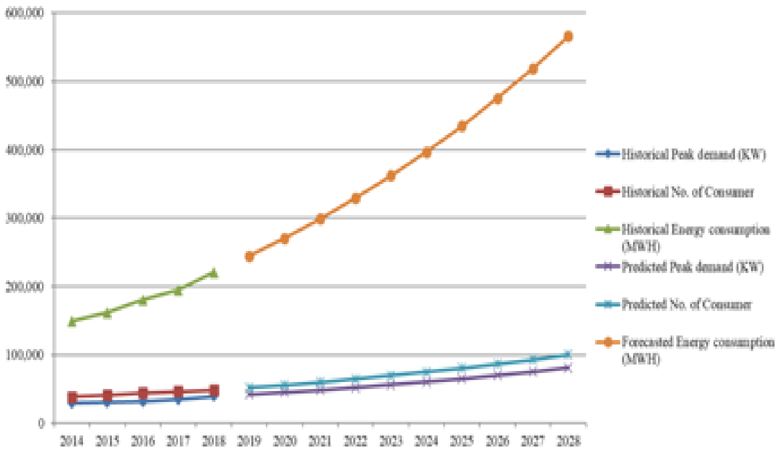

Fig 2: Historical and Forecasted Energy Consumption for the Year 2014-2028

IV. CONCLUSION

The utilization of multiple linear regression was applied to energy consumption of Puerto Princesa Distribution system from 2014 to 2028. The results obtained are summarized in Table 2 and presented as a graphical form in

Figure 2. The annual peak demand and the number of consumers recorded from the electric utility between the years 2014 and 2018 were the variables used. Based on the regression results, the energy consumption by 2028 is expected to be 566,078,019.1 kWh. Moreover, the error performance test demonstrates a coefficient of determination (R^2) and an adjusted R^2 value of 0.995 and 0.991, respectively, with a mean average percent error of 0.74%, indicating that the multiple linear regression model is a good fit. Furthermore, results obtained from this study may be used to future studies applying different forecasting techniques such as exponential smoothing, artificial neural network and MatLab.

REFERENCES

[1]. M. Y. Khamaira, A. S. Krzma, and A. M. Alnass. **Long term peak load forecasting for the libyan network**, in *Proceedings of First Conference for Engineering Sciences and Technology*, vol. 1, 2018, pp. 186-189.

[2]. K. Dmitri, A. Maria, and A. Anna. **Comparison of regression and neural network approaches to forecast daily power consumption**, in *Proc. - 2016 11th Int. Forum Strateg. Technol. IFOST 2016*, 2017, pp. 247–250, 2017.

[3]. A. K. Srivastava, D. Singh, and A. S. Pandey. **Short term load forecasting using regression trees: random forest, bagging and M5P**, in *Int. J. Adv. Trends Comput. Sci. Eng.* vol. 9, no. 2, 2020, pp. 1898–1902.

[4]. M. K. Singla, J. Gupta, P. Nijhawan, and A. S. Oberoi. **Electrical load forecasting using machine learning**, in *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 3, 2019, pp. 615–619.

[5]. K. G. Tay, A. Huong, W. K. Tiong, and S. Saleh. **Electricity consumption forecasting using fuzzy inference system**, in *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 1, 2020, pp. 249–254.

[6]. Z. Mohamed and P. Bodger. **Forecasting electricity consumption in New Zealand using economic and demographic variables**, in *Energy*, Elsevier Ltd., vol. 30, 2005, pp. 1833–1843.

[7]. M. Krutsyak. **Forecasting demand on the Ukrainian electricity market using socio-economic variables,** in *Economics, Management and Sustainability*, vol. 4, no. 1, 2009, pp. 46-57.

[8]. O. Ade-Ikuesan, A. Oyedeji, and M. Osifeko. **Linear regression long-term energy demand forecast modelling in Ogun State, Nigeria,** in *Journal of Applied Sciences and Environmental Management*, vol. 23, no. 4, 2019, 753-757.

[9]. K. R. M. Supapo, R. V. M. Santiago, and M. C. Pacis. **Electric load demand forecasting for Aborlan-Narra-Quezon distribution grid in Palawan using multiple linear regression,** in *2017 IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, Manila, 2017, pp. 1-6.

[10]. S. R. Khuntia, J. L. Rueda, and M. A. M. M. van der Meijden. **Long term load forecasting considering volatility using multiplicative error model,** in *Energies*, vol. 11, 2018, 3308 (3-5).

[11]. R. Akan, S. N. Keskin, and S. Uzundurukan. **Multiple regression model for the prediction of unconfined compressive strength of jet grout columns,** in *Procedia Earth and Planetary Science*, Elsevier Ltd., vol. 15, 2015, pp. 299-303.

[12]. N. Amral, C. S. Özveren, and D. King. **Short term load forecasting using multiple linear regression**, in *Proc. Univ. Power Eng. Conf.*, 2007, pp. 1192–1198.

[13]. B. Dhaval and A. Deshpande. **Short-term load forecasting with using multiple linear regression**," *Int. J. Electr. Comput. Eng.*, vol. 10, no. 4, 2020, pp. 3911–3917.

[14]. N. A. B. Kamisan, M. H. Lee, S. Suhartono, A. G. Hussin, and Y. Z. Zubairi. **Load forecasting using combination model of multiple linear regression with neural network for Malaysian City** in *Sains Malaysiana*, vol. 47, no. 2, 2018, pp. 419–426.

[15]. R. B. Patel, M. R. Patel, and N. A. Patel. **Electrical load forecasting using machine learning methods, RNN and LSTM,** in *Journal of Xidian University*, vol. 14 no. 4, 2020, pp. 1376-1386.