International Journal of Innovation Scientific Research and Review
Vol. 05, Issue, 04, pp.4336-4339, April 2023
Available online at http://www.journalijisr.com
SJIF Impact Factor 2023: 6.599

ISSN: 2582-6131

Research Article

PREDICTION OF ELECTRICITY CONSUMPTION USING MLR

*** Jatupat Mekparyup and Kidakan Saithanu**

Burapha University, Department of Mathematics, Faculty of Science,169 Muang, Chonburi, Thailand.

Received 15th February 2023; Accepted 16th March 2023; Published online 22th April 2023

ABSTRACT

Multiple linear regression (MLR) method was applied in this research to predict electricity consumption in Chonburi province, Thailand, using three independent variables; number of electricity customers, rainfall and temperature to fit the equation. All data was separated into two data sets; (1) training data set for generating the prediction equation and (2) validation data set for validating the accuracy performance by mean absolute percentage error (MAPE). All assumptions of MRL analysis were tested by Anderson-Darling statistic (AD), Durbin-Watson statistic (DW), Breusch-Pagan statistic (BP) and Variance inflation factor (VIF). Finally, the prediction performance was illustrated by the regression standard error, the adjusted coefficients of determination and MAPE with 98,453 kWh, 0.928 and 4.65%, respectively.

Keywords: electricity consumption, multiple linear regression.

INTRODUCTION

Chachoengsao is a province in the eastern region of Thailand having the area next to the Gulf of Thailand about 12 kilometers. Chachoengsao has a power plant of the Electricity Generating Authority of Thailand, namely Bang Pakong Power Plant which has an installed capacity (Gross Capacity) of 1,450 megawatts and is expected to generate a total production capacity of 4,070 megawatts in the future. Provincial Electricity Authority Chachoengsao Province provided 11 power stations in the past. According to the report of Provincial Electricity Authority Chachoengsao Province since 2013 to 2019, the highest amount of electricity is in Muang district while the least amount of electricity is in Ratchasan district as seeing in Figure1 (Provincial Electricity Authority Chachoengsao Province, 2022). It clearly appeared that the electricity consumption tended to increase steadily every year. Therefore, studying the prediction of electricity consumption is quite necessary to plan for the production and use of electricity in Chachoengsao. Many research articles focused on studying prediction of electricity consumption (Bianco, Manca, and Nardini, 2009; Bianco, Manca, and Nardini, 2013; Shine *et al.,* 2018; Satre-Meloy, 2019; Kim, Kim, and Srebric, 2020). For this research, multiple linear regression or MLR will be utilized to generate a model for forecasting electricity consumption in Chachoengsao.

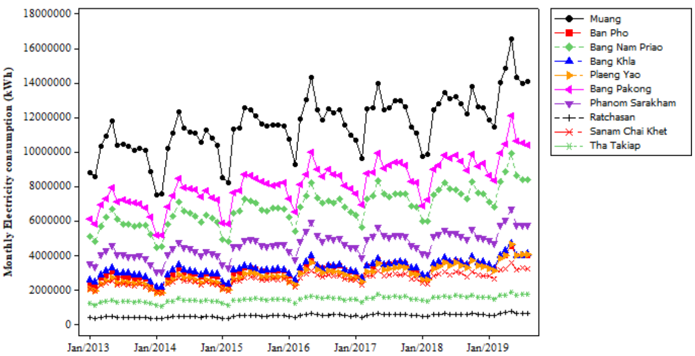

Figure 1: Monthly Electricity consumption of all districts in Chachoengsao since January 2013 – August 2019

MATERIALS AND METHODS

Data was gathered by District 2 Central Electricity Authority in Chonburisince January 2013 to August 2019 consisting of monthly average of electricity consumption as a dependent variable (y) while monthly average number of electricity customers, monthly averagerainfall and monthly averagetemperature as respectively three independent variables (x_1, x_2 and x_3). There are two data sets; training and validation data sets, subsequently displayed in Table 1.

Table 1: Data set

Data set	Size	%
Training (January 2013 – December 2017)	60	75
Validation (January 2018 – August 2019)	20	25
Total	80	100

Scatter plots between the one dependent and the three independent variables alsoPearson correlation coefficientswere monitored the linear relationship before fitting the model of prediction. MLR model of prediction as of Equation (1) was used to generate the model of prediction with stepwise method.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \quad (1)$$

where$y$ is the monthly average of electricity consumption (kWh), x_1 is the monthly average number of electricity customers, x_2 is the monthly average rainfall (millimeters), x_3 is the monthly average temperature (Celsius), β_0 is y-intercept of the model, β_1, β_2, β_3 are regression coefficients and εrepresents for an error term.After building the model, all MLR assumptions are consecutively investigated.

(1) The error terms must be normally distributed detected by Anderson-Darling statistic (AD) (Anderson and Darling, 1952).

$$\text{AD} = -n - \sum_{i=1}^{n} \left(\frac{2i-1}{n} \right) \{ \ln F(y_i) + \ln[1 - F(y_{n+1-i})] \} \quad (2)$$

(2) Auto-correlation between the error terms must be monitored by Durbin-Watson statistic (DW) (Durbin and Watson, 1950).

$$\mathrm{DW} = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \tag{3}$$

(3) Homoscedasticity (constant variance of error terms) must be tested by Breusch-Pagan statistic (BP) (Breusch and Pagan, 1979).

$$\mathrm{BP} = \frac{SSR^{*} / 2}{(SSE / n)^2} \sim \chi^2_{(k)} \tag{4}$$

(4) No correlation among the independent variables must be examined by variance inflation factor (VIF).

$$\mathrm{VIF}_j = C_{jj} = \frac{1}{1 - R^2_{j|\text{others}}};\ j = 1,\ 2,\ \ldots,\ k \tag{5}$$

Then, the indexes of prediction accuracy were severally explored by the three criteria.

(1) Mean absolute percentage error (MAPE)

$$\mathrm{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|\hat{y}_i - y_i|}{y_i} \tag{6}$$

(2) Mean absolute error (MAE)

$$\mathrm{MAE} = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{7}$$

(3) Root mean squared error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \tag{8}$$

RESULTS

Descriptive statistics was calculated by average (mean), standard deviation (SD), minimum value (min) and maximum value (max) split by types of variable and data set presented in Table 2.

Table 2: Data summary

Variables	Data set	Mean	SD	Min	Max
y	All	3,701,177	566,950	2,448,160	5,357,482
	Fitted	3,590,945	366,477	2,448,173	4,157,092
	Validated	4,244,497	492,240	3,255,942	5,357,482
x_1	All	16956	962	15,345	18,669
	Fitted	16,552	713	15,345	17,747
	Validated	18,221	278	17,787	18,669
x_2	All	10.226	7.128	0.000	41.900
	Fitted	11.040	7.100	0.000	41.900
	Validated	10.290	7.260	0.000	28.060
x_3	All	27.263	1.400	22.348	30.367
	Fitted	27.350	1.306	22.894	30.367
	Validated	27.370	1.092	25.329	29.343

***Corresponding Author: Jatupat Mekparyup,**
Burapha University, Department of Mathematics, Faculty of Science,169 Muang, Chonburi, Thailand.

There from, linear relationship was examined by scatter plot and sample Pearson correlation coefficients illustrated in Figure 2.

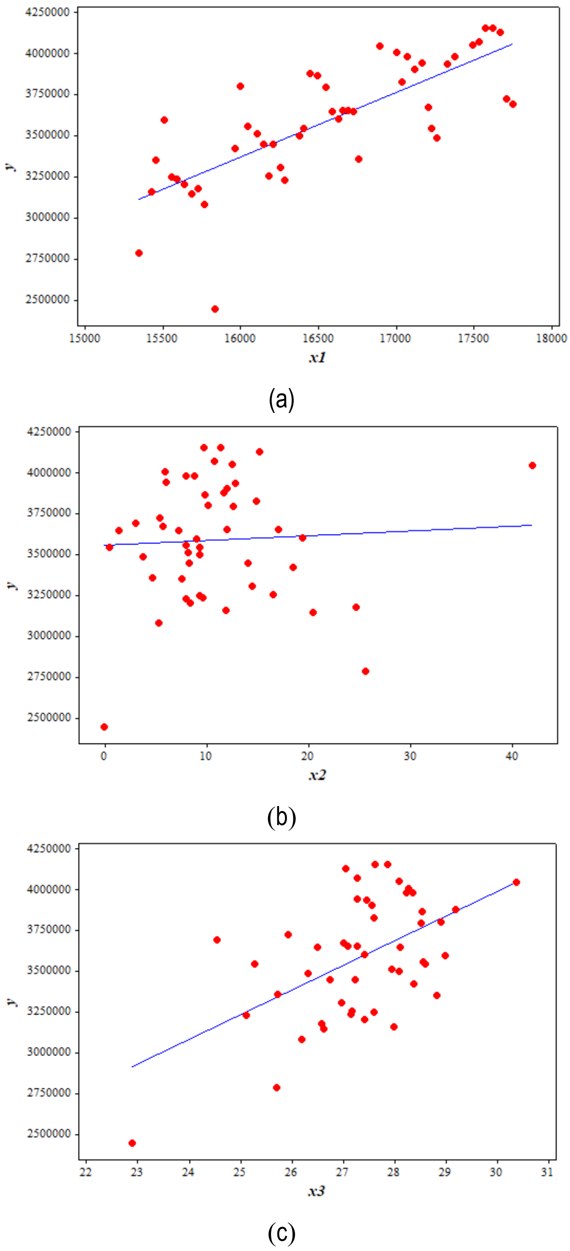

Figure 2: Scatter plot between (a) y and x_1, (b) y and x_2, (c) y and x_3

Linear relationship between y and $x_1$wasaffirmed by sample Pearson correlation coefficients of 0.769 (p-value =0.000) as the relationship between y and x_3 was 0.540 (p-value = 0.000) presented in Table 3. In conclusion, MLR method appropriated to fit a model for prediction of electricity consumption in Chachoengsao province.

Table 3: Pearson correlation coefficients among four variables

Variables	y	x_1	x_2
x_1	0.769		
	(0.000)		
x_2	0.054	-0.165	
	(0.710)	(0.258)	
x_3	0.540	-0.055	0.382
	(0.000)	(0.708)	(0.007)

p-values in bracket

Stepwise method was applied to generate the prediction model as of Equation (9).

$$\hat{y} = -7,698,517 + 411.40x_1 + 163,794x_3 \qquad (9)$$

The model was later testified by analysis of variance (ANOVA) demonstrated in Table 4 with *F*-statistic value of 309.54 (p-value = 0.000). The testing was found that the model was reasonable with adjusted coefficients of determination (r_{adj}^2) of 0.928 and standard error of regression (*S*) of 98,453 kWh.

Table 4: ANOVA

Source of variation	Degree of freedom	Sum of square	Mean of square	*F*	p-value
Regression	2	6.0008×10^{12}	3.0004×10^{12}	309.54	0.000
Residual	46	4.4588×10^{11}	9,692,997,154		
Total	48	6.4467×10^{12}			

Respectively, all parameters (except β_0) in the MLR model were verified by *t*-test with *t*-statistic value of 20.62 (p-value = 0.000) for testing β_1 *t*-statistic value of 15.03 (p-value = 0.000) for testing β_3. The result of testing was confirmed that β_1 and β_3 were not statistically significantly equal to zero.

Table 5: Regression coefficients

Predictor	Coefficient values	*t*-value	p-value
$\hat{\beta}_0$	-7,698,517	-16.85	0.000
$\hat{\beta}_1$	411.40	20.62	0.000
$\hat{\beta}_3$	163,794	15.03	0.000

According to Table 6, the MLR assumptions were examined; (1) normality testing of the error terms with AD of 0.397 (p-value = 0.355), (2) auto-correlation testing among the error terms with DW of 2.2788 (D_U = 1.6257), (3) homoscedasticity testing of the error terms with BP of 0.889 (p-value = 0.359), (4) independence testing among independent variables with VIF (x_1) of 1.00 and VIF (x_2) of 1.00. All assumptions were validated. Finally, all indexes of prediction accuracy consequently were calculated with MAPE = 0.0465, MAE = 185,478.9225 and RMSE = 286,976.0756.

Table 6: Assumption testing

Assumptions	Test statistic	Critical value	p-value
Normality	AD = 0.397		0.355
Auto-correlation	DW = 2.2788	D_U = 1.6257	
Homoscedasticity	BP = 0.889	5.991	0.359
Multicollinearity	VIF (x_1) = 1.00		
	VIF (x_3) = 1.00		

Graphical prediction performance was revealed by time series plot between real and predicted monthly averageof electricity consumption in Chachoengsao province as of Figure 3.

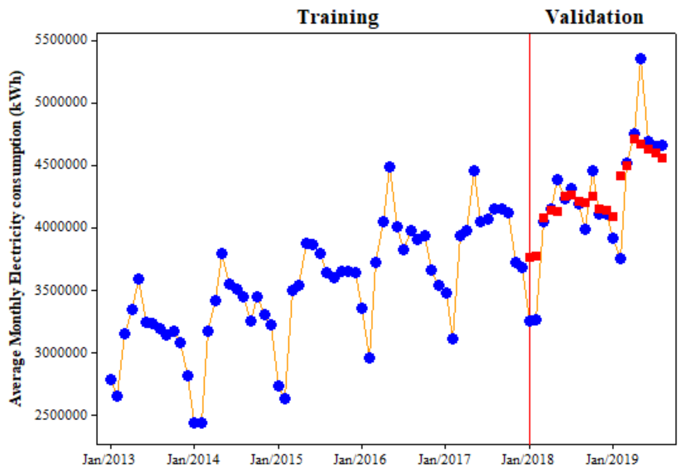

Figure 3: Time series plot between real and predicted data of monthly average of electricity consumption in Chachoengsao province

CONCLUSION AND DISCUSSION

The MLR equation for prediction of electricity consumption in Chachoengsao province, Thailand was established with stepwise technique. The number of electricity customers and temperature were meaning factors having an effect on the electricity consumption. The adjusted coefficient of determination was 0.928 and the standard error of regression was 98,453 kWh with mean absolute percentage error of 4.65%.

ACKNOWLEDGEMENTS

Specially, the authors were appreciated to provide the secondary data from District 2 Central Electricity Authority in Chonburi.

REFERENCES

Anderson, T. W., & Darling, D. A. (1952). Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes. The annals of mathematical statistics, 193-212.

Bianco, V., Manca, O., & Nardini, S. (2009). Electricity consumption forecasting in Italy using linear regression models. Energy, 34(9), 1413-1421.

Bianco, V., Manca, O., & Nardini, S. (2013). Linear regression models to forecast electricity consumption in Italy. Energy Sources, Part B: Economics, Planning, and Policy, 8(1), 86-93.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. Econometrica: Journal of the econometric society, 1287-1294. Durbin, Durbin, J., & Watson, G. S. (1950).Testing for serial correlation in least squares regression: I. Biometrika, 37(3/4), 409-428.

Energy Regulatory Commission. (2016).Overview of electricity consumption by economic sector. Available: https://www.erc.or.th/ERCWeb2/default.aspx

Kim, M. K., Kim, Y. S., & Srebric, J. (2020). Predictions of electricity consumption in a campus building using occupant rates and weather elements with sensitivity analysis: Artificial neural network vs. linear regression. Sustainable Cities and Society, 62, 102385.

Koivisto, M., Heine, P., Mellin, I., & Lehtonen, M. (2012).Clustering of connection points and load modeling in distribution systems. IEEE Transactions on Power Systems, 28(2), 1255-1265.

Lam, J. C., Wan, K. K., Liu, D., & Tsang, C. L. (2010). Multiple regression models for energy use in air-conditioned office buildings in different climates. Energy Conversion and Management, 51(12), 2692-2697.

Montgomery, D. C., & Askin, R. G. (1981). Problems of nonnormality and multicollinearity for forecasting methods based on least squares. AIIE Transactions, 13(2), 102-115.

Provincial Electricity Authority of Chachoengsao province. (2022). electricity consumption statistics in Trat. Available: https://www.pea.co.th/

Sanquist, T. F., Orr, H., Shui, B., & Bittner, A. C. (2012). Lifestyle factors in US residential electricity consumption. Energy Policy, 42, 354-364.

Shine, P., Scully, T., Upton, J., & Murphy, M. D. (2018). Multiple linear regression modelling of on-farm direct water and electricity consumption on pasture based dairy farms. Computers and Electronics in Agriculture, 148, 337-346.

Satre-Meloy, A. (2019).Investigating structural and occupant drivers of annual residential electricity consumption using regularization in regression models. Energy, 174, 148-168.

Varisara, T. (2017).Inverse modeling for energy consumption prediction in higher educational buildings: a case study of Chulalongkorn University. Master’s thesis, Department of Architecture, Faculty of Architecture, Chulalongkorn University

****