

PAPER • OPEN ACCESS

Energy consumption model improvement with principal component analysis method for small village community

To cite this article: Subhan Petrana *et al* 2018 *IOP Conf. Ser.: Earth Environ. Sci.* **105** 012014

View the [article online](#) for updates and enhancements.

You may also like

- [Adaptive Gaussian process regression for efficient building of surrogate models in inverse problems](#)
Phillip Semler and Martin Weiser
- [Experimental Researches on the Durability Indicators and the Physiological Comfort of Fabrics using the Principal Component Analysis \(PCA\) Method](#)
L Hristian, M M Ostafe, L R Manea et al.
- [Competency test clustering through the application of Principal Component Analysis \(PCA\) and the K-Means algorithm](#)
Raditya Danar Dana, Dodi Soilihudin, Ryan Hamonangan Silalahi et al.



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Energy consumption model improvement with principal component analysis method for small village community

Subhan Petrana¹, Eko Adhi Setiawan² and M Yunus H A¹

¹Department of Electrical Engineering, Universitas Khairun, Ternate, Indonesia

²Tropical Renewable Centre, Faculty of Engineering, Universitas Indonesia, Depok, Indonesia

Email : subhanpetrana@gmail.com

Abstract. Daily electrical energy consumption is a key parameter to determine balancing electricity supply and demand particularly in remote area where is not yet being electrified. The absence of daily electrical energy consumption data is a constraint for developing microgrid systems based on renewable energy. The purpose of this study is to improve electrical energy consumption model with adding some variables such as family income and family expenses, the number of family members and the load factor as independent variable, and electrical energy consumption assumption as dependent variable. Principal Component Analysis (PCA) method is used to find the relationship between independent and dependent variables. The electric energy consumption model $Y = (0.69) + (1.77e^{-7}x A_1) - (1.54e^{-7}x A_2) + (0.037x A_3) + (7.45 x A_4)$, with determination factor (R-Square) is 97%.

1. Introduction

Currently, the number of villages throughout Indonesia that have not been electrified reach 2.519 [1]. It is a challenge in development of microgrids system based on renewable energy, especially in village community. Consumption data is important parameter to determine balancing electricity supply and demand particularly in remote area where is not yet being electrified. The purpose of this study is to improve electrical energy consumption model with adding some variables such as family income and family expenses, the number of family members and the load factor as independent variable, and electrical energy consumption assumption as dependent variable. The PCA method is used to predict electric power burden as a basic for planning electrical network [2]. While [3] using PCA to predict electrical load by considering economic condition in the region. Beside PCA, Multiple Regression Method and Grey Models are also used to predict household electrical consumption. Multiple Regression Method is used by [4] to predict electrical energy consumption in Delhi by including the influence of weather change to electrical consumption. The Gray models method by [5] is used to predict electrical consumption for using as regional electrical consumption model in villages.



2. Methods

2.1. Principal Component Analysis (PCA)

PCA method is able to simplify observed variables by reducing its dimensions, by means eliminating the correlation between independent variables through the transformation of original independent variables into new variables that are not correlated to each other or known as free of multicollinearity. These variable will be regression to be seen its influence to dependent variable [3].

2.2. Mathematical Model of PCA

Mathematical model of PCA is as follows [3]:

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} = [X_1 \quad X_2 \quad \cdots \quad X_p] \quad (1)$$

where:

$$X_j = [X_{1j} \quad X_{2j} \quad \cdots \quad X_{nj}]^T ; j = 1, 2, \dots, p$$

the linear matrix combination is as follows:

$$\begin{aligned} PC_1 &= a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ PC_2 &= a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ &\vdots \\ PC_p &= a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p \end{aligned} \quad (2)$$

to maximize the ratio of PC1 variance to total Variance, with the limit that:

$$a_{i1}^2 + a_{i2}^2 + \dots + a_{ip}^2 = 1 \quad (3)$$

Principal Component is as follows:

$$PC_i = a_{1i}X_1 + a_{2i}X_2 + \dots + a_{pi}X_p \quad ; i = 1, 2, \dots, k \quad (4)$$

and matrix PC is as follow:

$$F = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1k} \\ f_{21} & f_{22} & \cdots & f_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ f_{n1} & f_{n2} & \cdots & f_{nk} \end{bmatrix} \quad (5)$$

The equation (1) to (5) are representation of PCA mathematical model.

2.3. PCA Calculation Steps

The steps of PCA calculation system and its explanation is [2]:

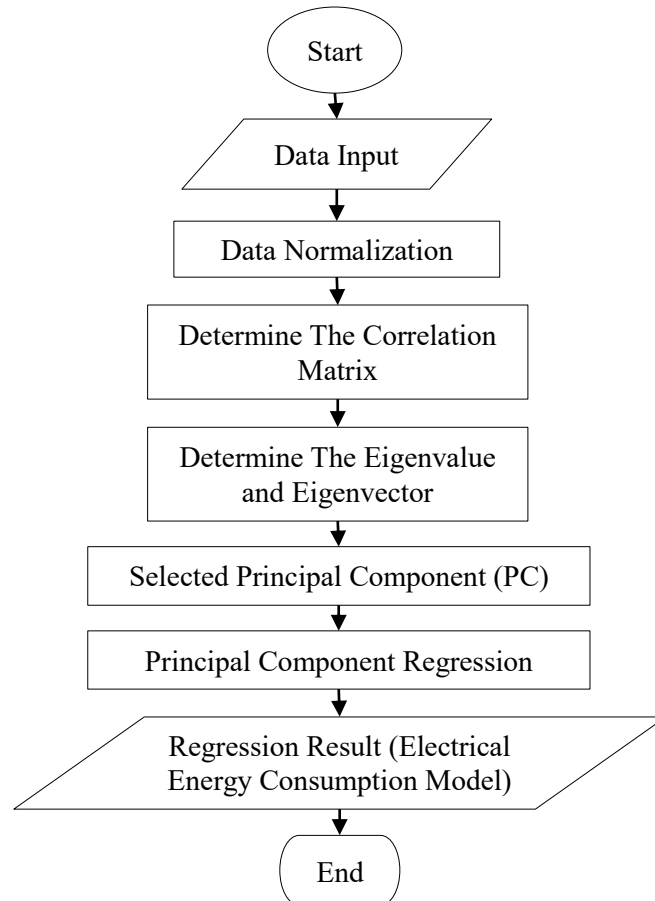


Figure 1. Calculation steps

1. The first steps of PCA calculation is to normalize the initial data as assumed in the matrix below:

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} \quad (6)$$

Its matrix size is $n \times p$, where n is number of observation and p represent of five variables. Furthermore, data normalization process is done by procedure that each observed data on a variable is subtracted with mean of its variable and divided with its standard deviation. The equation is as follow:

$$x_{ik}^1 = [x_{ik} - \bar{x}_k] / S_k \quad ; i = 1, 2, \dots, n ; k = 1, 2, \dots, p \quad (7)$$

Where x_{ik} is observation data, \bar{x}_k is mean, and S_k is deviation standard. The equation of mean and deviation standard is as follow :

$$\bar{x}_k = 1/n \sum_{i=1}^n x_{ik} \quad ; S_k = \sqrt{1/n - 1 \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2} \quad (8)$$

In this equation (8) the variant variable is 1 and mean is 0.

2. After normalizing, the next step is to determine correlation matrix with equation as follows:

$$r_{ij} = \frac{\sum_{k=1}^n x_{kr}^1 x_{kj}^1}{n-1} \quad ; i, j = 1, 2, \dots, p \quad (9)$$

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1p} \\ r_{21} & r_{22} & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{pj} & \dots & r_{pp} \end{bmatrix}$$

3. Next step is the determine of non-negative eigenvalue that uses characteristic equation $|R - \lambda I| = 0$ and eigenvector corresponded to the eigenvalues from R matrix:

$$C_i = (C_{1i}, C_{2i}, \dots, C_{mi})^T \quad ; i = 1, 2, \dots, m \quad (10)$$

4. Furthermore, M principal components arranged of the eigenvectors using the formula below:

$$PC_i = c_{1i}X_1 + c_{2i}X_2 + \dots + c_{pi}X_p \quad ; i = 1, 2, \dots, m \quad (11)$$

Principal components $PC_1, PC \dots, PC_m$ are uncorrelated variables to each other.

3. Result and Discussion

3.1. Research Flowchart

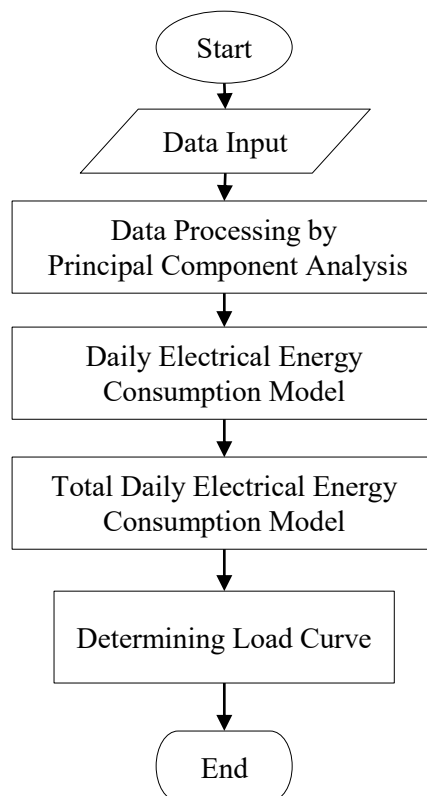


Figure 2. Research flowchart

Figure 2 shows the process of determining the daily electrical energy consumption model using the PCA method. The input data consisted of five variables were not considered before such as family income, family expenses, total family members, load factor is independent variable and electrical energy

consumption assumption as dependent variable. To find daily electrical energy consumption model, the input data is processed as shown in figure 1, and HOMER software is applied to form load profile.

3.2. Electrical Energy Consumption Assumption

In this research, electrical energy consumption assumption is determined because the proposed research object is a remote village and it has no electrical energy consumption data. The electrical energy consumption assumption calculation is as follow:

$$\cos \varphi = P/S \quad (12)$$

where $\cos \varphi$ is Power Factor, while P, S as real power and apparent power, equation (12) can be written as:

$$P = \cos \varphi \times S \quad (13)$$

so that:

$$P = 0.85 \times 450 \text{ VA} = 382.5 \text{ Watt}$$

Based on calculation result above, 382,5 Watt is a power available for each house. While, peak energy consumption is power that used of each house for 24 hours (hr) is determined using formula below:

$$\begin{aligned} \text{Peak Energy Consumption} &= P \times 24 \text{ hr} \\ \text{Peak Energy Consumption} &= 382.5 \times 24 \text{ hr} = 9180 \text{ Wh} = 9.18 \text{ kWh} \end{aligned} \quad (14)$$

Electrical energy consumption assumption is determined by use load factor 0.4, 0.6 and 0.8 that grouped based on total family income. its calculation shown in equation below:

$$\text{Electrical Consumption Assumptions} = \text{Daily Load Peak} \times \text{Load Factor} \quad (15)$$

Thus, based on equation (15) the electrical energy consumption assumption for three type of load factor above are 3.672 kWh, 5.508 kWh and 7.344 kWh.

3.3. Daily Electrical Energy Consumption Model Determination

The correlation matrix (R) is resulting after normalizing process from the initial data using equation (7) to (9), correlation matrix as shown below:

$$R = \begin{bmatrix} 1 & 0.825 & 0.085 & 0.815 & 0.814 \\ 0.825 & 1 & 0.238 & 0.717 & 0.717 \\ 0.085 & 0.238 & 1 & 0.293 & 0.294 \\ 0.815 & 0.717 & 0.293 & 1 & 0.999 \\ 0.814 & 0.717 & 0.294 & 0.999 & 1 \end{bmatrix}$$

The correlation matrix shows that there is a correlation among each variable. For example, if in first column of the matrix elements R21, R41 and R51 have strong correlation among them. The value of their matrix elements 0.825, 0.815 and 0.814. In R31 matrix element, value 0,85 has weak correlation with the three of matrix elements R21, R41 and R51. But, there is no multicollinearity occur because variance inflation for each independent variable are less than 10, as shown table below:

Table 1. Variance inflation factor (VIF)

Independent variable	Symbol	Variance Inflation
Family Income	A_1	5.383
Family Expenses	A_2	3.439
Total Family	A_3	1.278
Load Factor	A_4	3.507

The next step is to calculate eigenvalue from correlation matrix. The result of the calculation is shown in the table 2.

Table 2. Eigenvalue of correlation matrix

Principal Component	Eigenvalue	Condition Number
1	2.648	1.00
2	0.955	2.77
3	0.281	9.44
4	0.116	22.83

The result of eigenvalue calculating shows that there is no multicollinearity occur, because the condition of number for each eigenvalue are less than 100. Its calculation showing there are four principal component variables that become linier combination which has no correlation each other's. Subsequently, regression of four Principal Components resulting daily electrical energy consumption model with R-square as determinant factor 97 %. The model of daily electrical energy consumption is as follows $Y = (0.69) + (1.77e^{-7}x A_1) - (1.54e^{-7}x A_2) + (0.037x A_3) + (7.45 x A_4)$.

Consumption model as shown the equation above that the largest factor contributed to electrical energy consumption is load factor with the coefficient 7.45. Family income and expenses have small coefficient that are $1.77e^{-7}$ and $1.54e^{-7}$ but they have significant influence because the multiplying factor are in hundreds and million rupiahs. While number of family members with coefficient 0.037 has less influence in the equation above. Based on the model, total daily electrical energy consumption is 22.829 kWh/day.

3.4. Load Profile

Daily electrical energy consumption model that is 22,828 kWh/day is shows in the load curve profile below:

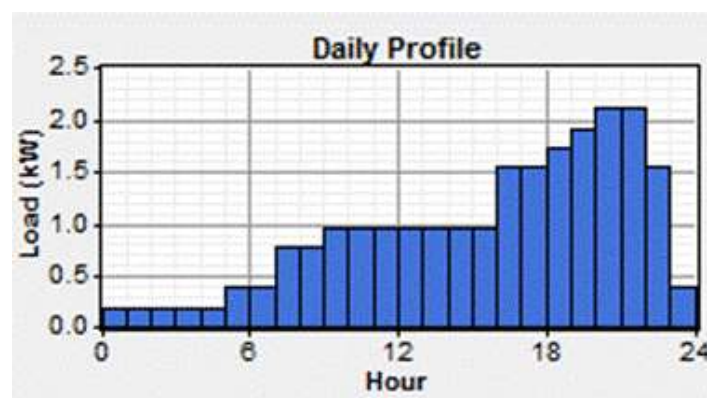
**Figure 3.** Load Profile

Figure 3 shows a change in the electrical load pattern for one day. Increase electric load occurs in the afternoon to evening, start from 4 pm to 10 pm with power 1.535 kW, 1.727 kW, 1.918 kW, 2.110 kW and 1.535 kW and peak load is 2.110 kW occurs at 8 pm to 9 pm. Increasing electric loads at night are caused by most family member doing their activity at home, such as praying, watching TV, and children learning activity. During the day, the range of load variation is from 0.767 kW to 0.959 kW because most people in the village doing their activity in outside their home.

4. Conclusion

By PCA method, the research successful determine mathematical model of electrical energy consumption $Y = (0.69) + (1.77e^{-7} \times A_1) - (1.54e^{-7} \times A_2) + (0.037 \times A_3) + (7.45 \times A_4)$. with determination factor 97%, and daily total energy around 22.829 kWh. In calculation process involving several variables such as family income, family expenses, number of family members, load factor, and electrical energy consumption assumption. This result will be used to design microgrid system with optimum configuration.

5. ACKNOWLEDGEMENT

The Authors would like to thank This work supported by “Hibah Penelitian Kerjasama Antara Perguruan Tinggi” Kemenristek Dikti. The Authors would also like to thank for all member in Tropical Renewable Energy Centre (TREC) Universitas Indonesia.

6. References

- [1] Badan Pusat Statistik. 2014. Statistik Potensi Desa Indonesia. *Badan Pusat Statistik*, Jakarta-Indonesia.
- [2] Jun-long F, Yu X, Yu F, Yang X, Guo-liang L. 2015. Rural Power System Load Forecast Based on Principal Component Analysis. *J Northeast Agric Univ* :67–72. doi:10.1016/S1006-8104(15)30034-9.
- [3] Lingli S, Changhong D, Qiushi X, Chu Z. 2012. Load Forecasting Considering The Regional Economic Environment. *Int. Conf. Autom. Control Artif. Intell.* **1** 1310–4.
- [4] Nidhi T. 2014 Determinants of the Household Electricity Consumption : A Case Study of Delhi. *Int J Energy Econ Policy* **4** 337–48.
- [5] Wu guangwei et al. 2011. The Application of Several Gray Models in Prediction of Electricity Consumption in Rural Areas. *Int. Conf. Bus. Manag. Electron. Inf., IEEE* 2–5. doi:10.1109/ICBMEI.2011.5920938.