Least Square Regression in Power Consumption Forecasting in Brgy. Bahay, San Miguel

Bruses, Ram Aliester P., Tajo, Denzel Breian I., Tolibas, Jerome Y.
BSECE 4A, ECE Department, Eastern Visayas State University, Tacloban City

***Abstract*— Power or energy consumption forecasting is crucial in order for an adequate supply of electricity and meet the demands of the consumer. In this paper, least square regression was used to model and forecast the electric power consumption for 30 months of 10 houses in Brgy. Bahay, San Miguel, Leyte. For the most part, the least square method based model was able to provide good estimates of the electric consumption of the succeeding month, given that the standard deviation is low, the higher the deviation of the data, the higher the error gets. The procedure done was only for short term forecasting.**

I. Introduction

Forecasting is important, it helps determine when an event will occur or a need arise, so that the necessary actions may be taken. [1] Household power consumption falls under time series wherein it is a set of observations x_t each one at a specific time t. [2] Power consumption estimation can help both the consumers and providers. Consumers can be more prepared on their possible electric bill and providers can estimate energy demand of the consumers, in order to provide a more stable electricity supply. [3] An accurate forecast is important especially in electric consumption, wherein an under or overestimate of the forecast can have serious economic effect. [4] A common problem in modeling is level of requirement to make a reliable prediction, either lack of information or abundance of noise or useless data. [5] But with thorough knowledge and awareness of historic data and application of techniques, some observations in the past may repeat themselves, even though change is constant and nothing stays the same. [1]

Among traditional statistical methods, the most used in forecasting is the Regression (30.3%). [6] Least square method is a type of linear regression, and will be used to model the data. Linear regression is used to describe the relationship between the independent and the dependent variable, whether it contains an upward or downward trend. [7] The values made from the models will then be assessed by finding the Mean Absolute Error (MAE) and Mean Square Error (MSE), which are used as forecast accuracy measures. Wherein the predicted values will be subtracted from the actual values for that specific month.

The study was aimed to help in estimation and forecasting of future power consumption for consumers and providers, especially for residents and government officials in Brgy. Bahay, San Miguel, Leyte.

II. Theoretical Background

A. Least Square Method

$$y = a_0 + a_1 x + e \tag{1}$$

The least square method is an estimation with the minimum sum of the squared errors. [1] The equation of the line eq. 1, a_0 the intercept, a_1 the slope and e the residual or error. [8]

B. MAE and MSE

MAE and MSE are both commonly used metric in assessing the forecast accuracy of a model. [9] [10]

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|yi - \hat{yi}| \tag{2}$$

The mean absolute error MAE, to make the errors positive, it gets the absolute value of the errors and then its average. [1] Measures the average error for the entire test.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(yi - \hat{yi})^2 \tag{3}$$

The mean squared error MSE, also makes the errors positive but instead of getting its absolute value, it gets the square each error and then its average. [1] It measures the quality of the model, a lower score is better.

III. Methodology

The data were gathered from Leyte Electric Cooperative III (LEYECO III). They provided the monthly power consumption of 10 random households from Barangay Bahay, San Miguel, Leyte. The data

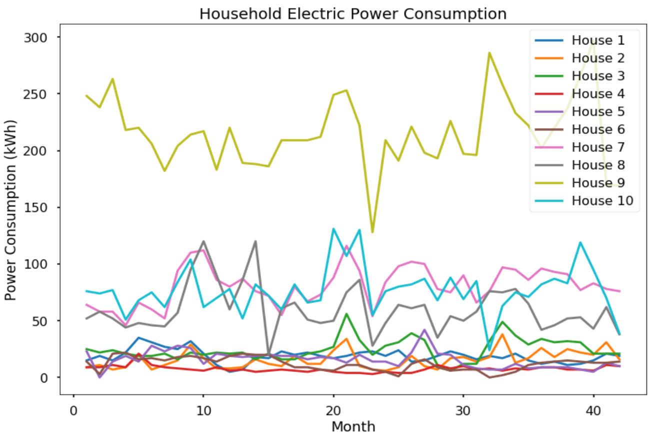

Fig. 1 Power Consumption of the 10 Households

consisted of 42 months or 3 ½ years. The first year will be initially used to model and predict the first month of the second year. To predict the succeeding months, the previous month will be included in the modeling and the first month of the modeling set that was previously used will be removed. This is to keep the modeling set updated and be adjusted to the changing data. [7] Testing would involve the regression metrics MAE and MSE. Fig. 1 show the power consumption of the 10 houses.

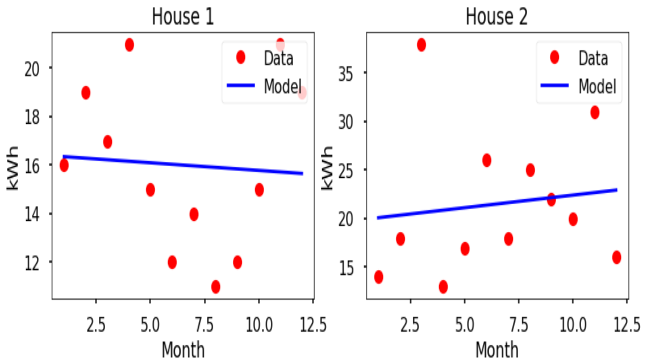

Fig. 2 Model of the first 12 months of House 1 & 2

A. Modeling

Least square method will be the method used to model the line. Training the model starts with 1 year of data, and predicts the next month of data and is compared with the actual data for that month. Fig. 2 shows the least square regression of house 1 & 2. A modeling set of 12 months was selected because power consumption is seasonal and monthly variations occur. [3] [4] The next model will also consist of 12 months of data but removing the first month used previously and adding the succeeding month that was previously tested on. This will continue until it reaches the last month of data. 42 months in total, 12 months for the initial model, and 30 months to be used for testing each of which will have a chance to be used for training the model.

1	2	3	4	5	6	7	8	9	10	11	12	13
Train												Test

2	3	4	5	6	7	8	9	10	11	12	13	14
Train												Test

Fig. 3 The procedure to be used to model

B. Testing

Each household would produce 30 models, each model predicting the next adjacent month. The results would be assessed base on its performance by using MAE and MSE on the actual data and the predicted values. The standard deviation of the raw data per household will also be recorded.

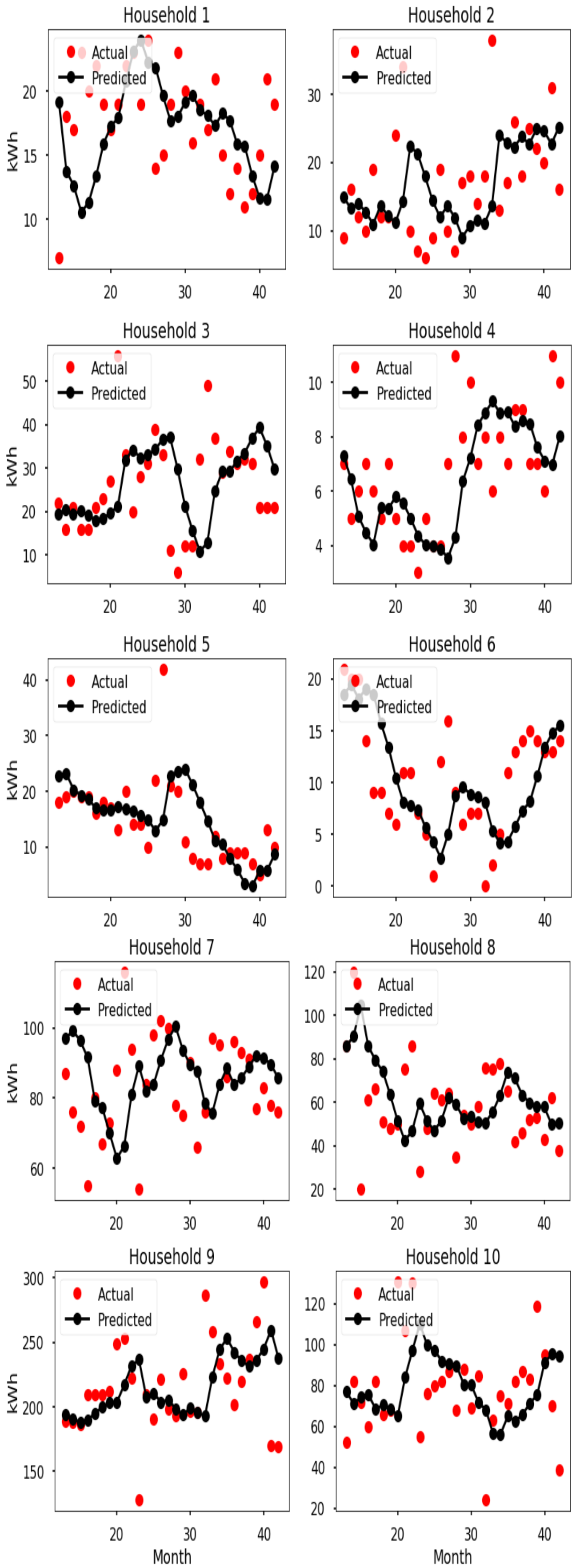

Fig. 4. Predicted values compared to the actual data

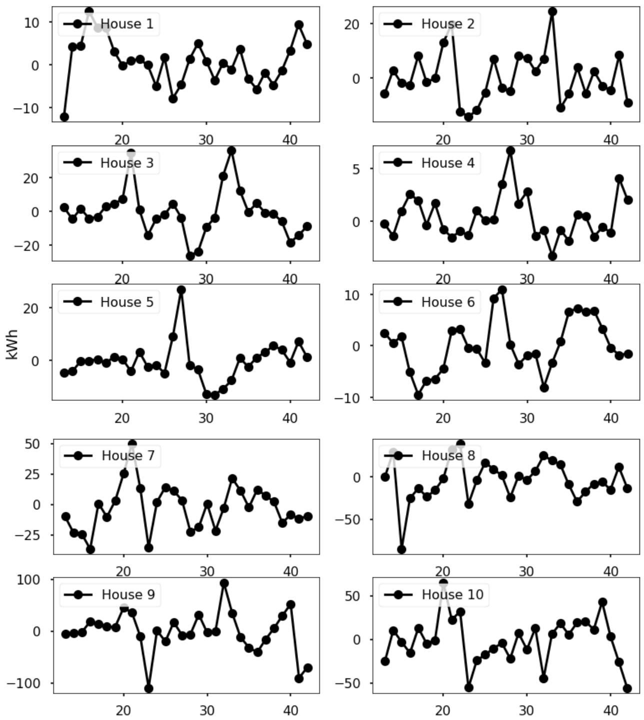

Fig.5 Magnitude of the Residuals

IV. RESULTS

Fig. 4 displays the predicted values and the actual values, while fig. 5 shows the difference between the two. The MAE and MSE were tabulated in *Table 1* showing how well the method of forecasting performed. MAE shows the average error present throughout the testing in kWh, and MSE shows how one performed in comparison to other households. Base on fig. 5 and table 1, houses 7, 8, 9, 10 performed poorly compared to the rest with very high difference in kWh between the actual and the predicted, this may be due to the large variance between the values.

Household	Std Dev	MAE	MSE
1	6.05	4.20	28.68
2	7.83	7.27	81.47
3	9.66	9.42	184.65
4	3.01	1.61	4.44
5	7.66	4.74	52.58
6	5.84	4.07	25.77
7	16.96	14.31	339.41
8	20.97	17.71	576.44
9	32.20	27.29	1567.07
10	20.72	20.46	697.66

Table 1. Results table, MAE and MSE

V. CONCLUSION

The least square regression used performed well in forecasting. The estimates that were produced did well for data with low standard deviation although struggled with higher standard deviation. The method also produced an MAE that is very close to the standard deviation, where the larger the standard deviation, the MAE also increases and has a value that is near to or close to the standard deviation.

VI. ACKNOWLEDGMENT

We would like to share our gratitude to first and foremost our dear loving heavenly Father. To our classmates and friends who helped us along the making of this study. To our instructor Engr. Ryan Reas for giving advices and guiding us to finish this project.

VII. REFERENCES

[1] S. G. Makridakis, S. C. Wheelwright and R. J. Hyndman, Forecasting: Methods and Applications, 3rd Edition, Wiley, December 1997.

[2] P. J. Brockwell and R. A. Davis, Introduction to Time Series and Forecasting, 2nd Ed., Springer, 2002.

[3] D. W. W. S. Seungho Lee, "Electricity Demand Profile of Australian Low Energy Houses," *Energy Procedia,* p. 91–100, December 2014.

[4] G. Masala and S. Marica, "Electricity load modeling: an application to Italian market," *Investment Management and Financial Innovations,* vol. 12, no. 1, pp. 35-46, 2015.

[5] V. Bianco, O. Manca and S. Nardini, "Electricity consumption forecasting in Italy using linear regression models," *Energy,* vol. 34, p. 1413–1421, 2009.

[6] G. R. T. Estevesa, B. Q. Bastos, F. L. Cyrino, R. F. Calili and R. C. Souza, "Long Term Electricity Forecast: A Systematic Review," *Procedia Computer Science* , vol. 55, p. 549 – 558, 2015.

[7] "Time Series Regression Model for Forecasting Malaysian Electricity Load Demand," *Asian Journal of Mathematics and Statistics,* vol. 1, no. 3, pp. 139-149, 2008.

[8] S. Chapra and R. P. Canale, Numerical Methods for Engineers, Sixth Ed., 1221 Avenue of the Americas, New York, NY 10020: McGraw-Hill, 2010.

[9] S. SAIGAL and D. MEHROTRA, "PERFORMANCE COMPARISON OF TIME SERIES DATA USING PREDICTIVE DATA MINING," *Advances in Information Mining,* Vols. Volume 4,, no. Issue 1, pp. pp.-57-66, 2012.

[10] W. W. a. Y. Lu, "Analysis of the Mean Absolute Error (MAE) and the Root Mean Square," in *IOP Conference Series: Materials Science and Engineering*.