

Micro-Spatial Electricity Load Forecasting Using Clustering Technique

*Christine Widyastuti**, *Adri Senen*, and *Oktaria Handayani*

Electrical Engineering Departement, Institute of Technology-PLN, Jakarta, Indonesia

Abstract. Low growth of electricity load forecast eliminates cost opportunity of electricity sale due to unserviceable load demands. Meanwhile, if it is exorbitant, it will cause over-investment and incriminate investment cost. Existing method of sector load is simplified and easy to implement. However, the accuracy tends to bias over one area of which data is limited and dynamic service area. Besides, the results of its forecast is macro-based, which means it is unable to show load centres in micro grids and failed to locate the distribution station. Therefore, we need micro-spatial load forecasting. By using micro-spatial load forecast, the extrapolated areas are grouped into grids. Clustering analysis is used for grouping the grids. It generates similarity matrix of similar data group. Clustering involves factors causing load growth at each grid; geography, demography, socio-economic, and electricity load per sector. Results of every cluster consist of different regional characteristics, which later the load growth is projected as to obtain more accurate forecast..

Keywords : Forecast, Micro-spatial, Grid, Cluster.

1 Introduction

Electricity power supply chain-generator, transmission and distribution started from electricity power supply planning. Electric utility forecast is one pivotal step in planning of electricity development. Forecasting is employed through anticipation model to fulfill electricity needs of electricity supply at certain period of time in an observed area [8].

In general, forecasting method is macro based model. As consequence, it is unable to show load centers in micro grids and failed to locate the distribution station. Thus, macro model for forecasting cannot be applied in creating master plan of electricity distribution. For this reason, micro-spatial load forecasting needs to be implemented in micro grids.

Micro-spatial load forecasting methods fall into two categories; trending and land use simulation. Multivariate simulation analysis is defined as a technique to simulate interactions of factors causing loads growth using historical data and condition of existing area by setting the area into certain grids [9]. This technique has no issue in the data availability at the onset. However, data availability becomes a problem in developing

* Corresponding author: christinewidyastuti@gmail.com

country and is worsen by the fast demands of electricity supply for improving financial level of the area. Clustering is the process of grouping a set of physical or abstract objects into classes of similar objects. Clustering analysis is a multivariate technique which aims to categorize objects based on their characteristics and solves the issue on large calculation process.

2 Micro-Spatial Load Forecast

2.1 Time Series Method

Micro-spatial load forecast using trending is exploration of historical data and past data growth to predict future load growth.

Baxter (2001) defined time series as a set of observations at sequence of time. Time series which is indexed as Z_t conceptually is interpreted as random variable. Sequence $\{Z_1, Z_2, \dots\}$ or $\{\dots, Z_0, Z_1, \dots\}$ is generated from stochastic process defined on probability mechanism, which is formulated in the following function:

$$Z_t = F_t + e_t \quad (1)$$

Here, we define Z_t as observation of time series, F_t as a set component of trend, cyclic, seasonality and statistics, and e_t as error.

2.2 Gompertz Method

Mathematical model of load forecasting using *Gompertz method* takes growth feature of electricity demands in certain area, where the area in smaller space tends to form unexpected ascending line and reach saturation. Willis called this a progression following curve 'S', in which asymmetric to its points of inflection. Mathematical growth curve according to (Draper and Smith, 1998) is denoted in the following equation:

$$\omega = \alpha \exp \{-\beta e^{-kt}\} \quad (2)$$

with

α : Asymptotic at value t approaching infinity

β : Integration constant;

k : Growth rate;

ω : Body weight at time t ;

t : age (days).

2.3 Multivariate Simulation Method

Spatial load prediction using multivariable analysis is a method of projecting load type and density of area development based on changes of existing and future land use [6].

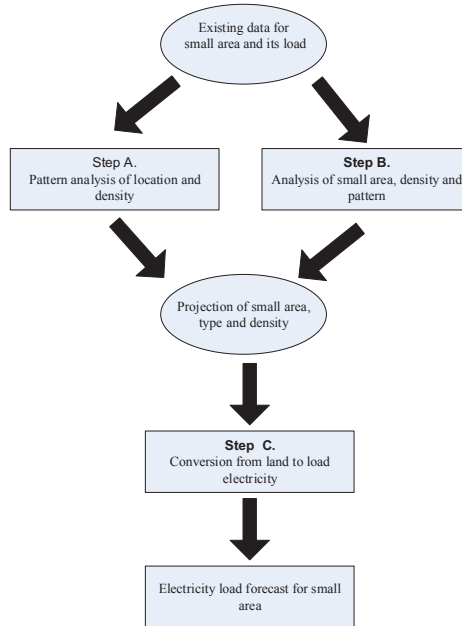


Fig. 1. Three procedures of Multivariate Analysis Prediction

3 Clustering For Micro-Spatial Load Forecasting

3.1 Clustering

Clustering aims to classify grids into relatively homogenous group of similar characteristics. Hence, load forecast is implemented in a grid model of its cluster. Cluster similarity in clustering process is measured by distance between object vectors (*Euclidean distance*). The smaller the Euclidean distance between objects, the greater their similarity [3].

Euclidean distance matrix is an $N \times N$ matrix representing the space between N objects.

$$D = \{d_{ij}\}; Ij = 1,2,3,\dots,N \tag{3}$$

The commonly used method of distance revision is *Euclidean distance* representing distance between objects.

$$d_{ij} = \sqrt{\sum_{k=1} (v_{ik} - v_{jk})^2} \tag{4}$$

“The smaller the *Euclidean distance*, the closer the grid relation.”

Grouping result and grouping advantage (C_j) are displayed in dendrogram. Dendrogram can determine total of cluster and its members. Clustering process attempts to minimize objective function, set in clustering process. The goal is generally to minimize within cluster variance and maximize between clusters variance. In other words, grouping data of similar characteristics into one cluster and grouping data of dissimilar characteristics into other group

3.2 Principal Component Analysis (PCA)

PCA is used to preserve important variables of each cluster obtained in clustering process. PCA depends on the type of original dataset. If original variables have the same units, the principle component is derived from covariance matrix. Whereas when original variables have different units of measurement, the principle component is derived from correlation matrix R. Thus, this needs transformation of original data into standard form:

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (5)$$

Total of principle components are based on the proportion of cumulative variance of 75% or more than total variance.

3.3 Factor Analysis

Factor analysis used to describe the structure of observed variables by generating factors in lower number than the observed variables. It consists of two analyses; *component analysis* and *common factor analysis*. The score of correlation coefficient between original variables and principle component variables is written as a function as follows:

$$r_{ij} = a_{ij} \sqrt{\lambda_j} \quad (6)$$

3.4 Mathematical Model

The model is formulated in multiple regressions based on the following mathematical model [2].

$$Y = b_1 + b_2 X_2 + b_3 X_3 + \dots + b_k X_k + e \quad (7)$$

Simplified into matrix:

$$Y = Xb + e \tag{8}$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_k \end{bmatrix} \quad e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_n \end{bmatrix}$$

Y : load density matrix n x 1
 X : variables matrix n x k
 B : regression coefficient matrix k x 1
 e : error matrix k x 1

To obtain the values of b, the number of squared deviation is minimized:

$$\sum e_i^2 = e'e = (Y - Xb)'(Y - Xb) \tag{9}$$

3.5 Trend Variables

To obtain load density growth at each year based on previous load density, we need to calculate trend of respective variable (except land use variables). The most appropriate trend variable to select is based on the lowest mean absolute percentage error (MAPE) [10].

Trend of land use change is taken from spatial planning and regional planning of involved area (RT/RW). As the data of neighborhood (RT) and hamlet (RW) were only available until 2010, we refer to trend of regional development in past years as to determine land use change from 2011 to 2016.

3.6 Peak Load Forecast

After trend of each variable is obtained (except for land use using RT/RW and historical data), variables trend growth are used for forecasting the load density following past load density model.

Results of load density forecast at each year are obtained from this cluster. Further, the results are used to calculate load density of respective sector of similar cluster. With the density of each sector, we can estimate the capacity of power by sector and district. This is obtained by multiplying load density per sector with the length of sector of its district in the cluster. Meanwhile, changes of length of sector at each year is adjusted to spatial regional planning. Further process is to calculate total power of district by summing total energy per sector (housing, business, industry and social), which is formulated in the following equation:

$$P_{Totalofneighborhood}(t) = C_f(P_R(t) + P_B(t) + P_I(t) + P_S(t)) \tag{10}$$

3.7 Flowchart of microspatial load forecast

Here is flowchart of all methods, displayed as follows:

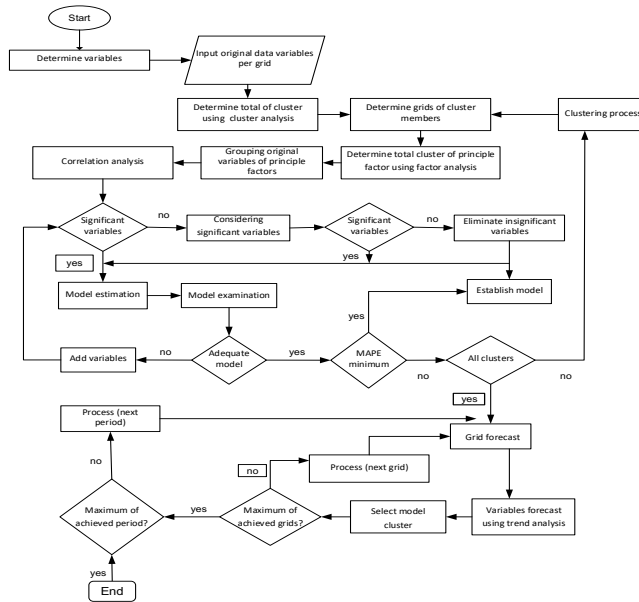


Fig 2. Micro-Spatial Flowchart Based On Multivariate Analysis Result And Discussion

4 Result and Discussion

4.1 Data

Clustering process involves a wide range of electricity variables and non-electricity variables. Total of variables are 12, and total of districts are 114. However, here we only provide 10 districts as seen in the table below:

Table 1. Variables Of Demography And Finance

No.	District	Households	GRDP
1	Belendung	2436	15,92
2	Jurumudi Baru	529	19,70
3	Jurumudi	3697	12,60
4	Batu ceper	4436	36,75
5	Cibodassari	10744	7,15
6	Cibodas	8588	23,13
7	Rawa Buaya	11445	180,61
8	Kedaung Kaliangke	7299	167,93
9	Duri Kosambi	16228	208,98
10	Cengkareng Timur	26392	183,61

Table 2. Geographical Variables

No.	District	Land Use (Ha)			
		Residential	Industry	Business	Social
1	Belendung	69,01	2,14	4,98	102,08
2	Jurumudi Baru	64,55	7,55	17,61	12,37
3	Jurumudi	63,13	2,85	6,64	40,70
4	Batu ceper	89,89	15,71	36,67	8,84
5	Cibodassari	88,19	0,25	5,50	3,25
6	Cibodas	94,89	13,83	32,27	12,00
7	Rawa Buaya	267,53	67,00	126,72	5,65
8	Kedaung Kaliangke	288,82	73,91	96,98	7,19
9	Duri Kosambi	238,37	101,38	82,73	80,41
10	Cengkareng Timur	290,54	68,10	80,38	71,57

Table 3. Electricity Variables

No.	District	Load (kW)			
		Housing	Industry	Business	Social
1	Belendung	245,64	10,39	9,12	400,66
2	Jurumudi Baru	229,77	36,72	32,22	48,55
3	Jurumudi	224,71	13,85	12,15	159,75
4	Batu ceper	319,97	76,44	67,08	34,70
5	Cibodassari	313,91	1,22	10,06	12,76
6	Cibodas	337,76	67,28	59,03	47,10
7	Rawa Buaya	4295,55	6588,11	9013,76	114,78
8	Kedaung Kaliangke	4637,39	7267,58	6898,15	146,09
9	Duri Kosambi	3827,37	9969,13	5884,64	1633,80
10	Cengkareng Timur	4664,90	6696,18	5717,95	1454,14

4.2 Building Cluster

Area taken as the object of research is area network including some parts of Tangerang, Bogor and West Jakarta. There are 114 districts taken as grids for clustering. According to data mining, two methods of clustering are hierarchical clustering and non-hierarchical clustering [6]. Hierarchical clustering is algorithm that group two or more similar objects into cluster. The process is repeatedly executed until clusters merged together producing tree-like diagram, which shows the hierarchical relationship between objects. Dendogram is usually the output to describe the hierarchical process [6].

Clustering is employed by grouping objects (districts) into clusters, in which every cluster consists of district with relatively homogenous characteristics. Objects grouping are implemented by clustering technique. Results of correlation matrix process between district variables are described below:

Table 4. Agglomeration Schedule

Cluster Combined		Coefficients	Stage Cluster First Appears	
Cluster 1	Cluster 2		Cluster 1	Cluster 2
34	40	1090367.000	0	0
72	76	1533934.000	0	0
30	41	1709857.000	0	0
45	46	2755732.000	0	0
34	39	12083793.500	1	0
24	52	17070329.000	0	0
30	34	19409947.833	3	5
37	38	20949377.000	0	0
47	51	22859369.000	0	0
18	22	25862474.000	0	0



Fig. 1. Forming Cluster

From results of clustering, the total of 114 districts is grouped into 5 clusters, as seen below:

Tabel 5. Results of Clustering

Cluster				
1	2	3	4	5
Belendung	Jurumudi	Cibodassari	Rawa Buaya	Duri Kosambi
Jurumudi Baru	Batu ceper	Cibodas	Kedaung Kaliangke	Cengkareng Timur
Caringin	Jatiuwung	Kroncong	Kapuk	Cengkareng Barat
Babakan	Pedurenan	Karang Tengah	Kamal Muara	Kalideres
Serpong	Garendeng	Gembor	Kedoya Utara	Kamal
Gunung Sindur	Pabuaran Tumpeng	Babakan	Kembangan Utara	
Sukamulya	Karang Sari		Semanan	
Mekarsari	Gebang Raya		Pegadungan	
Cikunar	Sukasari		Tegal Alur	
Parung Panjang	Kemuning			

When clustering is applied, every cluster has relatively homogenous characteristics for each district, as depicted below:

Tabel 6. Homogenous Cluster Characteristic District

Variables	Cluster 1			
	Min	Max	Mean	StDev
Total of households(KK)	5169	10744.0	7518	1941.0
Size of housing (Ha)	20.6	94.9	67.8	28.5
Size of industry (Ha)	0.25	52.0	26.79	24.3
Size of Business (Ha)	3.6	116.3	62.4	54.5
Size of Social (Ha)	3.25	60.2	15.76	22.1
Gross regional domestic product (x 10 billion)	6.9	104.4	55	47.4
Average load of household (kW)	22.5	104.0	74.3	31.2
Average industrial load (kW)	1.4	291.4	150.2	136.1
Average business load (kW)	6.1	195.6	105	91.7
Average social load (kW)	2.9	53.7	14.08	19.7
Load Density (kW/Ha)	1.14	2.7	1.807	0.6

Hierarchical clustering is able to find the similar data into one hierarchy and dissimilar data into another group. To explain the distinctions of characteristics' illustration of respective cluster at each dimension, we use discriminant analysis. Procedure started after cluster is determined.

4.3 Calculation Of Load Growth Of Each District

Based on the results, trend of each variable (except for land use using RT/RW) is used to forecast load density. Further process is to calculate total power of districts by summing total energy per sector (housing, business, industry and social).

Table 7. Results Of Load Forecats Year 1 - 5

No.	Grid	Load growth				
		1	2	3	4	5
		Peak Load	Peak Load	Peak Load	Peak Load	Peak Load
		(MVA)	(MVA)	(MVA)	(MVA)	(MVA)
1	Belendung	1,23	1,34	1,46	1,60	1,76
2	Jurumudi Baru	0,71	0,77	0,84	0,92	1,01
3	Jurumudi	0,51	0,56	0,61	0,68	0,76
4	Batu ceper	0,68	0,74	0,82	0,91	1,01
5	Cibodassari	0,45	0,49	0,54	0,59	0,66
6	Cibodas	0,70	0,77	0,84	0,93	1,03
7	Rawa Buaya	41,08	44,94	49,19	53,90	59,13
8	Kedaung Kaliaske	41,08	44,94	49,19	53,90	59,13
9	Duri Kosambi	43,42	47,43	52,04	57,33	63,40
10	Cengkareng Timur	44,08	48,16	52,83	58,20	64,38

Table 8. Results Of Load Forecast Year 6 – 10

No.	Grid	Load growth				
		6	7	8	9	10
		Peak Load	Peak Load	Peak Load	Peak Load	Peak Load
		(MVA)	(MVA)	(MVA)	(MVA)	(MVA)
1	Belendung	1,23	1,34	1,46	1,60	1,76
2	Jurumudi Baru	0,71	0,77	0,84	0,92	1,01
3	Jurumudi	0,51	0,56	0,61	0,68	0,76
4	Batu ceper	0,68	0,74	0,82	0,91	1,01
5	Cibodassari	0,45	0,49	0,54	0,59	0,66
6	Cibodas	0,70	0,77	0,84	0,93	1,03
7	Rawa Buaya	41,08	44,94	49,19	53,90	59,13
8	Kedaung Kaliangke	41,08	44,94	49,19	53,90	59,13
9	Duri Kosambi	43,42	47,43	52,04	57,33	63,40
10	Cengkareng Timur	44,08	48,16	52,83	58,20	64,38

After load growth at every grid is obtained, we can estimate load centers (distribution station) at each grid along with its geographical structure. Accumulation from load growth of each grid is spatial load growth (macro). Finally, results of micro-spatial are converted as the data for distribution system master plan.

From the load forecast at each area in the next ten years, we can calculate the total of energy and total of connected load for the whole area we put into account. The results are displayed in the table below:

Table 9. Energy And Demand Forecast

Year	Peak Load (MVA)	Energy Sales (GWh)	Connected Load (MVA)
1	1137,32	5579,23	2419,83
2	1198,43	5879,04	2549,86
3	1265,86	6209,79	2693,32
4	1340,30	6574,97	2851,70
5	1422,55	6978,44	3026,69
6	1513,55	7424,87	3220,32
7	1614,16	7918,43	3434,39
8	1725,56	8464,89	3671,40
9	1848,97	9070,31	3933,98
10	1985,80	9741,54	4225,11

Results of energy growth are elaborated from load per sector, which demands in industrial sector is higher, followed by business sector of 562.2 MVA and housing sector. Meanwhile, social sector has the lowest energy demands.

We can see that average percentage of industrial growth is more dynamic compared to other sector loads. System growth increases with linear raise with the average of 6.39 %. Average growth in other sectors; housing 7.05 %, industry 6.63 %, business 6.0 %, and social 6.93 %. On the final note, it shows that the average growth of each sector is almost similar, that is around 6 %.

5 Conclusion

Micro-spatial load forecasting is able to identify and classify area into smaller group of homogenous characteristics from each district by clustering technique. Using clustering, 114 districts are reduced into 5 clusters. Results of forecasting is an elaboration of macro-spatial load forecasting which provides solution such as information of load size, time and location with higher accuracy level. For this reason, this method is appropriate for basic planning of electricity network distribution development.

The accuracy level of this method will be better if the extrapolated area is smaller and the data validity provided is accurate. As the basis of load variables are load density, this method can accommodate immediate load changes based on allocation of spatial based on sector service.

Due to its easy programming system, this technique is feasible to be perfected by enhancing its ability in patter recognition through fuzzy application or other intelligent systems.

Reference

- [1] JP Carvallo, PH Larsen, AH Sanstad, CA Goldman, "Load Forecasting in Electric Utility Integrated Resource Planning," *osti.gov* 6, (2017)
- [2] X. Sun, Z. Ouyang and D. Yue, "Short-Term Load Forecasting Based on Multivariate Linear Regression," *IEEE Conference on Energy Internet and Energy System Integration (EI2)*, Beijing (2017)
- [3] Shahzadeh, Abbas & Khosravi, Abbas & Nahavandi, Saeid, "Improving load forecast accuracy by clustering consumers using smart meter data". **1-7**. 10.1109/IJCNN.7280393. (2015)
- [4] Laurinec, Peter & Lucka, Maria. "New clustering-based forecasting method for disaggregated end-consumer electricity load using smart grid data". **210-215**. 10.1109/INFORMATICS.8327248. (2017)
- [5] Krzysztof Gajowniczek, Tomasz Ząbkowski, "Simulation Study on Clustering Approaches For Short-Term Electricity Forecasting", *Complexity, Complex Optimization and Simulation in Power System*, Volume **2018**
- [6] Raza, Muhammad Qamar & Nadarajah, Mithulananthan & Li, Jiaming & Lee, Kwang, "Multivariate Ensemble Forecast Framework for Demand Prediction of Anomalous Days". *IEEE Transactions on Sustainable Energy*, p. **6**, (2018)
- [7] Jianwei Mi, Libin Fan, Xuechao Duan , and Yuanying Qiu, "Short-Term Power Load Forecasting Method Based on", *Hindawi, Mathematical Problems in Engineering*, Volume **2018** Improved Exponential Smoothing Grey Model (2017)

- [8] Mujiati Dwi Kartikasari and Arif Rohmad Prayogio. “Demand Forecasting of Electricity in Indonesia with Limited Historical”, *J. Phys.: Conf Data* (2018)
- [9] A Senen, “Studi Prakiraan Beban Listrik Secara Mikrosposial Berdasarkan Simulasi Tata Guna Lahan” *Jurnal Media elektro* (2013)
- [10] Adri Senen, Titi Ratnasari, “Studi Peramalan Beban Rata – Rata Jangka Pendek Menggunakan Metoda Autoregressive Integrated Moving Average (ARIMA)”, *Jurnal Ilmiah SUTET*, Vol.7, No. 2, 201