# A Multiple Linear Regression Model Based on Spatial Temperature and Humidity Clustering for Building Energy Use Intensity Forecasting

**Imam Arif Rahardjo**
Department of Electrical Engineering, Faculty of Engineering, University of Indonesia, Indonesia
imam.arif21@ui.ac.id

**Iwa Garniwa**
Department of Electrical Engineering, Faculty of Engineering, University of Indonesia, Indonesia
iwa@eng.ui.ac (corresponding author)

**Budi Sudiarto**
Department of Electrical Engineering, Faculty of Engineering, University of Indonesia, Indonesia
budi.sudiarto@ui.ac.id

**Faiz Husnayain**
Department of Electrical Engineering, Faculty of Engineering, University of Indonesia, Indonesia
faiz.h@ui.ac.id

**Fahmi Firdaus Angkasa**
Department of Electrical Engineering, Faculty of Engineering, University of Indonesia, Indonesia
fahmi.angkasa@ui.ac.id

**Pidanic Jan**
Department of Electrical Engineering and Informatics, University of Pardubice, Czech Republic
jan.pidanic@upce.cz

**Nurwan Reza Fachrurrozi**
Department of Electrical Engineering, Telkom University, Jakarta, Indonesia
nurwan@telkomuniversity.ac.id

*Received: 24 February 2026 | Revised: 24 March 2026 and 3 April 2026 | Accepted: 4 April 2026*

 *DOI: https://doi.org/10.48084/etasr.18284*

## ABSTRACT

**Accurate building Energy Use Intensity (EUI) is crucial for improving building energy efficiency, particularly in countries with significant climatic diversity such as Indonesia. However, many current regression-based energy prediction models rely on national-scale climate data, implicitly assuming a uniform relationship between climate variables and energy consumption across regions. This assumption may reduce prediction accuracy in geographically heterogeneous environments. To address this limitation, this study focuses on office buildings, whose operational cooling demand is strongly influenced by local climatic conditions, necessitating the integration of spatially clustered temperature and humidity data with Multiple Linear Regression (MLR) modeling. Provincial climate data from 38 Indonesian provinces are first classified into homogeneous climate zones using the Fuzzy C-Means (FCM) clustering method. FCM clustering was chosen because it can produce smoother transitions in differences across climate data and handle overlapping data through membership degrees in spatial clustering, compared to k-means clustering. Subsequently, cluster-specific MLR models are developed to predict EUI within each identified climate group. The performance of the cluster-scale model is compared and validated using the coefficient**

**of determination ($R^2$), along with Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), to assess the variance in EUI explained by climate, relative to a conventional national-scale regression model. The results show that the cluster-specific regression models consistently improve the forecasting accuracy of the national-scale models across all climate clusters. The MAE decreased from 0.19–0.34 at the national scale to 0.04–0.13 at the cluster scale, while the RMSE decreased from 0.25–0.43 to 0.05–0.16. Overall, the findings indicate that aligning regression modeling with climate-based clustering significantly improves accuracy. Furthermore, these findings provide a basis for policymakers and building managers to develop climate-specific energy efficiency strategies. Although climate variables were the primary predictors in this study, operational factors, such as occupancy, HVAC control settings, and building envelope characteristics, could also be integrated in future studies.**

***Keywords-energy use intensity; cluster; temperature and humidity; multiple linear regression; heterogenity; accuracy***

# I. INTRODUCTION

The building sector contributes to global energy consumption and global emissions. It accounts for 30% of the total energy consumption and 27% of the total GHG emissions [1]. In [2], this energy use accounted for nearly 40% of the total energy use. This substantial contribution underscores the urgency of improving energy efficiency in the building sector to support sustainability objectives. Building energy efficiency performance is commonly evaluated using Energy Use Intensity (EUI), defined as the amount of energy consumed per unit building area [3, 4], which serves as a performance indicator for building energy benchmarking [5]. Consequently, the accuracy of EUI is critical to ensure reliable performance evaluation and informed decision-making. EUI can be determined either directly from measurements or through predictive approaches. However, given the importance of forward-looking analysis in planning and policy development, predictive methods are increasingly being adopted in research and energy-efficiency studies.

Building on the importance of accurate EUI estimation, numerous predictive studies incorporate microclimate variables as primary factors in assessing building energy performance [6]. Among these variables, temperature is the most influential factor, particularly in regions with significant cooling or heating demand [7, 8]. Warmer temperatures resulted in an average 16%–46% increase in space cooling and an 8%–17% reduction in space heating [9]. Every 1°C increase in ambient temperature leads to a 25% increase in cooling load [10]. Since temperature and humidity interact simultaneously within climatic systems, their combined effect further amplifies building energy consumption. In addition, the cooling energy consumption intensity in the Hot Summer and Cold Winter (HSCW) and Hot Summer and Warm Winter (HSWW) zones was twice as high as that seen in zone Cold (C), nine times higher than the value in zone Severe Cold (SC), and 36 times higher than the value in zone Moderate (M) [11].

Although previous research has incorporated climate variables into energy consumption predictions, several important gaps remain. First, spatial microclimate variation is often overlooked, as many studies apply a single aggregated or averaged climate representation across large geographical areas. Such an approach may mask regional heterogeneity and reduce predictive accuracy. To address this limitation, this study integrates spatial clustering of temperature and relative humidity with MLR to better capture localized climatic characteristics. The use of local climate characteristics builds on a previous study that showed that applying subdistrict-level weather data improved prediction accuracy by 14.54% compared to the Typical Meteorological Year (TMY) [12].

Second, many studies focus on predicting total energy consumption with various models. An investigation of energy consumption in hospitals in Poland, stratified by climatic zone, showed that energy consumption was higher in the warmest zone than in the coldest, as determined by multivariate backward stepwise regression [13]. Previous approaches have also integrated indoor and outdoor climate conditions to capture the complexity of energy demand influenced by building characteristics and external weather conditions, using the BAIT index with several deep neural network models (LSTM, GRU, Bi-LSTM, Bi-GRU) [14]. In [11], using the bottom-up method through consideration of spatial differences in zones, differences were detected in the intensity of urban residential buildings, where SC was 4.07 kgce/($m^2$·a), C was 6.8 kgce/($m^2$·a), HSCW was 6.53 kgce/($m^2$·a), HSWW was 6.5 kgce/(m2·a), and M was 5.21 kgce/($m^2$·a). Likewise, a GAM-based M&V methodology was applied to study the impact of weather parameters, such as temperature and humidity, on energy consumption [15]. Similarly, authors in [16] tried to forecast electricity consumption in school buildings using the MLR model and reported $R^2$ values of 74% for the training set and 77% for the testing set. Limited attention has been given to directly forecasting EUI as the primary performance indicator. This study, therefore, develops a framework, specifically designed for EUI prediction within spatially coherent climate clusters using MLR. The geographical context remains underexplored, particularly in tropical archipelagic countries such as Indonesia, where substantial climatic diversity exists across provinces, ranging from hot, humid coastal regions to cooler highland areas. These microclimatic differences naturally lead to variations in building energy demand, highlighting the importance of spatially clustered, differentiated modeling and analysis. Similarly, research on buildings with varying shapes, heights, temperatures, and wind speeds, analyzed by spatial clusters in each region, showed different EUI values [17]. Table I presents a detailed comparison of the proposed approach with existing works.

This research proposes a climate-aligned predictive framework tailored to Indonesia, which comprises 38 provinces characterized by heterogeneous temperature and humidity profiles. The key contributions of this study include identifying spatial variances in temperature and humidity across

Indonesian provinces and applying clustering techniques to group regions with similar climatic characteristics, thereby achieving homogeneous climate zones. Based on these resulting homogeneous clusters, cluster-specific regression models are developed to generate EUI predictions at the cluster scale, enabling modeling that reflects localized climatic conditions. Finally, a comparative evaluation with national-scale modeling is conducted, in which the predictive performance of the cluster-based models is systematically compared against a conventional non-cluster (national-scale) regression model to quantify the improvement in accuracy.

TABLE I. COMPARISON OF PREVIOUS STUDIES AND THE PROPOSED MODEL APPROACH

| Literature | Climate variables considered | Treatment of spatial climate | Modeling approach | Output | Spatial scale | Country |
|---|---|---|---|---|---|---|
| [11] | Climate zone | Considered as a zone (SC, C, HSCW, HSWW, and M) | Bottom-up simulation | Energy consumption intensity | Countries and regions | China and developed countries (the USA, EU, and Japan) |
| [12] | Temperature, solar radiation, and relative humidity | Localized meteorological from WRF Simulation | UBEM simulation | Energy consumption | District | Gangnam, Seoul |
| [13] | Not considered, only the zone of climate | Climate zone–based division (4 NUTS-2 zones) | Multiple regression - backward stepwise | Electrical and thermal consumption | National | Poland |
| [14] | Temperature (indoor/outdoor) | Not explicitly considered | Deep neural network | Load forecasting | City/regional | Singapore |
| [15] | Measurement of temperature and humidity | Not explicitly considered | GAM | Energy saving | Regions | Chongqing, China |
| [16] | Measurement of temperature and humidity | Not explicitly considered | MLR | Electricity consumption | Region | France |
| [17] | Temperature, humidity, and wind speed | Implicit (pre-defined climate zones) | Regression-stepwise prediction | Heating EUI | Regions | Harbin, China |
| This research (proposed) | Temperature and relative humidity | Explicit spatial clustering of microclimate zones (Fuzzy C-Means (FCM) on T-H patterns) | Cluster-specific MLR for EUI forecasting | EUI prediction | Multi-regional/ national | Indonesia |

## II. METHODOLOGY

### A. *Overview of the Methodological Framework*

This study considers meteorological heterogeneity, particularly in temperature and humidity variables, across 38 provinces in Indonesia. Therefore, given the heterogeneity of the data, a clustering approach for the temperature and humidity variables was chosen to provide more accurate predictions of EUI in office buildings. This study was conducted in three stages: data preparation, climate-based regression, and model comparison (Figure 1).

### B. *Data Preparation*

Data preparation begins with the collection of raw data from various sources. Data types include temperature, humidity, and EUI data for each province. Temperature, humidity, and energy consumption data were obtained through direct measurements (field experiments) using temperature and humidity meters and energy meters installed in the observed buildings. Measurements were conducted continuously to capture variations in temperature and humidity, as well as actual energy consumption patterns in the buildings. Temperature and humidity data are the independent variables in this study. EUI is the dependent variable, obtained by dividing a building's electrical energy consumption by the area of the building that uses electricity:

$$\mathrm{EUI} = \frac{E}{A} \tag{1}$$

where $E$ is the energy consumption, and $A$ is the area of the building. The collected EUI data are then normalized to ensure a more uniform scale across provinces, making the data more stable.

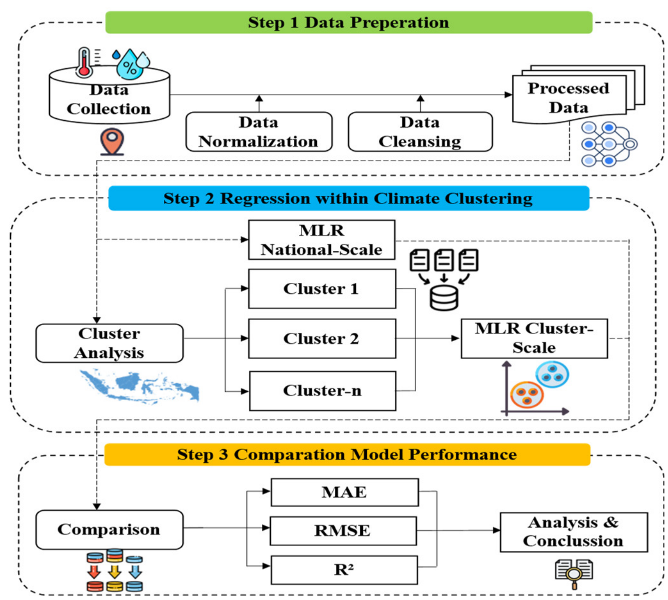


Fig. 1. Data flow and analytical steps of research.

### C. *Regression with Climate*

The processed data are then regressed onto the dependent variable using the independent variables. The regression of the independent variables, namely temperature and humidity, which form the EUI as the dependent variable, is carried out using two models: the Scale Cluster regression model and the Scale non-Cluster (national) regression model. In the Scale Cluster regression model, the first step is to analyze the clusters. Cluster analysis is used to determine the number of clusters using the FCM method, based on the homogeneity of the temperature and humidity variables across the provinces of Indonesia. Cluster determination in a province is based on the highest membership value. Thus, each cluster will represent a collection of provinces with homogeneous temperature and humidity characteristics. The provinces are grouped by cluster, and then regressed against temperature and humidity to predict EUI. This will result in as many regressions as there are clusters formed. Meanwhile, in the Scale non-Cluster (national) model, regression is performed directly on temperature and humidity values from all provinces to predict EUI, without first creating clusters. The general form of the regression model is expressed in:

$$EUI = \beta_0 + \beta_1 T + \beta_2 H \quad (2)$$

where $T$ and $H$ are independent variables, with $T$ representing temperature and $H$ humidity, $\beta_0$ is the intercept at $T$=0 and $H$=0, and $\beta_1$ and $\beta_2$ are weight coefficients.

### *D. Comparative Model Performance and Validation*

Each model was evaluated using the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination ($R^2$) metrics. MAE is an error metric calculated as the average of the absolute differences between the predicted and actual values. MAE is defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}[y_i - \hat{y}_i] \quad (3)$$

RMSE is an error metric calculated as the square root of the mean squared error between the predicted and actual values, and is defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[y_i - \hat{y}_i]^2} \quad (4)$$

$R^2$ is a metric that assesses the variation between predictions and actual values, and is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}[y_i - \hat{y}_i]^2}{\sum_{i=1}^{n}[y_i - \bar{y}]^2} \quad (5)$$

The results of this evaluation are then compared to determine the best-performing model. A good model will have low MAE and RMSE and high $R^2$.

## III. RESULTS

### *A. Province-Based Climate Clusters Based on Temperature and Humidity*

Temperature and humidity data from all 38 provinces in Indonesia show that these two variables are spatially heterogeneous across regions, as depicted in Table II. A closer examination of these two variables reveals that humidity dominates regional heterogeneity across Indonesia. Considering this heterogeneity, temperature and humidity cannot be used directly for predictions. If used, the prediction results will be inaccurate. This means that temperature and humidity variables at the national or state level in Indonesia will bias the prediction variables. Given the differences in temperature and humidity across Indonesia's provinces, a cluster analysis is required to group the 38 provinces into homogeneous clusters. Cluster analysis based on temperature and humidity shows that five clusters form a homogeneous province, as illustrated in Figure 2. This cluster separation aims to strengthen the cluster-based regression analysis and categories, as displayed in Figure 3.

TABLE II. CLIMATE-BASED LOCATION CATEGORIES DERIVED FROM PROVINCIAL TEMPERATURE, HUMIDITY, AND EUI CHARACTERISTICS

| Province | X1 | X2 | X3 | X4 | X5 |
|---|---|---|---|---|---|
| Aceh | 0.025 | 0.041 | 0.073 | 0.157 | 0.704 |
| North Sumatra | 0.016 | 0.021 | 0.030 | 0.084 | 0.850 |
| West Sumatra | 0.005 | 0.006 | 0.007 | 0.036 | 0.945 |
| Riau | 0.023 | 0.036 | 0.046 | 0.550 | 0.345 |
| Riau Islands | 0.009 | 0.011 | 0.013 | 0.073 | 0.894 |
| Jambi | 0.016 | 0.034 | 0.035 | 0.850 | 0.065 |
| Bengkulu | 0.015 | 0.021 | 0.032 | 0.093 | 0.839 |
| South Sumatra | 0.009 | 0.043 | 0.903 | 0.026 | 0.019 |
| Bangka Belitung | 0.022 | 0.044 | 0.054 | 0.760 | 0.120 |
| Lampung | 0.009 | 0.050 | 0.897 | 0.027 | 0.018 |
| Jakarta | 0.013 | 0.061 | 0.867 | 0.034 | 0.025 |
| West Java | 0.018 | 0.860 | 0.082 | 0.027 | 0.013 |
| Central Java | 0.020 | 0.905 | 0.036 | 0.029 | 0.011 |
| Yogyakarta | 0.060 | 0.781 | 0.062 | 0.070 | 0.026 |
| East Java | 0.032 | 0.803 | 0.098 | 0.045 | 0.021 |
| Banten | 0.106 | 0.351 | 0.338 | 0.119 | 0.086 |
| Bali | 0.013 | 0.079 | 0.836 | 0.046 | 0.026 |
| West Nusa Tenggara | 0.036 | 0.247 | 0.553 | 0.110 | 0.055 |
| East Nusa Tenggara | 0.022 | 0.111 | 0.761 | 0.065 | 0.041 |
| West Kalimantan | 0.080 | 0.142 | 0.095 | 0.571 | 0.113 |
| Central Kalimantan | 0.019 | 0.040 | 0.033 | 0.858 | 0.050 |
| South Kalimantan | 0.023 | 0.040 | 0.064 | 0.276 | 0.597 |
| East Kalimantan | 0.020 | 0.039 | 0.043 | 0.791 | 0.107 |
| North Kalimantan | 0.046 | 0.045 | 0.053 | 0.167 | 0.688 |
| North Sulawesi | 0.064 | 0.076 | 0.061 | 0.616 | 0.182 |
| Central Sulawesi | 0.041 | 0.043 | 0.039 | 0.654 | 0.223 |
| South Sulawesi | 0.028 | 0.065 | 0.062 | 0.763 | 0.082 |
| Southeast Sulawesi | 0.019 | 0.028 | 0.022 | 0.881 | 0.050 |
| Gorontalo | 0.045 | 0.042 | 0.031 | 0.779 | 0.104 |
| West Sulawesi | 0.027 | 0.031 | 0.029 | 0.739 | 0.174 |
| Maluku | 0.034 | 0.035 | 0.037 | 0.343 | 0.551 |
| North Maluku | 0.043 | 0.041 | 0.036 | 0.687 | 0.193 |
| West Papua | 0.766 | 0.070 | 0.036 | 0.081 | 0.046 |
| Southwest Papua | 0.173 | 0.090 | 0.073 | 0.349 | 0.315 |
| Papua | 0.569 | 0.107 | 0.053 | 0.191 | 0.081 |
| South Papua | 0.929 | 0.022 | 0.010 | 0.026 | 0.013 |
| Central Papua | 0.560 | 0.213 | 0.077 | 0.096 | 0.055 |
| Highland Papua | 0.563 | 0.179 | 0.084 | 0.106 | 0.068 |

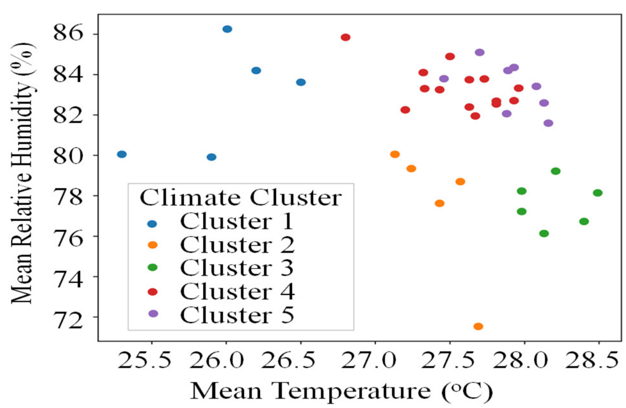


Fig. 2. Temperature and humidity feature space illustrating a cluster.

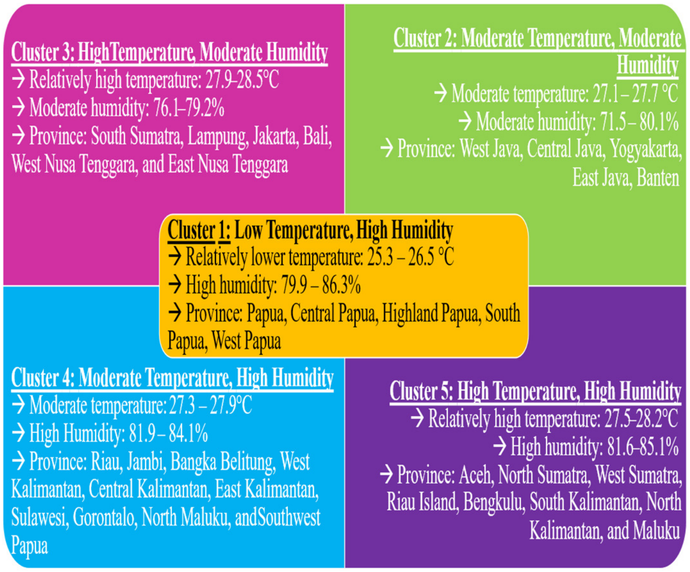


Fig. 3. Climate and energy-based cluster categories.

### *B. Development of EUI Prediction Equation Based on MLR Regression - Clustered*

EUI prediction is performed after clustering Indonesian provinces based on temperature and humidity, among the climate variables. This prediction uses MLR on the temperature and humidity in the formed provincial cluster groups. Because five cluster groups were formed, this study obtained five MLR models, one for each cluster, to predict the EUI. The resulting clusters are presented in Figure 4.

$$EUI_{Non\ Cluster} = 21.735 - 0.397\,T + 0.111\,H \quad (6)$$

### *C. Comparison of Prediction Performance Between Cluster and Non-Cluster (National) Models*

EUI predictions for each province, generated with the Scale Cluster and non-Cluster (national scale) models based on temperature and humidity variables, were evaluated for accuracy using MAE and RMSE. The results of the Scale Cluster model compared to the Scale Non-Cluster (national) model are shown in Figure 5.

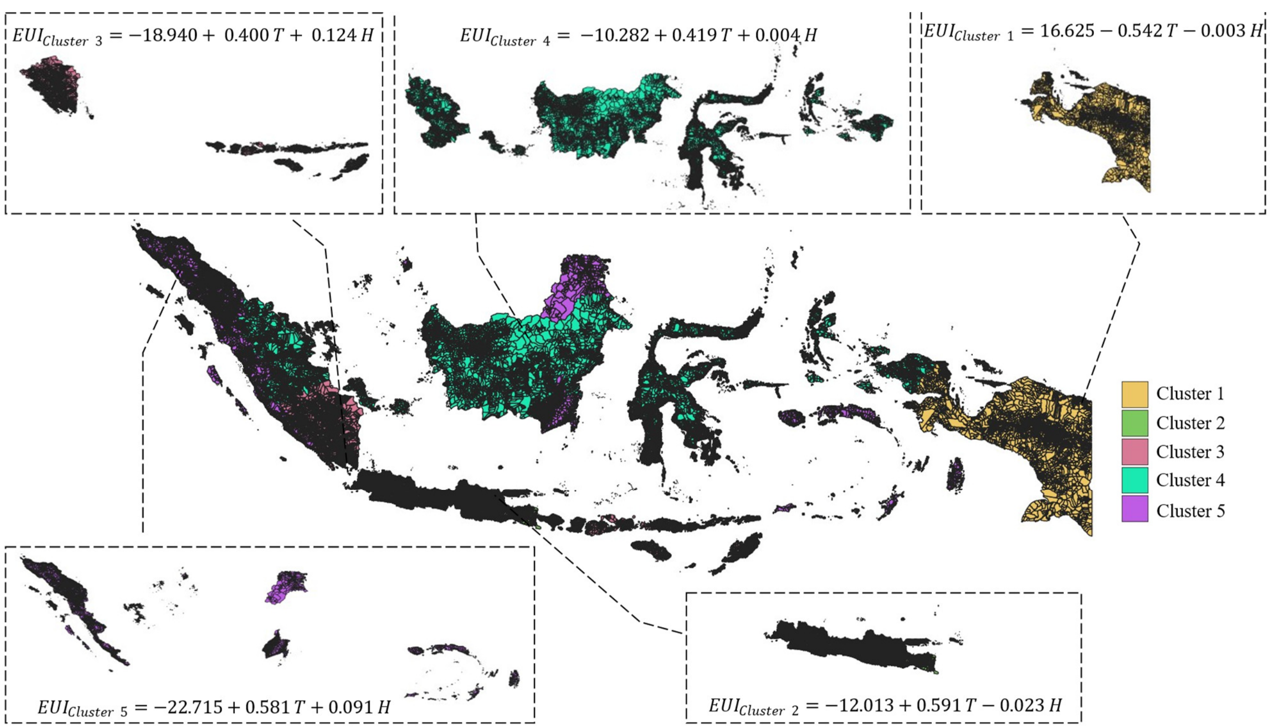


Fig. 4. Relationship between climate–energy cluster categories and the resulting MLR regression equation.

### *D. Mean Absolute Error*

The MAE metric for EUI predictions across 38 provinces is lower with the clustered MLR model than with the unclustered national MLR. The average provinces grouped in the scale cluster 1-5 have a lower MAE of 0.15018 (60%) compared to those not clustered on the national scale, ranging from 0.2465 to 0.0963.

### *E. Root Mean Square Error*

The RMSE results are lower for the clustered MLR model than for the unclustered national MLR. The average RMSE for provinces grouped in the scale cluster 1-5 was 0.18884, or 62% lower than that of provinces not clustered on the national scale, ranging from 0.30424 to 0.1154. Further analysis of each cluster demonstrated that the provinces in cluster 3 had the lowest MAE and RMSE values compared with those in the other four clusters. However, the provinces in the cluster consistently showed that the EUI prediction model for Indonesian provinces using regression based on clustered temperature and humidity variables performs better than the non-clustered EUI prediction model. A comparison of the MAE and RMSE values is shown in Figure 5 and Table III.

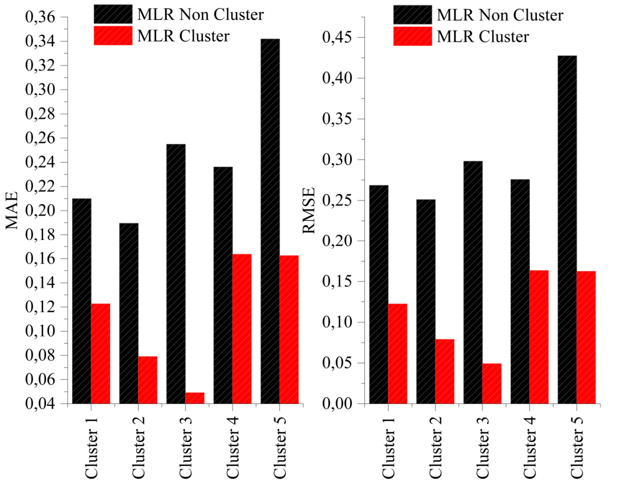


Fig. 5. Comparative predictive performance of cluster-scale MLR and national-scale (non-cluster) MLR.

### F. *Coefficient of Determination*

After obtaining the MAE and RMSE results, an $R^2$ comparison is conducted to further demonstrate that the model can predict values close to the actual ones. The $R^2$ value is obtained based on EUI predictions using temperature and humidity variables. The results of the graphic comparison show that the models using the scale cluster model across clusters 1-5 have higher $R^2$ values than the non-cluster or national-scale model. This indicates that the model with the scale cluster predicts EUI values closer to the actual values compared to the model with the non-cluster or national scale, as shown in Figure 6.

The lower MAE and RMSE values and the higher $R^2$ in the cluster-scale model, compared to the non-cluster or national-scale model, provide strong evidence that the cluster-based MLR approach improves performance. The proposed model in this study achieves greater accuracy than the previous ones.

TABLE III. COMPARISON OF THE MAE AND RMSE VALUES

| No | Cluster | Scale | MAE | RMSE |
|---|---|---|---|---|
| 1 | Cluster 1 | Differentiation | 0.0946 | 0.146 |
| | | % Differentiation | 45% | 54% |
| 2 | Cluster 2 | Differentiation | 0.1185 | 0.172 |
| | | % Differentiation | 63% | 68% |
| 3 | Cluster 3 | Differentiation | 0.2129 | 0.249 |
| | | % Differentiation | 84% | 84% |
| 4 | Cluster 4 | Differentiation | 0.1098 | 0.1122 |
| | | % Differentiation | 47% | 41% |
| 5 | Cluster 5 | Differentiation | 0.2151 | 0.265 |
| | | % Differentiation | 63% | 62% |
| 6 | Average | Differentiation | 0.15018 | 0.18884 |
| | | % Differentiation | 60% | 62% |

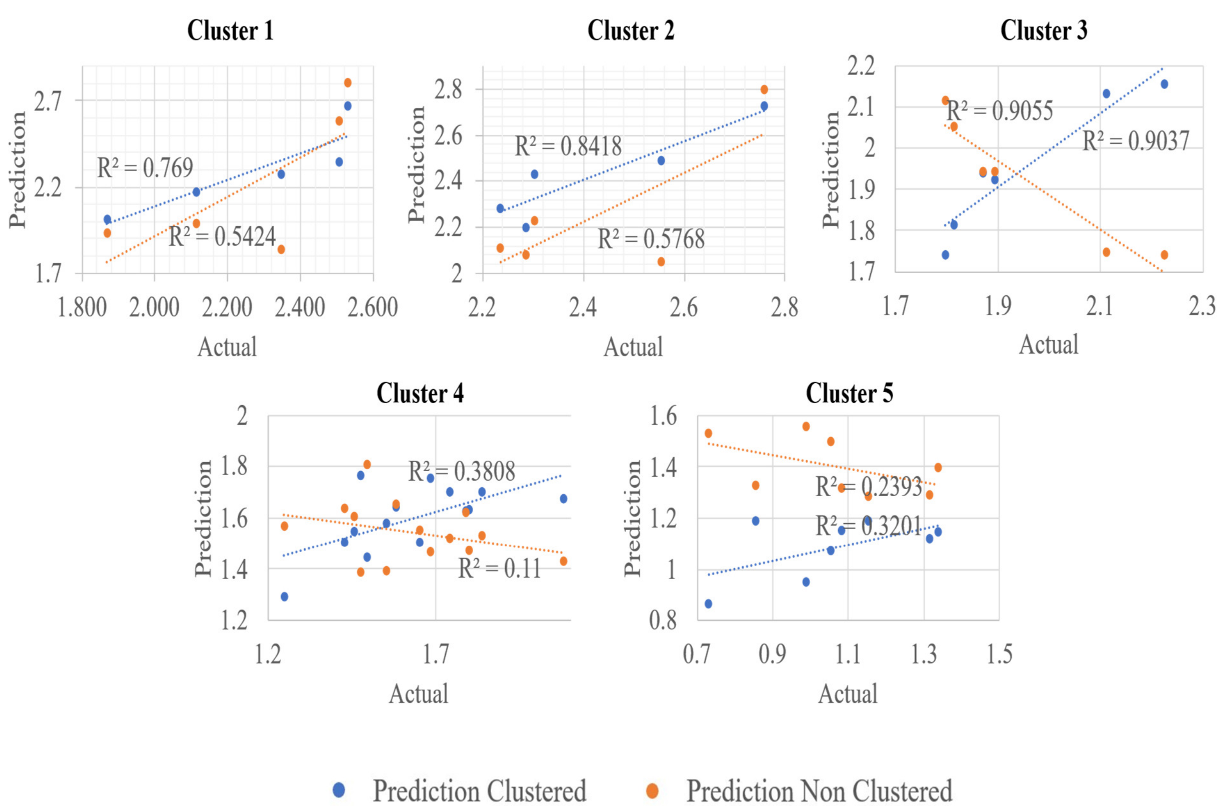


Fig. 6. Comparative coefficient of determination of cluster-scale MLR and national-scale (non-clustered) MLR.

Follow-up based on the results of this study can be developed in many areas. First, the geographic environment. This research can be further developed in countries/regions that encompass several provinces or administrative areas with heterogeneous temperature and humidity variables. Second, the development of variables, especially the independent variables, is important. The current study only uses two independent variables: temperature and humidity, which represent metrology. However, there are many other metrology variables, such as wind speed and solar radiation. In addition to adding or changing metrology variables, it is still possible to include independent variables beyond metrology to predict EUI. Third, the evaluation of EUI and/or space-comfort standards accounts for clustering, as different temperature and humidity levels affect energy consumption and comfort.

## IV. CONCLUSION

The temperature and humidity data across Indonesian provinces indicate heterogeneity. Therefore, the independent variables, temperature and humidity, within Indonesian provinces need to be grouped to form homogeneous clusters before generating regression predictions for Energy Use Intensity (EUI). Cluster analysis based on temperature and humidity indicates that five clusters form homogeneous provinces in Indonesia.

The cluster-scale framework improves the accuracy of conventional national-scale models by separating zones dominated by sensible cooling loads from those dominated by latent cooling loads. From a building physics perspective, high humidity forces the HVAC to cool and remove water vapor. This is because removing water vapor (latent heat) requires more energy. Similarly, rising temperatures will significantly increase the HVAC system's workload. Across all clusters, the cluster-scale regression model consistently outperforms the national-scale model, as evidenced by the lower Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). A higher coefficient of determination ($R^2$) further reinforces this finding. This improvement demonstrates that using scale-cluster modeling under homogeneous climate conditions can reduce prediction bias arising from aggregation in regions with heterogeneous variables. In addition to improving accuracy by reducing bias, this model reduces heteroscedasticity (variation in error rates) in the residuals across different climate contexts.

The resulting clusters can be used not only to improve EUI predictions but also to enhance cluster-based policy relevance, such as benchmarking EUI values by cluster. This allows each province to adopt different EUI standards aligned with its cluster group, rather than relying solely on a single national standard. Furthermore, the resulting clusters can be used to formulate strategic policies related to energy efficiency. The degree of membership in Fuzzy C-Means (FCM) clustering and policy design allows for more flexible policymaking interventions for provinces located on the border between two climate zones.

This study is still limited to using climate variables to predict EUI in office buildings. There are many other variables, such as differences in building typology, operating schedules, occupancy density, and building envelope characteristics, that can cause substantial differences in EUI even within the same climate group. This limitation suggests that future research could expand this framework by incorporating additional operational factors, such as occupancy, HVAC control settings, and building envelope characteristics, as well as climate or urban variables, and by examining finer temporal and spatial scales, or by integrating climate change projections to assess the long-term resilience of building energy systems under evolving climate conditions.

## DECLARATION OF COMPETING INTERESTS

The authors declare that there are no financial or personal conflicts of interest that could have influenced this paper.

## ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support for the article processing charge (APC) provided by the Unit of Research, Innovation, and Community Engagement, Faculty of Engineering, Universitas Indonesia under contract no NKB-1130/UN2.F4.D/PPM.00.04/2026.

## DATA AVAILABILITY

All data is available upon request from the corresponding author.

## REFERENCES

[1] S. Alnatheer and M. A. Ahmed, "Enhanced Energy Prediction of Next-Generation Urban Buildings Optimized with Pyramidal Dilation Attention Convolutional Deep Neural Networks," *Engineering, Technology & Applied Science Research*, vol. 15, no. 5, pp. 28394–28401, Oct. 2025, https://doi.org/10.48084/etasr.13097.

[2] A. Halim, M. N. Yusoff, and S. S. S. Ali, "Energy Audit and Energy Consumption Analysis of the Campus Building Operation of the State Polytechnic of Samarinda," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 25895–25901, Aug. 2025, https://doi.org/10.48084/etasr.10859.

[3] N.-M. Nguyen and M.-T. Cao, "Energy use intensity analysis of office buildings using green BIM-integrated Interpretable machine learning," *Journal of Building Engineering*, vol. 108, Aug. 2025, Art. no. 112760, https://doi.org/10.1016/j.jobe.2025.112760.

[4] K. Lee, H. Lim, J. Hwang, and D. Lee, "Development of building benchmarking index for improving gross-floor-area-based energy use intensity," *Energy and Buildings*, vol. 328, Feb. 2025, Art. no. 115103, https://doi.org/10.1016/j.enbuild.2024.115103.

[5] *Building Simulation Applications BSA 2024*, Bozen-Bolzano University Press, Bolzano, Italy, 2025.

[6] X. Xie, O. Sahin, Z. Luo, and R. Yao, "Impact of neighborhood-scale climate characteristics on building heating demand and night ventilation cooling potential," *Renewable Energy*, vol. 150, pp. 943–956, May 2020, https://doi.org/10.1016/j.renene.2019.11.148.

[7] R. Abu *et al.*, "Modeling influence of weather variables on energy consumption in an agricultural research institute in Ibadan, Nigeria," *AIMS Energy*, vol. 12, no. 1, pp. 256–270, Nov. 2023, https://doi.org/10.3934/energy.2024012.

[8] H. Z. Anonto *et al.*, "Optimizing Energy Consumption Prediction Using Hybrid LightGBM and XGBoost: Integrating Heterogeneous Data for Smart Grid Management," in *2025 IEEE Region 10 Symposium (TENSYMP)*, Jul. 2025, pp. 1–8, https://doi.org/10.1109/TENSYMP63728.2025.11144991.

[9] Y. Zhao, X. Ding, Z. Wu, S. Yin, Y. Fan, and J. Ge, "Impact of urban form on building energy consumption in different climate zones of

China," *Energy and Buildings*, vol. 320, Oct. 2024, Art. no. 114579, https://doi.org/10.1016/j.enbuild.2024.114579.

[10] S. V. I. R. V. Serasinghe, M. A. Wijewardane, and I. D. Nissanka, "A Study on Climate Change Impact on Cooling Energy Demand Patterns for an Existing Office Building," in *ICSBE 2020*, Singapore, 2022, pp. 115–130, https://doi.org/10.1007/978-981-16-4412-2_10.

[11] S. Chen *et al.*, "Prediction of urban residential energy consumption intensity in China toward 2060 under regional development scenarios," *Sustainable Cities and Society*, vol. 99, Dec. 2023, Art. no. 104924, https://doi.org/10.1016/j.scs.2023.104924.

[12] I. Adilkhanova, J. H. Jeong, and G. Y. Yun, "The role of geographic scale of weather data in urban building energy models," *Sustainable Cities and Society*, vol. 125, May 2025, Art. no. 106339, https://doi.org/10.1016/j.scs.2025.106339.

[13] M. Cygańska and M. Kludacz-Alessandri, "Determinants of Electrical and Thermal Energy Consumption in Hospitals According to Climate Zones in Poland," *Energies*, vol. 14, no. 22, Jan. 2021, Art. no. 7585, https://doi.org/10.3390/en14227585.

[14] Y. Wang, Q. Li, and Y. Xu, "Building Electricity Load Forecasting Considering Climate Change Impacts: A Multi-Factor Deep Learning Approach," in *2025 IEEE Power & Energy Society General Meeting (PESGM)*, Jul. 2025, pp. 1–5, https://doi.org/10.1109/PESGM52009.2025.11225602.

[15] J. Yang, M. He, X. Zhang, Q. Ning, Y. Chen, and M. A. Ziaei Mazinan, "Climate adaptive energy efficiency modeling using a generalized additive approach to optimize building performance across Chinese climate zones," *Scientific Reports*, vol. 15, no. 1, Jun. 2025, Art. no. 20088, https://doi.org/10.1038/s41598-025-04844-1.

[16] K. Run, F. Cévaër, and J.-F. Dubé, "Preliminary Multiple Linear Regression Model to Predict Hourly Electricity Consumption of School Buildings," in *Future Energy: Challenge, Opportunity, and, Sustainability*, X. Wang, Ed. Cham: Springer International Publishing, 2023, pp. 119–127.

[17] S. Song, H. Leng, H. Xu, R. Guo, and Y. Zhao, "Impact of Urban Morphology and Climate on Heating Energy Consumption of Buildings in Severe Cold Regions," *International Journal of Environmental Research and Public Health*, vol. 17, no. 22, Jan. 2020, Art. no. 8354, https://doi.org/10.3390/ijerph17228354.