AIMS Energy, 13(5): 1241–1272.
DOI: 10.3934/energy.2025046
Received: 26 June 2025
Revised: 06 September 2025
Accepted: 25 September 2025
Published: 10 October 2025

http://www.aimspress.com/journal/energy

Research article

An integrated multivariate analysis—GIS framework for accurate microspatial energy demand forecasting

Adri Senen[1,2,*] and Jasrul Jamani Jamian[1]

[1] Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Johor, Malaysia
[2] Faculty of Electricity and Renewable Energy, Institut Teknologi PLN, Jakarta 1, Indonesia

* **Correspondence:** Emial: adri@graduate.utm.my; Tel: +6287871794354.

Abstract: Current energy demand forecasting often cannot capture the complexity and diversity of electricity demand growth in smaller areas (microspatial). This challenge is exacerbated by variables affecting load growth that differ by location. Therefore, we proposed an integrated framework that blends multivariate analysis and geographic information systems (GIS) to generate more accurate and contextually relevant energy demand forecasts at the microspatial scale, employing a grid resolution of approximately 1.51 km × 1.51 km. Multivariate analysis was used to identify and select significant variables that represent the unique conditions of each area. The variables analyzed included demographic, geographic, economic, and sectoral electricity loads. The significant variables selected through this analysis were then used to form a predictive model of future energy demand growth. The results showed that four significant variables were identified from the initial ten variables for each cluster, with the model showing a high degree of accuracy (R^2) of 0.9796, and a mean absolute percentage error (MAPE) was 3.36%. Moreover, GIS integration provided visualization and spatial analysis that strengthened the understanding of load distribution in various areas, thus supporting more effective network planning and decision-making. This approach showed significant potential in improving accuracy and spatial resolution in energy forecasting to support adaptive and sustainable electricity system management.

Keywords: energy demand forecasting; microspatial; cluster; multivariate; geographic information systems (GIS)

1. Introduction

The demand for electrical energy continues to increase in line with economic growth and technological advancements [1]. Electricity energy forecasting is a crucial stage in power system planning, enabling the anticipation of shortages or surpluses in supply while maintaining operational reliability [2]. Conventional approaches only consider the temporal dimension and ignore spatial variations at the micro area. The application of spatial-temporal-based methods improves prediction accuracy by capturing different consumption patterns between regions [3]. In addition, a multivariate approach that integrates social, economic, and spatial variables effectively represents the distribution of energy consumption more comprehensively [4]. Analysis of electricity load characteristics emphasizes spatial patterns as a critical aspect, as electricity consumption can vary significantly between locations [3]. This spatial analysis can be strengthened by integrating artificial intelligence models to capture spatial-temporal dynamics [5]. The Load forecasting also requires methods capable of representing the complexity of interactions between variables in the long term [6]. The development of multivariate methodologies has proven effective for micro-spatial areas [7]. The integration of multivariate analysis with a GIS framework offers significant potential for improving the accuracy of microspatial energy forecasting. This approach also supports optimizing power system planning while enabling spatial and temporal visualization of load distribution.

Recent research in energy forecasting has focused mainly on univariate time series analysis, which provides essential insights into relatively stable conditions [8]. However, this approach is often limited in areas with insufficient historical data or experiencing rapid land use changes due to economic growth and other dynamic factors [9]. To overcome these limitations, multivariate analysis has emerged as a more reliable methodology, enabling the identification of key factors that influence electricity load patterns on a microspatial scale [10]. This method provides a comprehensive framework for understanding the interrelationships between variables that affect electricity consumption, including demographic characteristics [11], economic conditions [12], geographical features [13], and energy consumption per sector [9]. By considering these variables, the multivariate approach can reveal complex consumption patterns previously undetectable by univariate analysis [14]. Furthermore, GIS data strengthens spatial modelling, making microspatial energy load forecasting more accurate and supporting more effective power system planning [15].

The proposed framework in this study was designed with statistical and machine learning approaches that have been extensively employed in the literature. The analysis of spatial energy demand forecasting began with classical methods such as the Principal Component Analysis (PCA) method [16], which efficiently reduces data dimensions [17], followed by Stepwise Regression, systematically selects variables to construct a parsimonious model [18]. Multiple Regression is a fundamental approach to understanding the linear relationships between independent and dependent variables [19], while Geographically Weighted Regression (GWR) extends the regression framework to include spatial heterogeneity [20]. Moreover, recent advances in machine learning offer adaptive alternatives, including Long Short-Term Memory (LSTM), which is used to handle time series data [21], and Support Vector Machine (SVM) effectively classifies high-dimensional data [22]. Integrating these methods within the framework provides a comprehensive analysis and enhances forecasting accuracy.

The stepwise method is one of the multivariate analysis methods that enables the gradual selection

of the most significant independent variables, resulting in a simpler and more easily interpretable model [23]. This method is also an alternative method that enables the gradual selection of the most significant independent variables [24]. In the context of regression, partial correlation is calculated from the residual results of regressing one independent variable against the other, thereby reinforcing the selectivity in variable selection [25]. General criteria in model selection involve evaluating the R-squared value, which tends to be stable, and the residual standard error (s) value, which approximates the data variance [26]. In spatial studies, the stepwise method has proven effective in filtering out insignificant variables, resulting in models that are more suited to regional characteristics and easier to interpret [27].

Although each method has significant advantages, the literature also highlights certain limitations in their application, particularly in micro-scale spatial energy forecasting. Table 1 summarizes a comparison of the strengths and weaknesses of the primary benchmark methods of this study in comparison to previous studies.

The summary, Table 1 illustrates that although each benchmark method has specific advantages, none can comprehensively address the complexity of spatial energy demand forecasting challenges. The proposed methodology, which integrates classical statistical techniques, modern machine learning algorithms, and GIS-based spatial analysis, is designed to bridge this gap. By leveraging the strengths of various approaches while minimizing their limitations, this framework offers a more comprehensive and innovative solution than a single strategy. This integration enhances the accuracy and robustness of the forecasting results and underscores the novelty of this research compared to previous literature. This research is expected to contribute to developing more accurate electricity consumption prediction models and support better decision-making in power system planning and management by identifying the most influential variables on electricity consumption.

In this research, the study area was discretized into regular square grids with a resolution of 1.51 km × 1.51 km to capture micro-level spatial variations in energy demand. It is essential to help manage and analyze spatial data in electricity system planning. GIS provides the ability to visualize and analyze spatial data in a geographic context, which allows for identifying patterns that are not visible through traditional analysis [28]. The integration of GIS in electricity energy demand forecasting provides more accurate mapping of load distribution based on geographic location and identifies spatial factors that influence electricity demand [29]. GIS can combine spatial data such as population, land use, and load per sector with historical electricity consumption data to create more comprehensive forecasting models [30]. To generate predictive models, the framework employs GWR, a statistical technique designed to capture local relationships between predictor and response variables across geographic locations [31]. GWR extends traditional global regression by incorporating spatial weighting functions that reflect the Influence of nearby observations [32], thereby improving model sensitivity to local conditions [33]. This spatially adaptive approach has demonstrated effectiveness in various domains, including environmental modeling and infrastructure planning, and is particularly suitable for forecasting electricity demand in microspatial contexts [34]. The use of GWR in electricity energy demand forecasting also allows the integration of multivariate data, such as demographic, economic, and environmental data [35]. By incorporating these variables, the forecasting model can capture the complex relationships between factors affecting energy consumption [29]. This improves forecasting accuracy and provides deeper insights into the dynamics of energy consumption in different regions.

By integrating GIS into spatial energy demand forecasting based on multivariate analysis, unique and complex spatial load patterns are expected to be identified. The results of this research can significantly contribute to the development of more accurate electricity consumption prediction models and provide more appropriate policy recommendations to improve energy efficiency and power system sustainability in a region.

Table 1. Benchmark publication and assessment of methods in GIS-integrated micro-spatial energy forecasting.

Method	Original Inventors/Authors	Advantages	Limitation
Principal Component Analysis [16].	Karl Pearson	Efficient in reducing dimensionality; Eliminates the issue of multicollinearity; Computation for big datasets is accelerated.	Direct interpretation is challenging; The original variable information may be lost.
Stepwise Regression [18].	M. A. Efroymson	Selecting variables automatically; Simpler models are produced, making identifying important factors easier.	Overfitting is more likely to occur when the data is unstable; significant variables could be excluded because of statistical criteria.
Multiple Regression [19].	D. Draper and H. Smith	Fundamental information about the linear relationships between independent and dependent variables is provided; Implementation and interpretation are simple.	This method cannot capture nonlinear relationships; Multicollinearity can affect it; Outlier sensitivity.
Geographically Weighted Regression [20].	Brunsdon, Fotheringham, Charlton	Spatial heterogeneity is captured; Local parameter estimate is provided; Works well with nonstationary Phenomena.	Choosing a bandwidth is difficult; Estimates with limited data could be erratic; it is intensely computational.
Long Short-Term Memory [21].	Hochreiter & Schmidhuber	Long-term dependencies are captured; Effective with time-series and sequential data; Often used in contemporary forecasting.	Large datasets are needed; Training is challenging and time-consuming; Tuning parameters is complex.
Support Vector Machine [22].	Vladimir Vapnik	Effectively handles nonlinear data; Works well for datasets with many dimensions; has a robust theoretical basis.	The kernel selection significantly impacts the adjustment of complex parameters; it is less effective with massive datasets.
Geographic Information Systems [28].	J. Rogers, M. J. Dawood, and A. A. Elkamel,	Enables space display in depth; Combines statistical and machine learning analysis with geographic data; Geographical variation in energy forecasting is captured.	Accurate and detailed spatial data is needed. There are limitations in handling complex load patterns. Results are highly reliant on grid resolution and map quality.

2. Methods

2.1. Microspatial energy *demand forecasting*

Microspatial energy demand forecasting is a method for predicting electricity demand at a particular level, where the area is divided into grids based on its characteristics (as shown in Figure 1). This method often uses GIS to analyze the spatial distribution of electricity demand and visualize the forecasting results. Microspatial analysis also integrates variables that affect energy consumption, such as economic factors, demographics, geographical location, and type of loads (residential, industrial, business, and social load) [7].

The area mapping process was used to help visualize the large and complex geographic data available, consisting of many related themes. The process begins with clusters of evaluated grids, which connect different data in a specific location, combine them, assess them, and finally map the results.

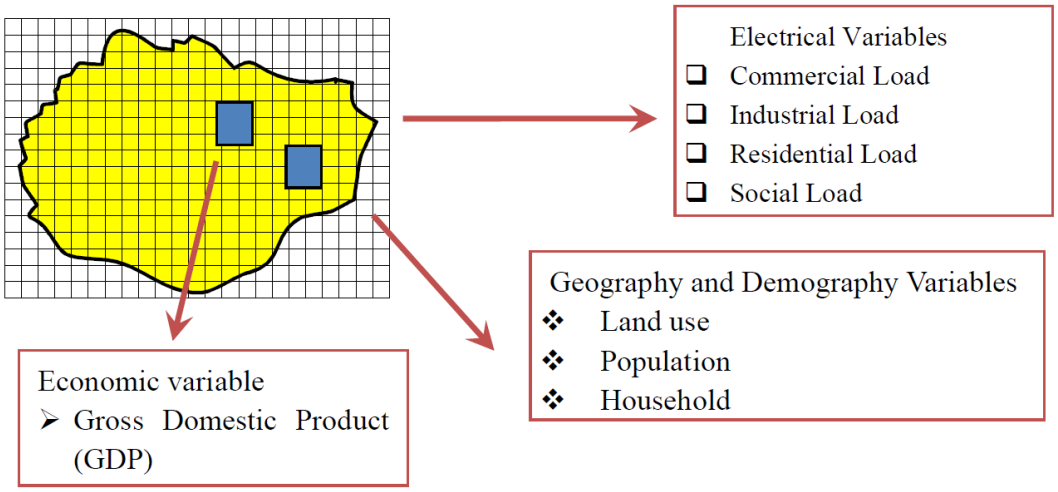


Figure 1. Microspatial based on grid division.

Spatial plotting is done as shown in Figure 1, where the land use and topography on the maps will have data attributes that contain information that can be adjusted to the electrical load profile. The spatial data can also be combined with other spatial data so that it becomes layers that contain complementary data [36], which will increase its accuracy and help facilitate the planning and development of distribution networks [37].

2.2. Adaptive clustering

Clustering methods aim to group data into clusters based on similar characteristics. The aim is to ensure that data in one cluster are highly similar, while data from different clusters have significant differences [38]. This step aims to minimize the complexity of energy demand forecasting models, with the final model determined in subsequent stages. Accordingly, this study employs the adaptive clustering method, which can autonomously determine the number of clusters with optimal

performance by analyzing attribute data and the membership degree of each data grid [39]. This is very different from the traditional clustering approach, which groups the smallest number of clusters randomly or sequentially, where the number of clusters is predetermined, making it less adaptive to complex variations in the data, especially in cases of spatial inhomogeneity. The clustering process is presented in Figure 2.

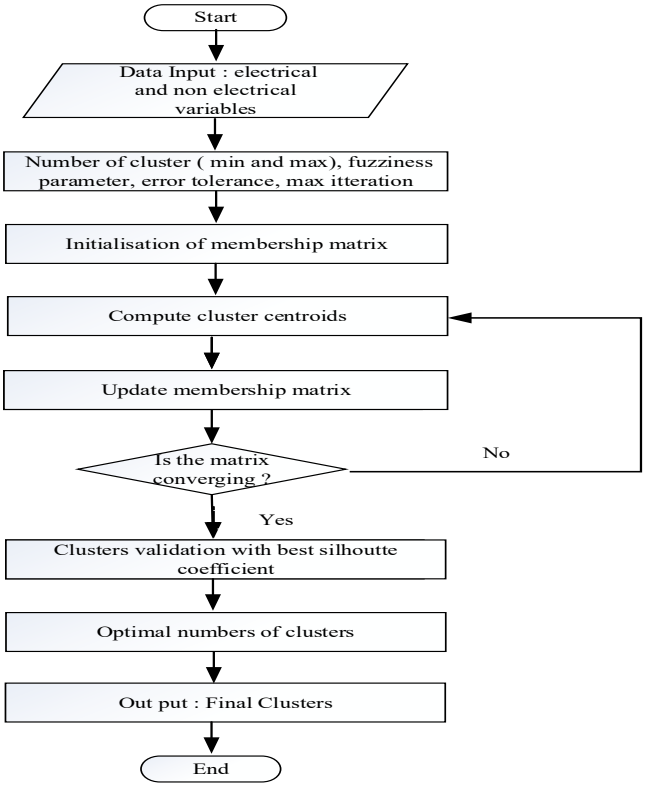


Figure 2. Flowchart of adaptive clustering.

The adaptive clustering method enables ambiguity in the boundaries between areas, making it suitable for capturing the complexity and variation that occurs in the field, which works by minimizing the fuzzy weighted objective function [40]. The objective function is formulated as follows:

$$P_m = \sum_{i=1}^{n} \sum_{j=1}^{c} \left(\mu_{ij}\right)^m \left\| x_i - c_j \right\|^2 \quad (1)$$

Where n is the number of data points, c represents clusters, xi denotes the i-th data point, c_j is the center of the j-th cluster, μ_{ij} indicates the membership degree of the i-th data point to the j-th cluster, and m is the fuzzifier parameter with a value greater than 1 that controls the degree of fuzziness in the clustering process.

The clustering process commenced with data normalization to ensure that all variables contributed equally to distance or similarity calculations, eliminating redundancy and data repetition. Subsequently, the number of optimal clusters is validated using the Silhouette algorithm [41], which evaluates cluster quality by comparing the proximity of a data point to its own cluster and its distance

to the closest other cluster. The Silhouette coefficient is calculated using the equation:

$$s(i) = \frac{b(i)-a(i)}{max(a(i),b(i)} \quad (2)$$

Where s(i) denotes the silhouette coefficient value for the i-th data point, a(i) denotes the average distance between the i-th data point and all members in its cluster, and b(i) denotes the average distance between the i-th data point and members in the nearest cluster.

2.3. Multivariate analysis using multiple regression (preliminary model)

Multiple regression is one of the most commonly used multivariate analysis techniques [42]. It enables one to model the linear relationship between one continuous dependent variable (in this case, load density) and several independent variables (e.g., load per sector, GDP and area) [43]. Mathematically, the relationship between these variables is shown in Eq (3) below.

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + e_i \quad (3)$$

Where Y as the dependent variable of size, X as the independent variable of size, and e as the random error after removing m independent variables. In vector and matrix notation, Eq (3) can be simplified into

$$Y=X\beta+e \quad (4)$$

Where

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \qquad X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} \qquad e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

2.4. Multicollinearity test

Multicollinearity can lead to unstable regression coefficient estimates and problems in the interpretation of the regression model because it is difficult to determine the effect of each independent variable separately. Therefore, it is necessary to conduct a multicollinearity test to identify the presence of a high correlation between two or more independent variables in the regression model. The method used for this test is the Variance Inflation Factor (VIF). The greater the VIF value, the higher the correlation between these and other variables. The standard acceptable VIF value is $1 < VIF < 10$. The formula can calculate VIF for a variable [33].

$$VIF(X_i) = \frac{1}{(1-R_i^2)} \quad (5)$$

R_i is the coefficient of determination, and X_i is a function of all other independent variables.

2.5. Variable selection

Variable selection is essential in determining the most relevant independent variables to predict the dependent variable. One method often used for variable selection is stepwise selection [27]. This method helps simplify the model by selecting significant variables and discarding insignificant ones [42]. This method gradually adds or removes variables from the model based on specific statistical criteria, resulting in the simplest yet most effective model to explain the variability in the dependent variable [44].

The stepwise method combines the forward (entering one by one the independent variables with the highest correlation) and backward (removing one by one the insignificant variables) methods. The steps of analysis with the stepwise method [26]:

- Calculating the correlation between each independent variable and the dependent variable.
- Entering the independent variable with the highest correlation into the regression model.
- Excluding insignificant independent variables based on specific criteria.
- Repeating steps 2–3 until there are no more variables that meet the criteria for inclusion or exclusion.

This process is done iteratively by adding or removing variables from the model based on the Akaike Information Criterion (AIC). AIC helps select the best model among several candidate models by balancing model complexity and model fit to the data [45]. Furthermore, its application extends across reliability and performance assessment domains, demonstrating its versatility in balancing parsimony and accuracy in predictive modeling [46]. In general, the mathematical equation is:

$$\text{AIC} = -2 \log (\text{L}(\theta|x) + 2\,\text{k} \quad (6)$$

Where θ is the parameter estimate that maximizes the likelihood, k is the number of parameters in the model Likelihood, and L(θ|x) is the probability of the observed data x, with unknown parameter θ.

In this context, L is a function that relates the model parameters to the observed data. In this case, the likelihood is the residual sum of squares (RSS), which is the sum of squares of the difference between the observed value and the value predicted by the regression model formulated as follows [47].

$$\text{RSS} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 - Y'X\hat{Y} \quad (7)$$

AIC considers not only the fit of the model (through RSS) but also the complexity of the model (through the number of parameters k). Models with lower AIC are considered better as they offer a balance between fit and simplicity.

2.6. Model determination

The method used to determine this model is GWR [48]. This method can create more accurate load forecasting results by considering spatial effects and local factors affecting load [49]. The GWR

model function can be written as:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \quad (6)$$

Where y_i is the observation value of the response variable at the i-th location, x_{ik} is the observation value of the predictor variable k at location i, $\beta_0(u_i, v_i)$ is the intercept value of the regression model, $\beta_k(u_i, v_i)$ is the regression parameter for the i-th location, (u_i, v_i) is the coordinate point (latitude, longitude) at the i-th location, and ε_i is the GWR model residual at the i-th location.

2.7. Selection of the best model

This stage selects the most appropriate model to explain the relationship between the dependent and independent variables, provides accurate estimates, and allows valid interpretation. The method used in this study is the coefficient of determination R-squared (R^2), which indicates how good the model is. A higher R^2 value indicates a better model. Model selection is also based on the smallest MAPE value, so the selected model can be considered feasible to represent the cluster model.

2.8. Mapping area with GIS

Mapping helps spatially analyze location-based area characteristics such as overlays and coordinates. These are then integrated with external data sources such as GIS to make mapping more informative and attractive to users. ArcGIS is one of the spatial data-based software used in applying GIS. It allows users to integrate geographic data sources, including satellite images, GPS data, survey data, and others [28].

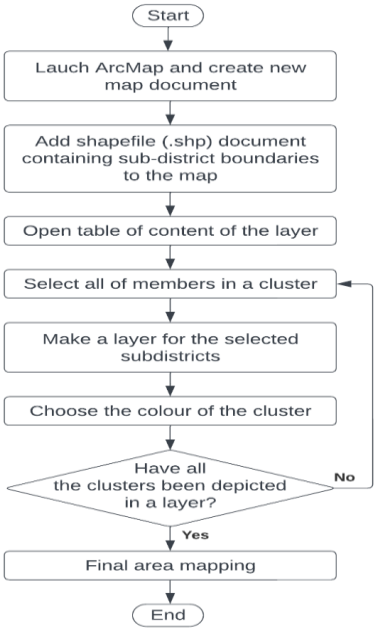


Figure 3. Geographic information systems flowchart.

In this study, ArcGIS is combined with the data extracted from variable selection in the form of data on the dominant parameters of the load profile characteristics. By linking the variable selection results with ArcGIS, it is possible to combine multivariate statistical analysis with geographic mapping, providing a deeper and more contextual insight into the electricity load data and the factors influencing its geographic characteristics. This enables better decision-making in power system planning, management, and understanding of a geographical context. The flowchart in Figure 3 illustrates the procedures for charting this region.

2.9. Load forecasting

Load forecasting is obtained by calculating the total power of each grid by summing the power in each sector (residential, business, industrial, and social) in the area based on the load density model obtained previously. The results of the load density model obtained are projected into the form of load for each coming year based on the land use in each grid concerned. The results of the load density forecasting per year obtained in this cluster are then used to calculate the load density of each sector in the same cluster. The density of each sector obtained and the amount of power per sector per village can be determined by multiplying the load density per sector by the sector area of the cluster. Moreover, the change in sector area per year is adjusted to the spatial plan and its layout.

3. Results

3.1. Study area and data

This study was conducted in the Kebayoran network area, which covers parts of Jakarta, Tangerang, and Depok, to validate the developed micro spatial forecasting model. The study area consists of 109 districts with ten independent variables. The geographical coverage of the Kebayoran network area used in this study is shown in Figure 4.

Figure 4. Mapping of the Kebayoran network area.

The data consists of 10 independent variables representing electricity, economy, geography, and demography characteristics collected from the Kebayoran network area, symbolized by X_1 to X_{10} as presented in Table 2. These independent variables are expected to affect load density as the dependent variable that Y symbolizes. The dataset is categorized into two major categories: Electrical and non-electrical variables. Electrical variables, which include household load (X_7), industrial load (X_8), commercial load (X_9), and social load (X_{10}), are obtained from the official annual average monthly records of the national electricity utility. Non-electrical variables, namely the number of households (X_1) and Regional Gross Domestic Product (X_6), are collected annually from the local government's Central Statistics Agency (BPS). Moreover, land use variables, which include residential area (X_2), industrial area (X_3), and commercial area (X_4), are obtained from regional spatial planning documents and local government data.

Table 2. Independent variables.

Variable	Description	Variable Representation
X_1	Total households	Demography
X_2	Residential area	Geography
X_3	Industrial area	Geography
X_4	Commercial area	Geography
X_5	Social area	Geography
X_6	Gross Domestic Product (GDP)	Economy
X_7	Residential load	Electricity
X_8	Industrial load	Electricity
X_9	Commercial load	Electricity
X_{10}	Social load	Electricity

3.2. Adaptive clustering

The clustering process yields four distinct clusters, as presented in Table 3. Each cluster contains varying members (grids), reflecting the grouping based on the similarity of variable characteristics. The clustering result is also depicted in Figure 3. The area involved in this study is illustrated in Figure 4.

Table 3. Total members for each cluster.

cluster 1	cluster 2	cluster 3	cluster 4
24	14	6	65

The results of the area clustering analysis using the adaptive clustering algorithm successfully group the data into four distinct clusters, as illustrated in Figure 2. Each cluster exhibits unique characteristics that represent the degree of membership of each region to a particular cluster. The assignment of data points to clusters is determined by calculating the membership degree, based on the distance between each data point and the cluster centers. A shorter distance corresponds to a higher membership degree, with values ranging from 0 to 1. A membership degree value approaching 1

indicates a strong association between the data point and the corresponding cluster. As an illustration, Table 4 provides an example of the formation of each cluster, along with the membership degree values of each data point.

Table 4. Membership degree of cluster 3.

Grid	Membership Degree			
	cluster 1	cluster 2	cluster 3	cluster 4
Tegal Parang	0.914	0.021	0.012	0.052
Cikoko	0.885	0.038	0.022	0.054
Pengadegan	0.913	0.024	0.013	0.050
Tanah Baru	0.174	0.658	0.054	0.114
Cinangka	0.203	0.446	0.137	0.214
Gandul	0.255	0.399	0.113	0.233
Rawa Barat	0.039	0.066	0.809	0.086
Paninggilan	0.040	0.062	0.827	0.072
Cipadu	0.0360	0.0591	0.8381	0.0668
Kebayoran Lama Selatan	0.204	0.132	0.106	0.557
Grogol Selatan	0.233	0.094	0.054	0.619
Pela Mampang	0.306	0.122	0.076	0.496

Furthermore, the clustering results are reconstructed and mapped using GIS. It is used in area plotting to display geographic boundaries and spatial data on a digital map. Plotting an area using GIS starts with preparing geospatial data of the region, such as shapefiles (.shp) that can be downloaded from official sources or created through manual digitisation. The data is then imported into GIS software (QGIS, ArcGIS) to ensure a proper coordinate system for accurate geographic positioning. After merging the data, users can begin visualizing information by labeling area names according to data properties and creating custom colour gradations by establishing boundary lines and colours. Consequently, as illustrated in Figure 5, the map displays a different color for each cluster.

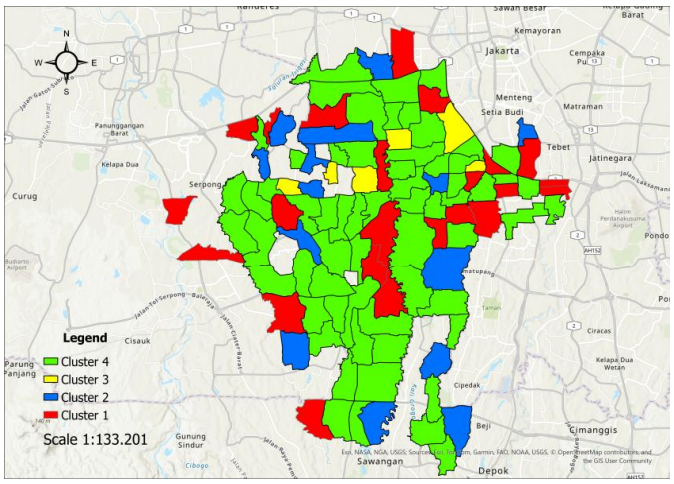


Figure 5. Study area based on clustering result.

As illustrated in Figure 5, the clustering results produces four clusters. To further narrow and focus on the study area, we select districts under clusters 1 and 4, comprising 24 and 64 grids out of 109 areas. These clusters are selected as they collectively represent 68% of the total area, making them the primary focus for load forecasting analysis. We focus on clusters 1 and 4 of 24 and 64 grids out of 109 areas, respectively. These clusters are selected as they collectively represent 68% of the total area, making them the primary focus for load forecasting analysis, as shown in Figure 6.

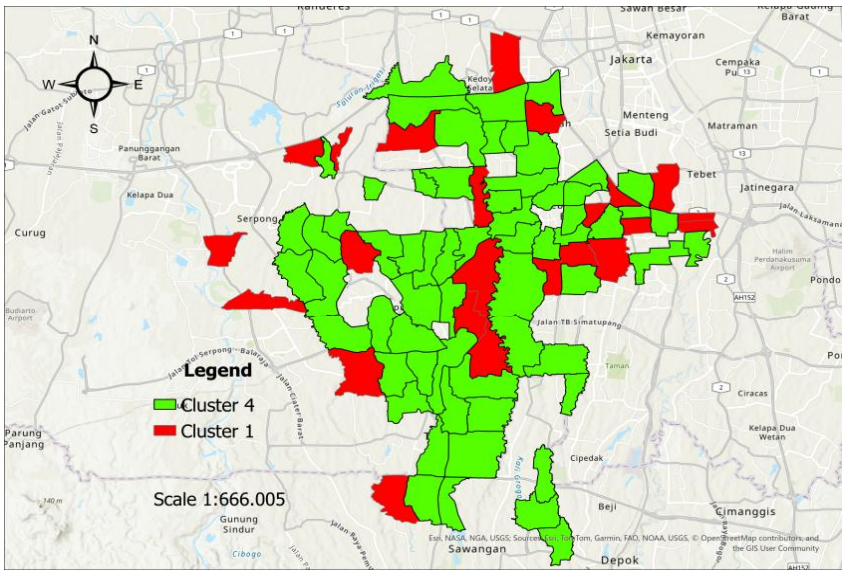


Figure 6. Study area that only involves clusteres 1 and 4.

3.3. Preliminary model

Table 5. Preliminary model results of cluster 1 and cluster 4.

Variables	cluster 1			cluster 4		
	Estimate	t value	Pr (>ItI)	Estimate	t value	Pr (>ItI)
Intercept	18.57884	2.232	0.0438	19.13	5.733	4.5e−07
X_1	0.000557	0.299	0.7697	−7.65e−05	−0.146	0.88456
X_2	0.007173	0.223	0.8274	0.00087	0.248	0.80480
X_3	−0.16749	−0.506	0.6214	0.04591	−1.687	0.09734
X_4	0.15162	0.292	0.7746	0.04634	1.156	0.25297
X_5	−0.09041	0.433	0.6720	0.03902	−2.620	0.01140
X_6	−0.08704	0.419	0.6822	0.06797	1.510	0.13925
X_7	−0.00253	−0.356	0.7275	0.001553	−1.296	0.20048
X_8	0.01449	1.301	0.2158	0.0013	0.575	0.56790
X_9	0.01077	1.409	0.1823	0.002448	2.959	0.00457
X_{10}	0.00762	0.747	0.4681	0,005902	2.945	0.00476

Table 6. Result of the multicollinearity test on the variables.

Variable	VIF	
	cluster 1	cluster 4
Household (X_1)	1.663	9.489
Land use: Residential (X_2)	4.756	92.983
Land use: Industry (X_3)	253.876	7650.073
Land use: Business(X_4)	242.887	9223.056
Land use: Social (X_5)	13.774	1139.455
GDP: (X_6)	11.743	67.424
Load: Residential (X_7)	9.896	93.794
Load: Industry (X_8)	13.651	62.223
Load: Business (X_9)	2.609	21.966
Load: Social (X_{10})	1.585	9.431

The preliminary model uses a multiple linear regression model to see the relationship between load density in each sub-district and the factors that influence it. All the independent variables involved (X_1–X_{10}) are used for the initial stage, as shown in Table 5.

Based on the calculation results in Table 5, cluster 1 shows that all independent variables do not significantly impact Load Density. This can be observed from the Pr(>|t|) values, as none of them are less than 0.05 (the 5% significance level) . The variables X_3, X_5, X_6, and X_7 have a negative effect on Load Density. Moreover, the variables X_1, X_2, X_4, X_8, X_9, and X_{10} positively affect load density. This differs from cluster 4, which has two significant variables, namely X_9 and X_{10}, with X_1, X_3, X_5, and X_7 having a negative effect on Load Density, and other variables positively impacting load density. This is likely to detect multicollinearity among the variables, which may cause problems interpreting the regression results. Therefore, it is necessary to carry out a multicollinearity test as shown in Table 6.

Based on the multicollinearity test results, as shown in Table 6, five variables have values above 10 and 200 in cluster 1, and only two variables have VIF values above 10. Therefore, the regression model obtained cannot be used as a load forecasting model representing each cluster. That way, multivariate analysis must be conducted to determine and reduce significant variables from each cluster.

3.4. Multivariate analysis

Multivariate analysis is a crucial stage in deciding the best model. We use the step-wise method to identify and determine, which independent variables are most influential on the growth of load density using AIC as an information criterion. The following is the process of selecting variables to be used as a model, as shown in Tables 7 to 12. The variable selection process started with eliminating one of the variables from each cluster from X_1 to X_{10}, as seen in Tables 7 and 8. After that, reducing one variable will continue with the reduction of two variables after obtaining the smallest AIC value (AIC value 113.01 for cluster 1 and 296.7). The process of reducing two variables is shown in Tables 8 and 9, and stops when the smallest AIC value obtained for cluster 1 is 111.07, and the funds for cluster 4 are 294.81. This process of variable reduction continues until the smallest AIC value is obtained. Furthermore, new variables can be created by eliminating and adding variables at

each step, enabling flexibility in finding the best model with the most relevant set of variables. This can be seen in Tables 10 and 11.

Table 7. The process of eliminating one of the variables of cluster 1.

Test	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	AIC
1	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	113.08
2	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	113.01
3	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	113.38
4	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	113.07
5	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	113.26
6	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	113.24
7	✓	✓	✓	✓	✓		×	✓	✓	✓	113.15
8	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	115.85
9	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	116.33
10	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	113.92

Table 8. The process of eliminating one of the variables of cluster 4.

Test	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	AIC
1	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	296.7
2	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	296.8
3	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	300.0
4	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	298.3
5	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	304.5
6	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	299.4
7	✓	✓	✓	✓	✓		×	✓	✓	✓	298.7
8	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	297.1
9	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	306.7
10	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	305.3

Table 9. The process of eliminating two of the variables of cluster 1.

Test	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	AIC
1	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	113.01
2	×	×	✓	✓	✓	✓	✓	✓	✓	✓	111.08
3	✓	×	×	✓	✓	✓	✓	✓	✓	✓	111.60
4	✓	×	✓	×	✓	✓	✓	✓	✓	✓	111.07
5	✓	×	✓	✓	×	✓	✓	✓	✓	✓	111.41
6	✓	×	✓	✓	✓	×	✓	✓	✓	✓	112.34
7	✓	×	✓	✓	✓	✓	×	✓	✓	✓	111.19
8	✓	×	✓	✓	✓	✓	✓	×	✓	✓	113.87
9	✓	×	✓	✓	✓	✓	✓	✓	×	✓	120.41
10	✓	×	✓	✓	✓	✓	✓	✓	✓	×	113.86

Table 10. The process of eliminating two of the variables of cluster 4.

Test	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	AIC
1	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	296.73
2	×	×	✓	✓	✓	✓	✓	✓	✓	✓	294.81
3	✓	×	×	✓	✓	✓	✓	✓	✓	✓	298.04
4	✓	×	✓	×	✓	✓	✓	✓	✓	✓	296.31
5	✓	×	✓	✓	×	✓	✓	✓	✓	✓	302.48
6	✓	×	✓	✓	✓	×	✓	✓	✓	✓	297.57
7	✓	×	✓	✓	✓	✓	×	✓	✓	✓	297.18
8	✓	×	✓	✓	✓	✓	✓	×	✓	✓	295.11
9	✓	×	✓	✓	✓	✓	✓	✓	×	✓	304.61
10	✓	×	✓	✓	✓	✓	✓	✓	✓	×	304.44

Table 11. A combination of reducing and adding variables of cluster 1.

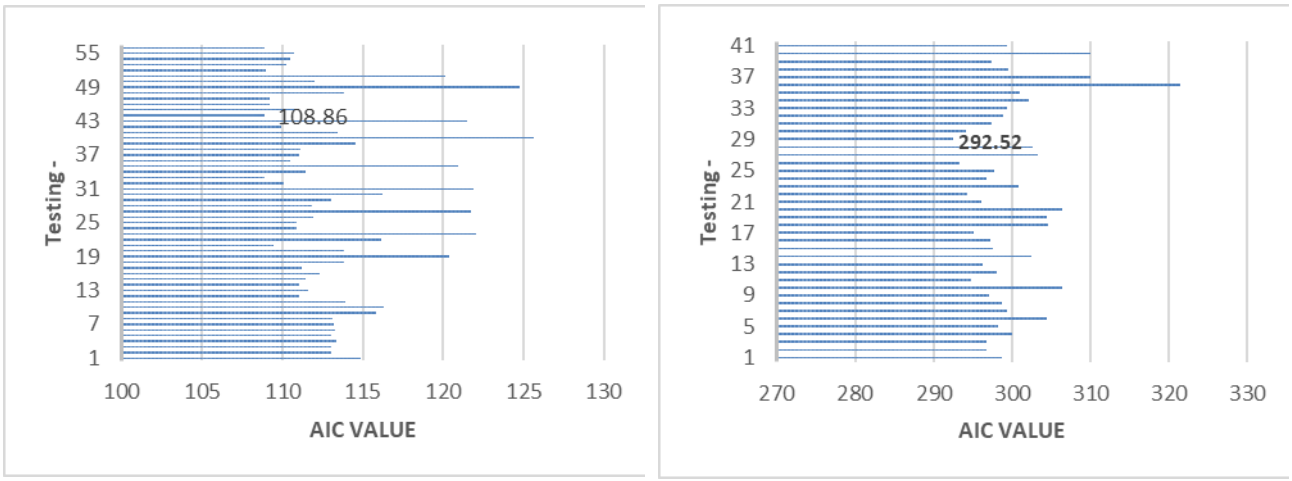
Test	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	AIC
1	×	×	✓	×	✓	✓	×	✓	×	×	120.1
2	×	×	×	×	✓	✓	×	✓	✓	×	113.8
3	×	×	✓	×	×	✓	×	✓	✓	✓	125.7
4	×	×	✓	×	✓	×	×	✓	✓	×	112.0
5	×	×	✓	×	✓	✓	×	×	✓	×	108.9
6	×	✓	✓	×	✓	✓	×	✓	×	×	109.0
7	×	×	✓	✓	✓	✓	×	✓	✓	×	110.2
8	×	×	✓	×	✓	✓	✓	✓	✓	×	110.5
9	✓	×	✓	×	✓	✓	×	✓	✓	×	110.7
10	×	×	✓	×	✓	×	✓	✓	✓	✓	110.1

Table 12. The process of eliminating one of the variables of cluster 4.

Test	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	AIC
1	×	×	✓	×	✓	✓	✓	×	✓	✓	292.52
2	×	×	✓	✓	✓	✓	✓	×	✓	✓	293.33
3	×	×	✓	×	×	✓	✓	✓	✓	✓	294.32
4	×	✓	✓	×	✓	×	✓	×	✓	✓	294.46
5	✓	×	✓	×	✓	✓	×	×	✓	✓	294.52
6	×	✓	✓	×	✓	✓	✓	×	✓	✓	297.45
7	×	×	×	×	✓	✓	✓	×	✓	✓	321.44
8	×	×	✓	×	✓	×	✓	×	✓	✓	299.46
9	✓	×	✓	×	✓	✓	✓	✓	×	✓	309.49
10	×	×	✓	×	×	✓	✓	×	✓	✓	309.93

Based on the variable selection process in Tables 11 and 12, four variables have a significant influence on load density in cluster 1, namely land use: industry (X_3), land use: social (X_5), GDP (X_6),

and business load (X_9), with the smallest AIC value of 108.9. Cluster four has six significant variables, namely land use: industry (X_3), land use: social (X_5), GDP (X_6), residential load (X_7), business load (X_9), and social load (X_{10}), with an AIC value of 292.52. At this value, the variable selection iteration has stopped; therefore, these variables are what we consider variables that represent each cluster. The overall results of variable selection experiments by reducing and adding variables based on their AIC values can be seen in the following chart in Figure 7.



(a) (b)

Figure 7. Result of AIC testing: (a) cluster 1 and (b) cluster 4.

Figure 7 shows that 56 trials are conducted to determine the variables significantly influencing the load density by combining the reduction and addition of variables until the smallest AIC value is obtained. The maximum AIC value is 125.7 with five significant variables, while the smallest AIC value is 108.86 with four significant variables in the experiment to reduce the number of variables. As for cluster 4, the maximum AIC value is 321.44 with the number of significant variables 5, while the smallest AIC value is 292.52 with 40 experiments carried out so that six significant variables are obtained, as shown in Figure 5b.

After obtaining significant variables, the multicollinearity test is carried out with the VIF indicator to ensure that the variables are suitable for use in the formation of further load forecasting modeling. Table 13 shows the complete result.

Table 13. Result of the multicollinearity test on variables.

cluster 1		cluster 4	
Variable	VIF	Variable	VIF
Land use: Industry (X_3)	3.5469	Land use: Industry (X_3)	1.7333
Land use: Social (X_5)	1.3459	Land use: Social (X_5)	1.621
GDP (X_6)	5.5653	GDP (X_6)	4.234
Load: Business (X_9)	4.5604	Load: Residential (X_7)	2.825
		Load: Business (X_9)	1.789
		Load: Social (X_{10})	1.446

Based on the multicollinearity test in Table 13 above, all obtained significant variables from each cluster have VIF values below 10. Cluster 4 has VIF values below 5.

3.5. Determination and selection of the best model

The resulting R^2 value indicates a model's goodness. An R^2 value greater than other models demonstrates that the model is better. Table 14 shows the measure of model goodness produced by linear regression and GWR models.

Table 14. Determination model.

Model	R^2	
	cluster 1	cluster 4
Linear Regression	0.785	0.799
GWR	0.998	0.979

Table 14 shows that the overall R^2 value generated by the GWR model is greater than that of the linear regression model. This indicates that the GWR model is better suited for modeling load density. Furthermore, GWR is applied to each location in each cluster based on the previously significant variables obtained by determining each region's coordinates. After that, each model is tested for each area until the best model that can represent each cluster's model based on the smallest MAPE value is obtained. Tables 15 and 16 below show each cluster's five sampling areas with the smallest MAPE value.

Table 15. Selection of the best GWR model of cluster 1.

Subdistrict	GWR Model Equation	MAPE
Menteng Dalam	$y = 36.99 - 0.03X_3 - 0.04X_5 - 0.14X_6 + 0.01X_9$	6.09%
Pengadegan	$y = 38.24 - 0.03X_3 - 0.04X_5 - 0.14X_6 + 0.01X9$	3.36%
Tegal Parang	$y = 35.81 - 0.03X_3 - 0.04X_5 - 0.134X_6 + 0.01X_9$	6.03%
Ulujami	$y = 32.68 - 0.03X_3 - 0.04X_5 - 0.08X_6 + 0.01X_9$	7.61%
Cikoko	$y = 38.39 - 0.03X_3 - 0.04X_5 - 0.14X_6 + 0.01X_9$	5.38%

Table 16. Selection of the best GWR model of cluster 4.

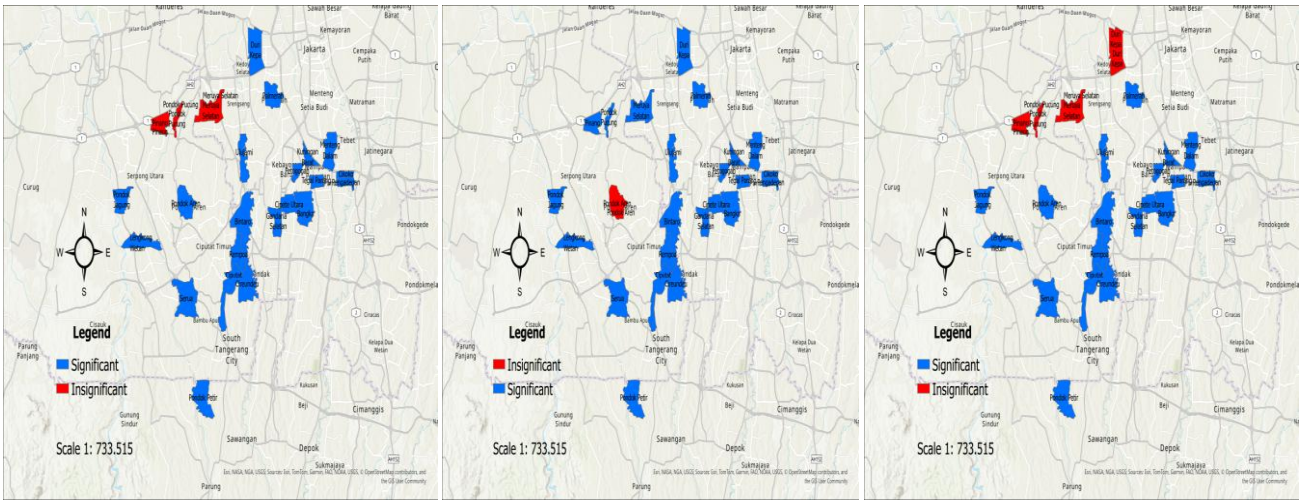
Subdistrict	GWR Model Equation	MAPE
Gandaria Utara	$y = 34.666 - 0.023X_3 - 0.031X_5 + 0.050X_6 - 0.003X_7 + 0.003X_9 + 0.002X_{10}$	8.50%
Melawai	$y = 34.881 - 0.024X_3 - 0.031X_5 + 0.055X_6 - 0.003X_7 + 0.003X_9 + 0.002X_{10}$	8.34%
Kebayoran Lama	$y = 33.648 - 0.022X_3 - 0.030X_5 + 0.068X_6 - 0.003X_7 + 0.002X_9 + 0.003X_{10}$	7.43%
Grogol Selatan	$y = 30.939 - 0.021X_3 - 0.029X_5 + 0.088X_6 - 0.003X_7 + 0.002X_9 + 0.003X_{10}$	4.52%
Pela Mampang	$y = 34.171 - 0.022X_3 - 0.031X_5 + 0.045X_6 - 0.003X_7 + 0.003X_9 + 0.002X_{10}$	9.37%

Based on Tables 15 and 16 above, the model used for cluster 1 is Pengadegan, and the model in cluster 4 is the Grogol Selatan. The model is subsumed by all areas in the cluster with the smallest MAPE value, which are 3.36% and 4.52%. An MAPE value below 5% is an excellent level of accuracy.

3.6. Mapping area with GIS

Based on the selection of variables, it is possible to map the significance distribution variables in each cluster with GIS, which can be seen in Figures 8 and 9. Using GIS, mapping areas based on significant variables can visualize, analyze, and understand how certain variables affect the load density in each cluster. In cluster 1, shown in Figure 8, the GDP variable significantly affects all regions in the cluster. Furthermore, industrial land use in the three areas does not have a significant effect on the cluster.

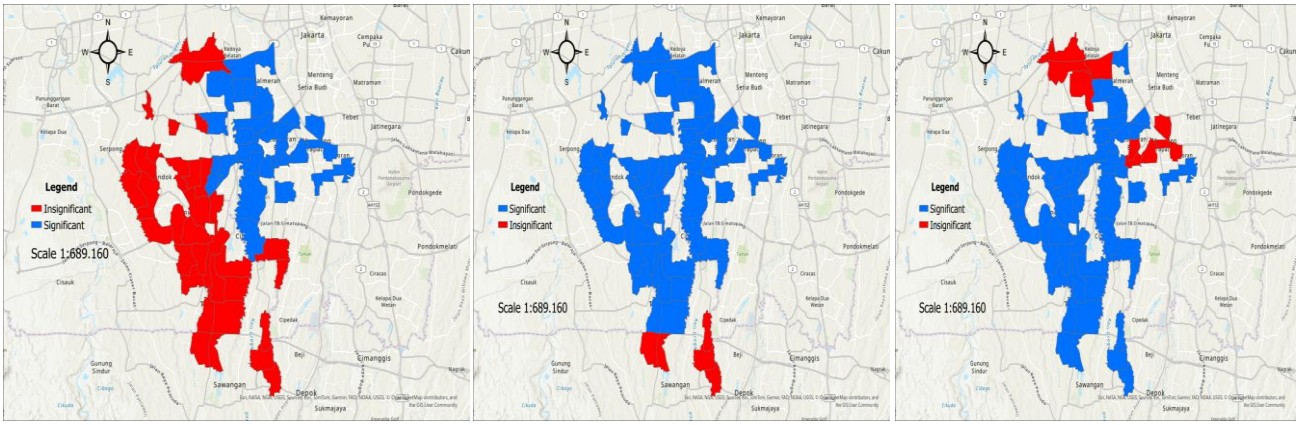
Figure 9 shows that the business sector load is a variable that significantly influences almost all areas in cluster 4; only six areas have a minor influence on the cluster. In Figure 8, the social sector load also shows the same situation; only nine regions have no significant effect. Additionally, where the previous area was not too significant, the influence of residential and social loads became an area where social loads influenced the electricity load. Besides that, the mapping conducted for the residential sector electricity load variable, as shown in Figure 9, shows that the distribution of residential load significantly influences 50% of the area in cluster 4.



(a) (b) (c)

Figure 8. Significance variables distribution of cluster 1: (a) Land use industry; (b) GDP and (c) Commercial load.

For the grid profile of the Pengadegan area as the modeled area in cluster 1, all variables significantly influence the load density, as shown in Figure 8. Pengadegan is geographically in the lowlands, with an average altitude of 15–20 meters above sea level. It is a large social area, with most of its electricity load influenced by commercial loads. In addition, the area has a high density consistent with the load density data, indicating a high value for this cluster.



(a) (b) (c)

Figure 9. Significance variables distribution of cluster 4: (a) Residential load; (b) Commercial load and (c) Social load.

On the other hand, the profile of the South Grogol area as a model area in cluster 4 shows that all significant variables influence load density, as shown in Figure 9. This scene is geographically in the lowlands, which have an average height of 5–15 meters above sea level. It is also a relatively large social area with an electrical load influenced by residential, commercial, and social loads. In addition, the area has a high density; this aligns with the load density data, including high values for areas in cluster 4.

3.7. Load forecasting

The load growth projection is based on the previously obtained GWR model, namely

$$y_{c\text{-}1} = 38.236 - 0.033\ X_3 - 0.042\ X_5 - 0.135\ X_6 + 0.007\ X_9$$

$$y_{c\text{-}4} = 30.939 - 0.021\ X_3 - 0.029\ X_5 + 0.088\ X_6 - 0.003\ X_7 + 0.002\ X_9 + 0.003\ X_{10}$$

Where $y_{c\text{-}1}$: Load density model equation cluster 1 and $y_{c\text{-}4}$: Load density model equation cluster 4.

To obtain the growth of load density each year based on the load density model obtained previously, it is first necessary to calculate the trend of each variable (except the land use variable) to get a growth model for each year of each variable. Based on the trend of each variable obtained, the variable growth trend model is used to forecast the load density using the load density model obtained previously ($y_{c\text{-}1}$ and $y_{c\text{-}2}$ equations), as shown in Figures 10 and 11.

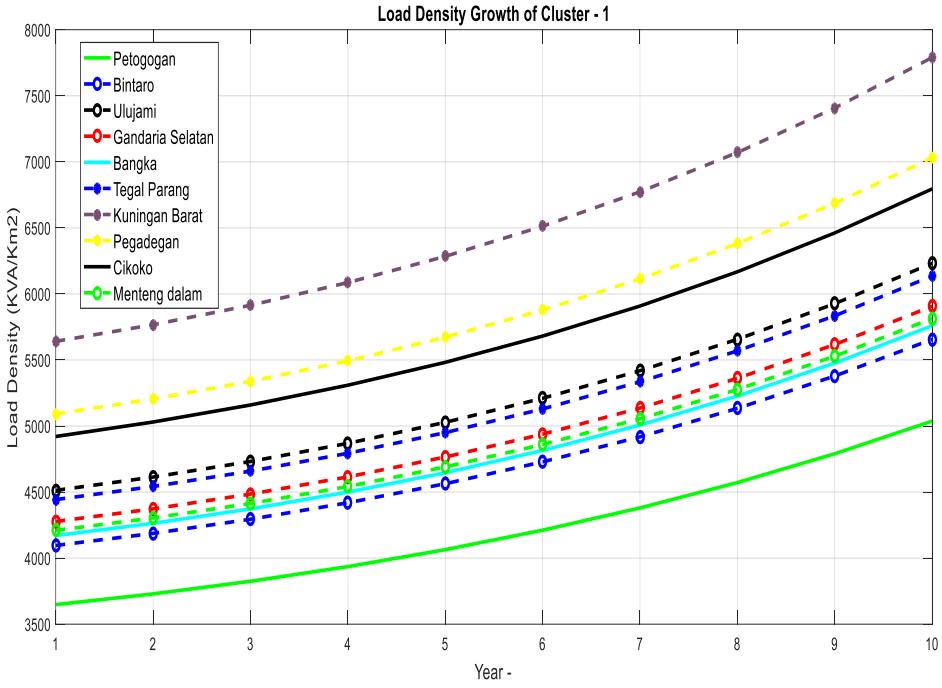


Figure 10. Grid load forecasting of cluster 1.

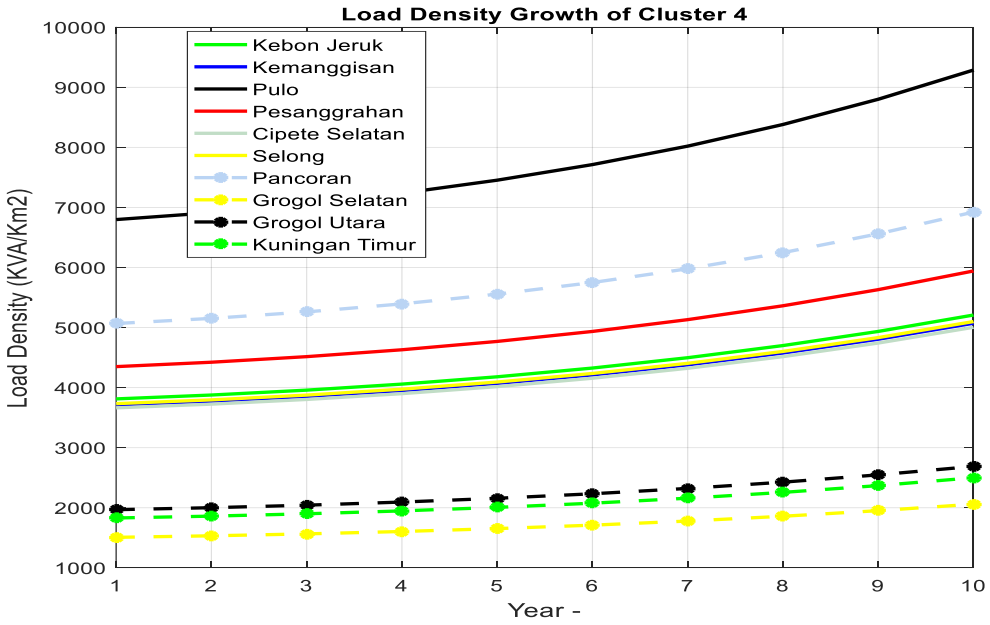
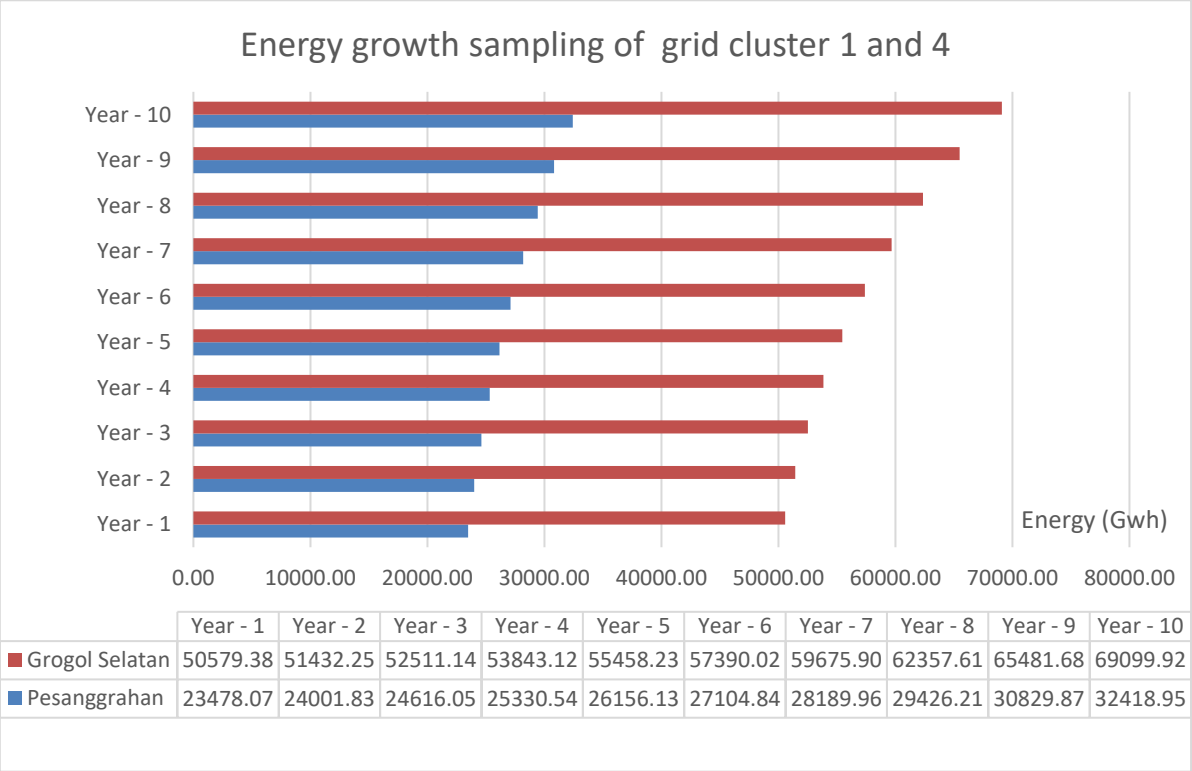


Figure 11. Load density growth of each cluster 4.

The results of this load density growth are further elaborated and broken down to obtain energy demand growth for each area in each cluster. Figure 12 shows a sampling of the energy demand growth results for each year for each grid according to the model representing the cluster, namely the Pengadegan area for cluster 1 and Grogol Selatan for cluster 4. Overall energy demand forecasting results for each area in each cluster can be seen in Appendix A.



	Year - 1	Year - 2	Year - 3	Year - 4	Year - 5	Year - 6	Year - 7	Year - 8	Year - 9	Year - 10
Grogol Selatan	50579.38	51432.25	52511.14	53843.12	55458.23	57390.02	59675.90	62357.61	65481.68	69099.92
Pesanggrahan	23478.07	24001.83	24616.05	25330.54	26156.13	27104.84	28189.96	29426.21	30829.87	32418.95

Figure 12. Energy demand growth in cluster 1 and cluster 4.

The overall load density growth for each cluster 1 and cluster 4 is shown in Figure 13 below:

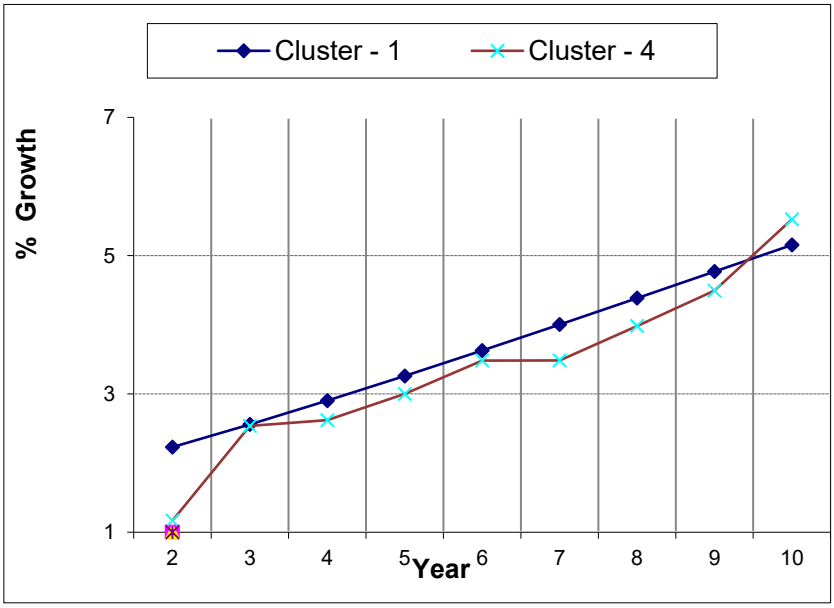


Figure 13. Load density growth in cluster 1 and cluster 4.

Based on the results of load growth forecasts through the micro spatial method, the average load density growth for cluster 1 in the next ten years is 3.65% and 3.37% for cluster 4. This growth

percentage is not much different compared to the Electricity Supply Business Plan by the State Electricity Company (RUPTL) for the Greater Jakarta area for 2021–2030, which is 3.724%. The load density results are then converted into load per area as shown in Figure 13.

The average growth of the regions is lower than the RUPTL because some of the areas in clusters 1 and 4 are in a different region, Tangerang, which has a load growth of 3.5%. Therefore, for cluster 4, the growth is lower because there are more Tangerang areas in cluster 4 than in cluster 1. However, if we calculate the MAPE value of the forecasting results with the existing RUPTL, the average error value is less than 2%. Besides that, the load forecast results show that areas in the same cluster will have the same distribution of energy demand growth patterns over the next 10 years, as shown in Table 17.

Table 17. Energy demand growth forecast for each cluster.

Year	Total of Cluster Energy Demand (GWh)	
	cluster 1	cluster 4
1	966522	2783902
2	990177	2830844
3	1017199	2890226
4	1046941	2963539
5	1080161	3052435
6	1119004	3158761
7	1162655	3284577
8	1212225	3432179
9	1268009	3604128
10	1331745	3803277

The energy demand growth of each cluster is obtained from the sum of energy from each region of each cluster. Figure 9 shows the distribution of load growth for each area in each cluster, which can be seen in Appendix A. According to the study, using GWR and GIS offers essential insights into electrical load's properties and regional distribution. This method makes it possible to estimate the number of load centers in each area precisely, reflecting the geographic structure, and to project load growth in smaller areas more accurately.

4. Conclusions

This micro-spatial energy demand forecasting method, based on multivariate analysis and GIS integration, is a development of the macro- and sectoral spatial load forecasting method. Spatial analysis identifies load distribution patterns by region. Multivariate analysis uncovers factors affecting load patterns. By identifying significant variables, accurate electricity consumption prediction models can be developed. Integrating GIS in load forecasting helps manage spatial data and identify spatial factors influencing electricity demand.

Based on the multivariate analysis, six of the ten variables significantly influence the load density. Those variables are land use: industry (X_3), land use: social (X_5), GDP (X_6), residential load (X_7), business load (X_9), and social load (X_{10}). The smallest AIC value is 292.52, and the feasibility of the

R-squared model reaches 0.9796.

Furthermore, GWR produces a suitable forecasting model by considering significant variables. GWR models are used to determine models that can represent clusters based on the selected model, namely the model in the South Grogol area, with the smallest MAPE value of 4.52%. Then, the model obtained is used to forecast energy demand, resulting in an average load density growth for the 64 Kebayoran network area in the next ten years of 3.54%. The proposed method is proven to show patterns that are not visible through traditional methods, can identify them, produce more accurate load distribution mapping, improve accuracy, and provide policy recommendations.

As a continuation of this research, further model development is directed at integrating a hybrid approach that combines GWR with machine learning algorithms such as Long Short-Term Memory (LSTM) and Random Forest to capture spatial-temporal dynamics simultaneously and more accurately. In spatial expansion, SVM-based clustering techniques are applied to adaptively classify new areas into existing clusters, along with settlement growth and changes in electricity system infrastructure. The model can be designed to be responsive to energy transition scenarios such as decarbonisation and electric vehicle adoption, and can be scaled through cloud computing to support electricity system planning and renewable energy integration in Java Island with precision and sustainability.

Use of AI tools declaration

The authors declare that the research conducted and presented in this article has not used AI tools at any stage of the research process.

Acknowledgments

The authors received no financial support for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that there are no conflicts of interest in the writing and publication of this work.

Author contributions

Adri Senen: Writing—original draft, visualization, mapping problem, methodology, investigation, formal analysis, data curation, analysis, conceptualization. Jasrul Jamani Jamian: Writing—review & editing, supervision, and conceptualization.

Appendix A.

Grid load growth of each cluster.

Cluster	Grid	Energi Growth (GWh)									
		Year-1	Year-2	Year-3	Year-4	Year-5	Year-6	Year-7	Year-8	Year-9	Year-10
1	Duri Kepa	72734	74356	76259	78473	81030	83969	87331	91161	95509	100432
	Meruya Selatan	49739	50848	52149	53663	55412	57422	59721	62340	65313	68680
	Palmerah	37934	38781	39773	40927	42261	43794	45547	47545	49813	52380
	Cipete Utara	34887	35665	36578	37640	38867	40276	41889	43726	45811	48173
	Petogogan	15390	15734	16136	16605	17146	17768	18479	19289	20210	21251
	Bintaro	91423	93463	95854	98637	101851	105546	109771	114585	120051	126239
	Ulujami	37635	38475	39460	40605	41928	43449	45189	47170	49420	51968
	Gandaria Seltan	37141	37969	38941	40071	41377	42878	44595	46550	48771	51285
	Bangka	67516	69022	70789	72843	75218	77946	81066	84621	88658	93228
	Tegal Parang	22886	23396	23995	24692	25496	26421	27479	28684	30052	31601
	Kuningan Barat	27118	27723	28432	29258	30211	31307	32560	33988	35610	37445
	Pengadegan	23478	24002	24616	25331	26156	27105	28190	29426	30830	32419
	Cikoko	17380	17767	18222	18751	19362	20064	20868	21783	22822	23998
	Menteng Dalam	53284	54473	55867	57489	59362	61515	63978	66784	69970	73576
	Pinang	1934	1977	2028	2087	2155	2233	2322	2424	2540	2671
	Pondok Pucung	58506	59811	61342	63122	65179	67543	70248	73328	76826	80786
	Pondok Aren	35111	35895	36813	37882	39116	40535	42158	44007	46106	48482
	Lengkong Wtan	36010	36814	37756	38852	40118	41573	43238	45134	47287	49724
	Pondok Jagung	2933	2998	3075	3164	3267	3386	3522	3676	3851	4050
	Sarua	46488	47525	48741	50156	51791	53669	55818	58266	61045	64192

Continued on next page.

Cluster	Grid	Energi Growth (GWh)									
		Year-1	Year-2	Year-3	Year-4	Year-5	Year-6	Year-7	Year-8	Year-9	Year-10
1	Cireundeu	39208	40083	41109	42302	43681	45265	47077	49142	51486	54140
	Cempaka Putih	29374	30030	30798	31692	32725	33912	35270	36816	38573	40561
	Rempoa	27459	28071	28790	29625	30591	31700	32969	34415	36057	37915
	Pondok Petir	39324	40201	41230	42427	43809	45398	47216	49287	51638	54299
2	Sukabumi Utara	28727	29212	29825	30581	31498	32596	33894	35417	37191	39246
	Kelapa Dua	24691	25107	25634	26284	27073	28016	29132	30441	31966	33732
	Kebon Jeruk	58726	59716	60969	62516	64391	66634	69288	72402	76029	80230
	Srengseng	84349	85772	87571	89792	92486	95707	99519	103991	109201	115235
	Meruya Utara	109706	111556	113896	116785	120288	124478	129436	135253	142029	149877
	Kembngan Sltan	96479	98106	100164	102705	105786	109470	113831	118946	124905	131807
	Kemanggisan	42458	43174	44079	45198	46553	48175	50094	52345	54967	58005
	Gandaria Utara	27704	28172	28763	29492	30377	31435	32687	34156	35867	37849
	Pulo	42353	43067	43970	45086	46438	48056	49970	52215	54831	57861
	Melawai	37947	38587	39396	40395	41607	43057	44771	46783	49127	51842
	Gunung	31622	32155	32830	33662	34672	35880	37309	38986	40939	43201
	Selong	25918	26355	26908	27591	28419	29408	30580	31954	33555	35409
	Senayan	48787	49610	50650	51935	53493	55356	57561	60148	63161	66651
	Pesanggrahan	45022	45781	46742	47927	49365	51085	53119	55506	58287	61508
	Petukangan Utra	57832	58807	60040	61563	63410	65619	68232	71299	74871	79008
	Lebak Bulus	73610	74851	76421	78360	80710	83522	86848	90751	95298	100564
	Pondok Labu	82522	83914	85674	87847	90482	93634	97363	101739	106836	112739
	Cipete Selatan	41871	42577	43470	44572	45910	47509	49401	51621	54207	57202
	Pela Mampang	33347	33909	34621	35499	36564	37837	39344	41112	43172	45558
	Mampng Praptn	14285	14526	14830	15206	15663	16208	16854	17611	18493	19515
	Kalibata	38093	38735	39548	40551	41767	43222	44944	46963	49316	52041

Continued on next page.

Cluster	Grid	Energi Growth (GWh)									
		Year-1	Year-2	Year-3	Year-4	Year-5	Year-6	Year-7	Year-8	Year-9	Year-10
2	Rawajali	42654	43373	44283	45406	46768	48397	50325	52587	55221	58272
	Pancoran	30327	30839	31486	32284	33253	34411	35782	37389	39263	41432
	Pondok Pinang	160646	163354	166781	171012	176141	182277	189537	198054	207977	219469
	Kebyoran Seltan	74685	75944	77537	79504	81889	84741	88117	92076	96689	102032
	Kebayora Utara	52500	53385	54505	55887	57564	59569	61942	64725	67968	71723
	Cipulir	33649	34216	34934	35820	36894	38180	39700	41484	43563	45970
	Grogol Selatan	50579	51432	52511	53843	55458	57390	59676	62358	65482	69100
	Grogol Utara	66029	67142	68550	70289	72398	74919	77904	81404	85483	90206
	Kuningan Timur	61393	62428	63738	65355	67315	69660	72435	75690	79482	83873
	Gedong	45145	45906	46869	48058	49500	51224	53264	55658	58446	61676
	Parung Serab	1237	1258	1284	1316	1356	1403	1459	1525	1601	1690
	Sudimara Timur	1164	1184	1209	1240	1277	1321	1374	1436	1508	1591
	Pedurenan	1061	1079	1101	1129	1163	1203	1251	1308	1373	1449
	Kereo	1804	1835	1873	1921	1978	2047	2129	2225	2336	2465
	Babakan	826	840	858	880	906	938	975	1019	1070	1129
	Parigi Baru	50865	51723	52808	54148	55772	57715	60013	62710	65852	69491
	Pndk Kacng Brt	41304	42000	42881	43969	45288	46866	48732	50922	53473	56428
	Pndk Kcg tmr	40175	40852	41709	42767	44050	45584	47400	49530	52012	54886
	Perigi Lama	63559	64631	65987	67660	69690	72118	74990	78360	82286	86832
	Jurang Barat	40216	40894	41752	42811	44095	45631	47448	49580	52064	54941
	Jurang Timur	34755	35341	36082	36997	38107	39434	41005	42848	44994	47481
	Pondok Karya	43604	44339	45269	46417	47810	49475	51446	53758	56451	59570
	Pondok Betung	29764	30266	30901	31685	32635	33772	35117	36695	38533	40663
	Pamulang Timur	42147	42858	43757	44867	46213	47823	49727	51962	54565	57580

Continued on next page.

Cluster	Grid	Energi Growth (GWh)									
		Year-1	Year-2	Year-3	Year-4	Year-5	Year-6	Year-7	Year-8	Year-9	Year-10
2	Pondok cabe	78975	80307	81992	84071	86593	89610	93179	97366	102244	107894
	Pondok Cabe Ilir	63700	64774	66133	67811	69845	72278	75156	78534	82468	87025
	Kedaung	40723	41410	42279	43351	44652	46207	48047	50207	52722	55635
	Bambu Apus	36366	36979	37755	38712	39873	41262	42906	44834	47080	49682
	Jombang	43678	44415	45347	46497	47892	49560	51534	53850	56547	59672
	Sawah Baru	36526	37142	37921	38883	40050	41445	43096	45032	47288	49901
	Sarua indah	25032	25454	25988	26647	27447	28403	29534	30861	32407	34198
	Sawah	31929	32467	33148	33989	35008	36228	37671	39364	41336	43620
	Ciputat	22478	22857	23336	23928	24646	25504	26520	27712	29100	30708
	Cipayung	30524	31038	31690	32493	33468	34634	36013	37632	39517	41701
	Pisangan	49553	50389	51446	52751	54333	56226	58465	61093	64153	67698
	Pondok Ranji	41635	42337	43225	44322	45651	47241	49123	51330	53902	56880
	Rengas	22350	22727	23204	23792	24506	25360	26370	27555	28935	30534
	Grogol	5966	6067	6194	6351	6542	6770	7039	7356	7724	8151
	Krukut	35507	36106	36863	37798	38932	40288	41893	43776	45969	48509
	Limo	61261	62294	63601	65214	67170	69510	72279	75527	79310	83693
	Pangkalanjati	1478	1502	1534	1573	1620	1676	1743	1822	1913	2019
	Mampang	24853	25272	25802	26457	27251	28200	29323	30641	32176	33954
	Serua	39291	39953	40791	41826	43081	44581	46357	48440	50867	53678
	Kedaung	35939	36545	37311	38258	39405	40778	42402	44308	46527	49098

References

1. Evangelopoulos VA, Georgilakis PS (2022) Probabilistic spatial load forecasting for assessing the impact of electric load growth in power distribution networks. *Electr Power Syst Res* 207: 107847. https://doi.org/10.1016/j.epsr.2022.107847
2. Thakare S, Bokde ND, Feijóo-Lorenzo AE (2023) Forecasting different dimensions of liquidity in the intraday electricity markets: A review. *AIMS Energy* 11: 918–959. https://doi.org/10.3934/ENERGY.2023044
3. Jiang H, Dong Y, Dong Y, et al. (2024) Power load forecasting based on spatial-temporal fusion graph convolution network. *Technol Forecast Soc Change* 204: 123435. https://doi.org/10.1016/j.techfore.2024.123435
4. Zhao Z, Zhang Y, Yang Y, et al. (2022) Load forecasting via grey model-least squares support vector machine model and spatial-temporal distribution of electric consumption intensity. *Energy* 255: 124468. https://doi.org/10.1016/j.energy.2022.124468
5. Zambrano-Asanza S, Morales RE, Montalvan JA, et al. (2023) Integrating artificial neural networks and cellular automata model for spatial-temporal load forecasting. *Int J Electr Power Energy Syst* 148: 108906. https://doi.org/10.1016/j.ijepes.2022.108906
6. Maswanganyi N, Ranganai E, Sigauke C (2019) Long-term peak electricity demand forecasting in South Africa: A quantile regression averaging approach. *AIMS Energy* 7: 857–882. https://doi.org/10.3934/energy.2019.6.857
7. Senen A, Widyastuti C, Handayani O, et al. (2021) Development of micro-spatial electricity load forecasting methodology using multivariate analysis for dynamic area in Tangerang, Indonesia. *Pertanika J Sci Tech* 29: 2565–2578. https://doi.org/10.47836/pjst.29.4.18
8. El khantach A, Hamlich M, Belbounaguia N (2019) Short-term load forecasting using machine learning and periodicity decomposition. *AIMS Energy* 7: 382–394. https://doi.org/10.3934/ENERGY.2019.3.382
9. Pérez-Chacón R, Asencio-Cortés G, Troncoso A, et al. (2024) Pattern sequence-based algorithm for multivariate big data time series forecasting: Application to electricity consumption. *Future Gener Comput Syst* 154: 397–412. https://doi.org/10.1016/j.future.2023.12.021
10. Tambunan HB, Hakam DF, Prahastono I, et al. (2020) The challenges and opportunities of renewable energy source (RES) penetration in Indonesia: Case study of Java-Bali power system. *Energies* 13: 1–22. https://doi.org/10.3390/en13225903
11. Trull O, García-Díaz JC, Troncoso A (2020) Initialization methods for multiple seasonal holt-winters forecasting models. *Mathematics* 8: 1–16. https://doi.org/10.3390/math8020268
12. Zhao P, Hu W, Cao D, et al. (2024) Enhancing multivariate, multi-step residential load forecasting with spatiotemporal graph attention-enabled transformer. *Int J Electr Power Energy Syst* 160: 110074. https://doi.org/10.1016/j.ijepes.2024.110074
13. Nie Y, Li P, Wang J, et al. (2024) A novel multivariate electrical price bi-forecasting system based on deep learning, a multi-input multi-output structure and an operator combination mechanism. *Appl Energy* 366: 123233. https://doi.org/10.1016/j.apenergy.2024.123233

14. Huang Y, Hasan N, Deng C, et al. (2022) Multivariate empirical mode decomposition based hybrid model for day-ahead peak load forecasting. *Energy* 239: 122245. https://doi.org/10.1016/j.energy.2021.122245
15. Arjmand A, Samizadeh R, Dehghani Saryazdi M (2020) Meta-learning in multivariate load demand forecasting with exogenous meta-features. *Energy Effic* 13: 871–887. https://doi.org/10.1007/s12053-020-09851-x
16. Pearson K (1901) LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science* 2: 559–572. https://doi.org/10.1080/14786440109462720
17. Hu Y, Xia X, Fang J, et al. (2018) A multivariate regression load forecasting algorithm based on variable accuracy feedback. *Energy Procedia* 152: 1152–1157. https://doi.org/10.1016/j.egypro.2018.09.147
18. Efroymson MA (1970). Multiple regression analysis. In A. Ralston & H.S. Wilf (Eds.), Mathematical Methods for Digital Computers, 191–203. New York: Wiley.
19. Draper NR, Smith H (1981). Applied regression analysis (2nd ed.). New York: Wiley.
20. Brunsdon C, Fotheringham AS, Charlton ME (1996) Geographically weighted regression: A method for exploring spatial nonstationarity. *Geogr Anal* 28: 281–298. https://doi.org/10.1111/j.1538-4632.1996.tb00936.x
21. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9: 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
22. Chen HF (2009) In silico log p prediction for a large data set with support vector machines, radial basis neural networks and multiple linear regression. *Chem Biol Drug Des* 74: 142–147. https://doi.org/10.1111/j.1747-0285.2009.00840.x
23. Zhang Q, Wang N, Cheng J, et al. (2020) A stepwise downscaling method for generating high-resolution land surface temperature from AMSR-E data. *IEEE J Sel Top Appl Earth Obs Remote Sens* 13: 5669–5681. https://doi.org/10.1109/JSTARS.2020.3022997
24. Żogała-Siudem B, Jaroszewicz S (2021) Fast stepwise regression based on multidimensional indexes. *Inf Sci* 549: 288–309. https://doi.org/10.1016/j.ins.2020.11.031
25. Wang G, Sarkar A, Carbonetto P, et al. (2020) A simple new approach to variable selection in regression, with application to genetic fine mapping. *J R Stat Soc Ser B Stat Methodol* 82: 1273–1300. https://doi.org/10.1111/rssb.12388
26. Zeng S, Yan Z, Yang J (2022) Stepwise algorithm and new analytical model for estimating multi-parameter of energy piles from thermal response tests. *Energy Build* 256: 111775. https://doi.org/10.1016/j.enbuild.2021.111775
27. Subramanyam T, Donthi R, Satish Kumar V, et al. (2021) A new stepwise method for selection of input and output variables in data envelopment analysis. *J Math Comput Sci* 11: 703–715. https://doi.org/10.28919/jmcs/5205
28. Zhuang J, Guo Z, Gong X, et al. (2023) Spatial load forecasting method based on GIS and generative adversarial network. *2023 8th Asia Conference on Power and Electrical Engineering (ACPEE)*, 1965–1969. https://doi.org/10.1109/ACPEE56931.2023.10135732

29. Katruksa S, Jiriwibhakorn S (2019) Electricity load forecasting based on a geographic information system. *2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST)*, 1–4. https://doi.org/10.1109/ICEAST.2019.8802591
30. Messoudi L, Gouareh A, Settou B, et al. (2024) Modeling and forecasting energy consumption in Algerian residential buildings using a bottom-up GIS approach. *Energy Build* 317: 114370. https://doi.org/10.1016/j.enbuild.2024.114370
31. Zhao R, Zhan L, Yao M, et al. (2020) A geographically weighted regression model augmented by Geodetector analysis and principal component analysis for the spatial distribution of $PM_{2.5}$. *Sustain Cities Soc* 56: 102106. https://doi.org/10.1016/j.scs.2020.102106
32. Yasin H, Purhadi, Choiruddin A (2024) Spatial clustering based on geographically weighted multivariate generalized gamma regression. *MethodsX* 13: 102903. https://doi.org/10.1016/j.mex.2024.102903
33. Li W, Xu Q, Yi J, et al. (2021) Prediction and evaluation of forest fire in Yunnan of China based on geographically weighted logistic regression model [Preprint]. *Res Square*. https://doi.org/10.21203/rs.3.rs-997415/v1
34. Wu Y, Tang Z, Xiong S (2023) A unified geographically weighted regression model. *Spat Stat* 55: 100753. https://doi.org/10.1016/j.spasta.2023.100753
35. Tephiruk N, Hongesombut K, Damrongkulkamjorn P, et al. (2023) Spatial electric load forecasting using a geographic information system: A case study of Khon Kaen, Thailand. *2023 IEEE 3rd International Conference on Sustainable Energy and Future Electric Transportation (SEFET)*, 1–7. https://doi.org/10.1109/SeFeT57834.2023.10245044
36. Liu Y, Li Y, Xu J, et al. (2021) Grid-based spatial load forecasting method based on POI information mining. *2021 IEEE 4th International Electrical and Energy Conference (CIEEC)*, 1–6. https://doi.org/10.1109/CIEEC50170.2021.9510563
37. Vieira DAG, Silva BE, Menezes TV, et al. (2020) Large scale spatial electric load forecasting framework based on spatial convolution. *Int J Electr Power Energy Syst* 117: 105582. https://doi.org/10.1016/j.ijepes.2019.105582
38. Si C, Xu S, Wan C, et al. (2021) Electric load clustering in smart grid: Methodologies, applications, and future trends. *J Mod Power Syst Clean Energy* 9: 237–252. https://doi.org/10.35833/MPCE.2020.000472
39. Senen A, Putri TWO, Jamian JJ, et al. (2023) Fuzzy C-means clustering based on micro-spatial analysis for electricity load profile characterization. *Indones J Electr Eng Comput Sci* 30: 33–45. https://doi.org/10.11591/ijeecs.v30.i1.pp33-45
40. Bezdek JC (1981) Pattern recognition with fuzzy objective function algorithms. New York: Springer. https://doi.org/10.1007/978-1-4757-0450-1
41. Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20: 53–65. https://doi.org/10.1016/0377-0427(87)90125-7
42. Aswanuwath L, Pannakkong W, Buddhakulsomsiri J, et al. (2023) A hybrid model of VMD-EMD-FFT, similar days selection method, stepwise regression, and artificial neural network for daily electricity peak load forecasting. *Energies* 16: 1860. https://doi.org/10.3390/en16041860

43. Sethi R, Kleissl J (2020) Comparison of short-term load forecasting techniques. *2020 IEEE Conference on Technologies for Sustainability (SusTech)*, 1–6. https://doi.org/10.1109/SusTech47890.2020.9150490
44. Balzanella A, Verde R, Tenorio de Carvalho FA (2025) A cluster-wise regression method for distribution-valued data. *Knowl-Based Syst* 326: 113989. https://doi.org/10.1016/j.knosys.2025.113989
45. Ingdal M, Johnsen R, Harrington DA (2019) The Akaike information criterion in weighted regression of immittance data. *Electrochim Acta* 317: 648–653. https://doi.org/10.1016/j.electacta.2019.06.030
46. Jing X, Qinggui C, Haiqin Q, et al. (2020) Assessment of aero-engine service reliability based on Akaike information criterion. *2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, 636–639. https://doi.org/10.1109/ICMCCE51767.2020.00141
47. Portet S (2020) A primer on model selection using the Akaike Information Criterion. *Infect Dis Model* 5: 111–128. https://doi.org/10.1016/j.idm.2019.12.010
48. Wu Y, Tang Z, Xiong S (2023) A unified geographically weighted regression model. *Spat Stat* 55: 1–25. https://doi.org/10.1016/j.spasta.2023.100753
49. Chen F, Leung Y, Wang Q, et al. (2024) Spatial non-stationarity test of regression relationships in the multiscale geographically weighted regression model. *Spat Stat* 62: 100846. https://doi.org/10.1016/j.spasta.2024.100846

AIMS Press