Journal of Geospatial Science and Analytics

Journal homepage:
https://ejournal.foundae.com/index.php/jgsa/index
e-ISSN: 3090-8086 | p-ISSN: 3090-532X

Multiscale geographically weighted regression with LASSO and Group LASSO: Review and application to micro and small enterprises revenue

Juli Yandi Rahman
Indonesia Defense University, Indonesia

Farell Fillyanno Zevic
University of Edinburgh, United Kingdom

Wawan Hendriawan Nur*
National Research and Innovation Agency (BRIN), Indonesia

Irsyad Ramli
Bukit Digital Technology, Indonesia

Article Info

Article history:

Received: July 12, 2025
Revised: Oct 22, 2025
Accepted: Nov 11, 2025

Keywords:

Group structured data
Kernel function
Micro Small Enterprises' income
Multiscale Spatial Modelling
Variable selection

DOI:
10.58524/jgsa.v1i3.55

Abstract

Micro and Small Enterprises (MSE) holds a crucial role in the economy because it contributes 55% of the state's income, but MSE still has a lot of deficiencies, so immediate optimization is vital. The purpose of this study is to model and map the MSE income at the regency level in West Java using Multiscale Geographically Weighted Regression (MGWR) with a selection variable process. MGWR is a method that is used to capture a spatial heterogeneity process by allowing effects to vary over space using "borrowed" nearby data that is controlled by various bandwidths for each variable. This research also adds variable selection processes such as LASSO and Group LASSO as an improvement of MGWR to model group-structured data. The response of this study is MSE income in 27 regencies/cities in West Java province, Indonesia, with 144 independent variables that will be selected using LASSO and Group LASSO to become predictor variables in MGWR model. The results of the spatial modelling show that the best model is MGWR with Group LASSO using bi-square kernel function. Based on this result, it can be seen that a group of important variables which significantly affect the MSE income are fertility, energy source, natural disaster, industry, and tourism. Fertility and energy source significantly affect the MSE income in all regencies, but fertility itself has no significant effect in big cities. Then, in the industry and tourism, the number of visiting foreign tourists has the most significant effect.

To cite this article: Rahman, J. Y., Fillyanno, F. F., Zevic, F., Nur, W. H., & Ramli, I. (2025). Multiscale geographically weighted regression with LASSO and Group LASSO: Review and application to micro and small enterprises revenue. *Journal of Geospatial Science and Analytics,* 1(3), 185–202.

***Corresponding Author:**
Wawan Hendriawan Nur, National Research and Innovation Agency (BRIN), INDONESIA,
Email: wawa020@brin.go.id

INTRODUCTION

The economy is one of the vital aspects in a country and it is also included in Indonesia's development plan also known as RPJPN 2025-2045. To maintain and develop economic growth in a country especially Indonesia, Micro and Small Enterprises (MSEs) can be one of the promising solutions. It is because MSE contribute 55% of the Gross Domestic Product and 65% job opportunities even in the high-incomes economic condition (Anshika et al., 2021). However, MSE still has many deficiencies for example a limited financial access and lack of decision-making ability (Esubalew & Raghurama, 2020; Rismawati, 2009; Sinambela et al., 2021; Suminah et al., 2022). The large amount of MSE in Indonesia's villages has made a huge potential for MSE to build and develop a better economic condition in Indonesia.

Many previous studies have conducted the MSE modelling. For example, the MSE modelling using hierarchical multiple regression in Lagos, Nigeria (Olayemi et al., 2022) using a questionnaire for gaining the MSEs' actors' perception about the ease of carrying out business reformation. Then, there is research conducted by (Ayambila, 2023) which uses the Quantile regression method for gaining more comprehensive understanding of relations among the variables. Next, (E. Colipano, 2022) in his research, he used multinomial logistic regression to understand the relations between variables where the dependent variable is categorical.

The previous research that was stated before has not considered the spatial effect and variable selection. In this study, variable selection operators such as Least Absolute Shrinkage Selection Operator (LASSO), Group LASSO and Multiscale Geographically Weighted Regression (MGWR) with various kernel functions are used for modelling the relation between important predictors that have been selected with LASSO or group LASSO and the response with different spatial scales for each relation. In this study, 143 predictors with various groups such as geography, economy, demography and civilization, education, health, disease, fertility, energy source, criminality, disaster, staple food production, vegetables and nuts production, fruit production, raw materials and spices, meat production, eggs and milks production, industry and tourism, infrastructure and GDRP are used to modelling the MSE's income in West Java with 27 observation per variable. The main objective of this study is to identify the important variables or group variables that affect the MSE's income in West Java, and also modelling and mapping the significancy of the spatial effect of those important variables from the best-selected model that was selected based on the goodness of fit criteria.

This article consists of four sections. Section 2 explains a brief review of the method that used in this study, and each sub-section introduces LASSO, group LASSO and MGWR as well as goodness of fit criteria. The results and discussions of this study are explained in Section 3, consists of the results of K-fold cross-validation, important variables, outliers, classic assumption and optimal bandwidth for each model also the best model determination and the discussion. Finally, the conclusions and suggestions derived from the discussion are given in Section 4.

METHOD

This study uses a lot of data on various aspects. It makes the data have various scales, and transformation is a must to prevent the heteroscedasticity in the model residuals. In this study, we used the scale() function in R, which is a normal transformation that has a formula such as

$$x_{k_{new}} = \frac{x_k - \overline{x_k}}{\sigma_k^2} \quad (1)$$

Where x_k is the data on the k-variable, $x_{k_{new}}$ is the new data after transformation, $\overline{x_k}$ is the mean of the variable, and σ_k^2 is the standard deviation of the variable (Kappal, 2019). The following sub-sections in this section cover approaches such as LASSO regression, group LASSO regression, and also Multiscale Geographically Weighted Regression.

Least Absolute Shrinkage Selection Operator (LASSO) Regression

LASSO approaches firstly introduced by (Pillay & Lin, 2023; Tibshirani, 1996) which has been preferred by many researchers because this approach can easily define the important predictors and exclude the unnecessary variables by reducing the coefficients to zero, even in conditions where the number of variables is larger than the number of observations (Schneider & Tardivel, 2020). The LASSO regression employs the L1 regularization and has a formula such as

$$\underset{\beta \in \mathbb{R}^k}{\text{minimize}} \left\{ \frac{1}{2N} \sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{k} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{k} |\beta_j| \right\} \quad (2)$$

where the λ is the control shrinkage parameter for LASSO to reduce the parameter coefficients. For a single predictor, we can estimate the β for solving the problem in equation (2) by

$$\hat{\beta} = \begin{cases} \frac{1}{N}\langle \mathbf{z}, \mathbf{y} \rangle - \lambda \; if \; \frac{1}{N}\langle \mathbf{z}, \mathbf{y} \rangle > \lambda \\ 0 \; if \; \frac{1}{N}|\langle \mathbf{z}, \mathbf{y} \rangle| \leq \lambda \\ \frac{1}{N}\langle \mathbf{z}, \mathbf{y} \rangle + \lambda \; if \; \frac{1}{N}\langle \mathbf{z}, \mathbf{y} \rangle < -\lambda \end{cases} \quad (3)$$

where $\tilde{\beta} = \frac{1}{N}\langle \mathbf{z}, \mathbf{y} \rangle$ is the ordinary least square estimator with standardized data. We can rewrite the estimator in equation (3) succinctly as

$$\hat{\beta} = S_\lambda \left(\frac{1}{N}\langle \mathbf{z}, \mathbf{y} \rangle \right) \quad (4)$$

By this logic, the coordinate-wise scheme for solve the full lasso problem for the objective below that can be written as

$$\frac{1}{2N} \sum_{i=1}^{N} \left(y_i - \sum_{l \neq j} x_{il}\beta_l - x_{ij}\beta_j \right)^2 + \lambda \sum_{l \neq j} |\beta_l| + \lambda |\beta_j| \quad (5)$$

Based on this problem, we can estimate the parameter by S_λ that can be reconstructed as

$$\hat{\beta}_j = S_\lambda \left(\frac{1}{N}\langle \mathbf{x}_j, \mathbf{r}^{(j)} \rangle \right) \quad (6)$$

$$r_i^{(j)} = y_i - \sum_{l \neq j} x_{il}\hat{\beta}_l \quad (7)$$

where r denotes the partial residual of the model.

Group LASSO

Regression problems usually involve variables with group structure and substantial correlation between variables within the same group. This issue prevents standard LASSO from operating efficiently. Group LASSO in (Hastie et al., 2015; Huang et al., 2024), which focuses on the variable's group rather than its individual values, can resolve this issue. A group's coefficients can all be concurrently reduced to zero (or nonzero) by group LASSO. Take a look at a linear regression model with G groups of variables. For $g = 1, \dots, G$, the covariates in group g are represented by the vector, $Z_g \in \mathbb{R}^{k_g}$. Group LASSO approach can solve a convex problem in equation (8) such as

$$\underset{(\theta_1, \dots, \theta_G)}{\text{minimize}} \left\{ \frac{1}{2} \sum_{i=1}^{N} \left(\mathbf{y} - \sum_{j=1}^{G} \mathbf{Z}_j \theta_j \right)^2 + \lambda \sum_{j=1}^{G} \left\| \theta_j \right\|_2 \right\} \quad (8)$$

Since we can center all the variables and the answer in practice, we can ignore the intercept θ_0. The zero subgradient equations for this issue have the following form:

$$-\boldsymbol{Z}_g^T\left(\boldsymbol{y}-\sum_{l=1}^{G}\boldsymbol{Z}_l\hat{\theta}_l\right)+\lambda\hat{s}_g=0; for\ g=1,\dots,G, \tag{9}$$

when $\hat{\theta}_g \neq 0$, then we can necessarily have $\hat{s}_g = \frac{\hat{\theta}_g}{||\hat{\theta}_g||_2}$, otherwise $\hat{\theta}_g = 0$, then $\hat{s}_g$ is any vector with $||\hat{s}_g||_2 \leq 1$. A technique to solve the zero sub-gradient equations is to fix all of the block vectors $\{\hat{\theta}_h, h \neq g\}$, and then solve the $\hat{\theta}_g$. The issue will inevitably converge to an ideal solution because it is convex and the penalty is block separable. With all $\{\hat{\theta}_h, h \neq g\}$ fixed, we write

$$-\boldsymbol{Z}_j^T(\boldsymbol{r}_j-\boldsymbol{Z}_j\hat{\theta}_j)+\lambda\hat{s}_j=0 \tag{10}$$

where $\boldsymbol{r}_g = \boldsymbol{y} - \sum_{h\neq g}\boldsymbol{Z}_h\hat{\theta}_h$ is the g^{th} partial residual. Considering the conditions that the sub-gradient $\hat{s}_g$ satisfies, we must have $\hat{\theta}_g = 0$ if $||\boldsymbol{Z}_g^T\boldsymbol{r}_g||_2 < \lambda$, and otherwise the minimizer $\hat{\theta}_g$ must satisfy

$$\hat{\theta}_j=\left(\boldsymbol{Z}_j^T\boldsymbol{Z}_j+\frac{\lambda}{||\hat{\theta}_j||_2}\boldsymbol{I}\right)^{-1}\boldsymbol{Z}_j^T\boldsymbol{r}_j \tag{11}$$

This update is comparable to how a ridge regression problem is solved, with the exception that the underlying penalty parameter is dependent on $||\hat{\theta}_g||_2$. Regretfully, there isn't a closed-form solution for $\hat{\theta}_g$ in equation (11) unless $\boldsymbol{Z}_g$ is orthonormal. In this particular case, the straightforward update is

$$\hat{\theta}_j=\left(1+\frac{\lambda}{||\boldsymbol{Z}_j^T\boldsymbol{r}_j||_2}\boldsymbol{I}\right)_+\boldsymbol{Z}_j^T\boldsymbol{r}_j \tag{12}$$

where $(t)_+ := \max\{0, t\}$ is the positive partial function.

Multiscale Geographically Weighted Regression

Geographically Weighted Regression (GWR) is an Ordinary Least Square (OLS) that was firstly introduced by (Fotheringham et al., 2017, 2023) where the regression process is performed spatially uniquely in every location, it is done by calibrating the model separately in every area with borrowing data from the nearest location, and weighting the data based on the distance of the regression point, so a bigger weight will be given to the closer location. The decrement rate of the weight is controlled by a bandwidth that is optimized by GWR calibration.

The GWR is formulated by

$$y_i=\beta_{i0}+\sum_{j=1}^{k}\beta_{ij}x_{ij}+\varepsilon_i,\ \ i=1,\dots,n \tag{13}$$

where y is the response, β_{i0} is the intercept, x_{ik} is the k-predictor, β_{ik} is the k-local coefficients for k-predictor, and ε_i is the residual in i location. In the matrix, the GWR estimator for local parameter estimation of i location is:

$$\hat{\beta}_i=[\boldsymbol{X^T}\boldsymbol{W}_i\boldsymbol{X}]^{-1}\boldsymbol{X^T}\boldsymbol{W}_i\boldsymbol{y} \tag{14}$$

where $\boldsymbol{X}$ is the $n \times k$ predictor matrix, $\boldsymbol{W}_i = diag[w_{i1}, \dots, w_{in}]$ is the $n \times n$ diagonal weight matrix which weights each observation based on the distance from i location that is measured using the kernel function and a certain bandwidth, $\hat{\boldsymbol{\beta}}$ is the $k \times 1$ parameter vector, and $\boldsymbol{y}$ is the $n \times 1$ response's observation vector.

By the definition above, the predicted values of each observation can be formulated by

$$\hat{y}_i=\boldsymbol{X}_i[\boldsymbol{X^T}\boldsymbol{W}_i\boldsymbol{X}]^{-1}\boldsymbol{X^T}\boldsymbol{W}_i\boldsymbol{y}=\boldsymbol{Sy} \tag{15}$$

$\boldsymbol{S}$ is the hat matrix, also called the projection matrix, because this matrix can estimate the $\hat{y}$ values by multiplying-them by the $\boldsymbol{y}$. Next, we can also calculate the covariance matrix of the estimated parameter $\boldsymbol{V}_i$ that could be constructed in a similar way as

$$\boldsymbol{C}_i=[\boldsymbol{X^T}\boldsymbol{W}_i\boldsymbol{X}]^{-1}\boldsymbol{X^T}\boldsymbol{W}_i \tag{16}$$

$$\boldsymbol{V}_i = \boldsymbol{C}_i\boldsymbol{C}_i^T\widehat{\boldsymbol{\sigma}}^2 \tag{17}$$

where $\hat{\sigma}^2$ is the estimated standard deviation of the residual model pattern, which formulated by

$$\widehat{\boldsymbol{\sigma}}^2 = \frac{\sum_i(y_i - \hat{y}_i)^2}{\boldsymbol{m} - tr(\boldsymbol{S})} \tag{18}$$

with m is the number of observations in the sample. Then, we can calculate the standard error of the estimated parameter using equation (19)

$$se(\widehat{\boldsymbol{\beta}}_i) = \sqrt{diag(\boldsymbol{V}_i)} \tag{19}$$

Then, five different kernel functions that will be used are shown in Table 1 below (Al-Hasani et al., 2021; De Carvalho et al., 2017; Fan et al., 2018; Nugroho & Slamet, 2018; Zhong et al., 2013)

Table 1. Kernel Function

Function	Specification	
Bi-square	$w_i = \begin{cases} \left(1 - \left(\frac{\lvert d_i\rvert}{b}\right)^2\right)^2, & \text{if } \lvert d_i\rvert < b \\ 0, & \text{others} \end{cases}$	(20)
Gaussian	$w_i = \exp\left(-\frac{1}{2}\left(\frac{\lvert d_i\rvert}{b}\right)^2\right)$	(21)
Exponential	$w_i = \exp\left(-\left(\frac{\lvert d_i\rvert}{b}\right)\right)$	(22)
Tricube	$w_i = \begin{cases} \left(1 - \left(\frac{\lvert d_i\rvert}{b}\right)^3\right)^3, & \text{if } \lvert d_i\rvert < b \\ 0, & \text{others} \end{cases}$	(23)
Boxcar	$w_i = \begin{cases} 1, & \text{if } d_i < b \\ 0, & \text{others} \end{cases}$	(24)

which bandwidth b is a determined parameter from cross-validation (CV) process that shows the neighbourhood of each location (Fotheringham et al., 2022). For simplicity, we can write the GWR algorithm as:

Algorithm 1

Geographically Weighted Regression Estimation Process

1. Initialize empty $\widehat{\boldsymbol{\beta}}(m \times k), \widehat{\boldsymbol{y}}(m \times 1), \boldsymbol{S}(m \times m), \boldsymbol{CCT}(m \times k), \boldsymbol{SE}(m \times k)$
2. Given a data-borrowing scheme, compute $\boldsymbol{W}$
3. For location 1 ... m, at each location i, calculate:
4. The diagonal matrix $\boldsymbol{W}_i$ based on a row of the spatial weight matrix $\boldsymbol{W}$
5. Parameter estimates $\widehat{\boldsymbol{\beta}}_i$ from equation (14); store in $\widehat{\boldsymbol{\beta}}$
6. Predictive value $\hat{y}_i$ from equation (15); store in $\widehat{\boldsymbol{y}}$
7. Row of hat matrix $\boldsymbol{S}_i$ from equation (15); store in $\boldsymbol{S}$
8. Diagonal of intermediate values $\boldsymbol{C}_i\boldsymbol{C}_i^T$ from equation (16); store in $\boldsymbol{CCT}$
9. End for
10. Calculate $\hat{\sigma}^2$ from equation (18)
11. Calculate standard errors $se(\widehat{\boldsymbol{\beta}}_i)$ in equation (19); store $\boldsymbol{SE}$
12. End for
13. End

The conventional GWR limitation is an assumption that every variable relation in the model has the exact same bandwidth of weights in the same spatial scale. MGWR covered this limitation, by using different bandwidths for each variable relation in the model. MGWR advantages are that it could explain spatial heterogeneity more accurately and reduce the bias of parameter estimation (Fotheringham et al., 2017). The definition of MGWR model is

$$y_i = bw_0(\beta_{0i}) + \sum_{j=1}^{k} bw_j(\beta_{ij}x_{ij}) + \varepsilon_i \tag{25}$$

where bw is every bandwidth that used for each model component. MGWR model could be calibrated by rearranging the formula as Generalized Additive Model (GAM) (Comber et al., 2024). The GAM model of MGWR has a formula that can be written as

$$\boldsymbol{y} = \boldsymbol{f}_{\boldsymbol{bw_0}} + \sum_{\boldsymbol{j=1}}^{\boldsymbol{k}} \boldsymbol{f}_{\boldsymbol{bw_j}} \boldsymbol{X_j} + \boldsymbol{\varepsilon} \tag{26}$$

where $\boldsymbol{f}_{\boldsymbol{bw_k}}(\boldsymbol{X_k}) = bw_k(\beta_k)\boldsymbol{X_k}$ is a smoothing function or data-borrowing scheme that includes kernel function selection, weight and bandwidth calculation. To estimate every $\boldsymbol{f}_{\boldsymbol{bw_k}}$ and calculate the β_k, equation (26) could be reconstructed and gives component condition expectation as follows

$$\boldsymbol{f}_{\boldsymbol{bw_k}} = E\left(\boldsymbol{y} - \textstyle\sum_{\boldsymbol{p \neq k}} \boldsymbol{f}_{\boldsymbol{bw_p}} - \boldsymbol{\varepsilon} \middle| \boldsymbol{X_k}\right) = \boldsymbol{A_k}\left(\boldsymbol{y} - \sum_{\boldsymbol{p \neq k}} \boldsymbol{f}_{\boldsymbol{bw_p}}\right) \tag{27}$$

where $\boldsymbol{A_k}$ is $\boldsymbol{E}(.|\boldsymbol{X_k})$ which is a hat matrix of univariate GWR model of $\boldsymbol{X_k}$ which mapping $\boldsymbol{y} - \sum_{\boldsymbol{p \neq k}} \boldsymbol{f}_{\boldsymbol{bw_p}}$ to $\hat{\boldsymbol{f}}_{\boldsymbol{bw_k}}$. GWR hat matrix $\boldsymbol{A_k}$ defined as

$$\boldsymbol{A_k} = \begin{pmatrix} x_{1k}\left(\boldsymbol{X_k^T W_1 X_k}\right)^{-1} \boldsymbol{X_k^T W_1} \\ \dots \\ x_{nk}\left(\boldsymbol{X_k^T W_n X_k}\right)^{-1} \boldsymbol{X_k^T W_n} \end{pmatrix}_{n \times n} \tag{28}$$

where $\boldsymbol{W_i}$ is a diagonal spatial weight matrix which calculated based on a basis of a bandwidth with specific covariate and a kernel function.

After initialize the initial β for each predictor, we define a residual of GAM model of MGWR for the back-fitting algorithm as follows

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \sum_{\boldsymbol{j=1}}^{\boldsymbol{k}} \hat{\boldsymbol{f}}_{\boldsymbol{j}} \tag{29}$$

Consider $\boldsymbol{A_k}$ is a hat matrix of the partial optimal model before, so

$$\hat{\boldsymbol{f}}_{\boldsymbol{k}}^{*} = \boldsymbol{A_k}(\hat{\boldsymbol{f}}_{\boldsymbol{k}} + \hat{\boldsymbol{\varepsilon}}) \tag{30}$$

where $\hat{\boldsymbol{f}}_{\boldsymbol{k}}^{*}$ is the updated $\hat{\boldsymbol{f}}_{\boldsymbol{k}}$ which is an additive term of the prediction result of the previous iteration.

At the final iteration, $\hat{y}$ predicted value and $\hat{\varepsilon}$ residual will be used to see the difference between the current and the previous values to guarantee that the back-fitting algorithm reaches convergence. We can use the Score of Change (SOC) value to check the convergence based on the Residual Sum of Square (RSS) of the MGWR model.

$$SOC_{RSS} = \frac{RSS_{new} - RSS_{old}}{RSS_{new}} \tag{31}$$

The convergence is reached when SOC has a smaller value than an η limitation which is commonly set as 10^{-5}. It is possible to compute a covariate-specific hat matrix $\boldsymbol{R}_k$ that maps the response $\boldsymbol{y}$ to each of the estimated model components $\hat{\boldsymbol{f}}_k$ in the back-fitting algorithm such that

$$\hat{\boldsymbol{f}}_k = \boldsymbol{R}_k \boldsymbol{y} \tag{32}$$

where $\boldsymbol{R}_k$ is calculated as

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{A}_1 & \dots & \boldsymbol{A}_1 \\ \boldsymbol{A}_2 & \boldsymbol{I} & \dots & \boldsymbol{A}_2 \\ \dots & \dots & \dots & \dots \\ \boldsymbol{A}_k & \boldsymbol{A}_k & \dots & \boldsymbol{I} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \\ \dots \\ \boldsymbol{A}_k \end{bmatrix} \tag{33}$$

After convergence is reached, the final values of $\boldsymbol{R}_k$ can be summed to obtain the overall hat matrix $\boldsymbol{S}$ for the model as follows

$$\boldsymbol{S} = \sum_{k} \boldsymbol{R}_k \tag{34}$$

In addition, we can use the $\boldsymbol{R}_k$ to compute the covariance-specific effective number of parameters (ENP) as follows

$$ENP_k = tr(\boldsymbol{R}_k) \tag{35}$$

and ENP of the model can be calculated by

$$ENP_{model} = \sum_k ENP_k \tag{36}$$

For simplicity, we can write the MGWR algorithm as can be seen in Algorithm 2.

Algorithm 2

Multiscale Geographically Weighted Regression Estimation Process

GWR model calibration to obtain the initial values for each $\hat{\boldsymbol{f}}_k, \hat{\boldsymbol{\varepsilon}}$ and $\boldsymbol{R}_k$
Initialize $SOC \gg \eta$
Do until $SOC < \eta$:
For each term k:
Calibrate univariate GWR model $(\hat{\boldsymbol{f}}_{\boldsymbol{k}} + \hat{\boldsymbol{\varepsilon}}) \sim \boldsymbol{X}_{\boldsymbol{k}}$ to obtain new $\hat{\boldsymbol{f}}_{\boldsymbol{k}}^*$ and $\hat{\boldsymbol{\varepsilon}}^*$
Update $\hat{\boldsymbol{f}}_k \leftarrow \hat{\boldsymbol{f}}_k^*$ and $\hat{\boldsymbol{\varepsilon}} \leftarrow \hat{\boldsymbol{\varepsilon}}^*$
Calculate the $\boldsymbol{R}_k$
End for
Calculate new SOC^* and update $SOC \leftarrow SOC^*$
End do
For each term k:
Calculate ENP_k
End for
Calculate $\boldsymbol{S}$

Moran's Index

In spatial modelling, Moran's I index is an important aspect that has to be concerned. Moran's I index was developed by Patrick Alfred Pierce Moran that used to compute the spatial dependency between observation and location. It has a value range between -1 to 1 that shown the spatial term characteristic such as clustered, random, or dispersed (Higazi et al., 2013). The formula for calculate the Moran's I index is written as follows

$$I = \left(\frac{N}{W}\right) * \frac{\Sigma_{i=1}^{N} \Sigma_{j=1}^{N} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\Sigma (x_i - \bar{x})^2} \tag{37}$$

Goodness of Fit

The goodness of fit for each model can be seen from how well the model explains the data. In this study, we use R^2 and AIC value to determine the best model. The R^2 value has a formula for computing as follows (Sharif & Kamal, 2018)

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \tag{38}$$

Then, the AIC value was introduced by Hirotugu Akaike on 1973 as an extension of Maximum Likelihood (Cavanaugh & Neath, 2019; El-Habil, 2012). The formula of Maximum Likelihood is written as follows

$$\mathcal{L}(\mu, \sigma | y_1, \dots, y_n) = f(y_1, \dots, y_n | \mu, \sigma) = \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n e^{-\Sigma_{i=1}^{n} \frac{(y_i - \mu)^2}{2\sigma^2}} \tag{39}$$

The important idea of AIC is to combine the estimation process with structural and dimension determination. The AIC formula is written as follows

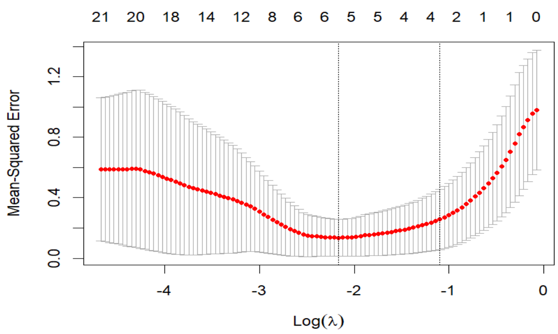
$$AIC = -2 \ln\left(\mathcal{L}(\hat{\theta}_{MLE} | y)\right) + 2k \tag{40}$$

RESULTS AND DISCUSSIONS

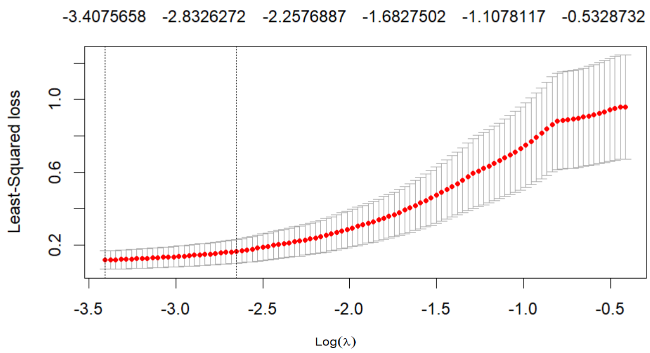
In this section, we present the results of the modelling and analysis process of MGWR model. We used data with various aspects including Geography, Economy, Demography, Education, Health, Disease, Fertility, Energy, Criminality, Disaster, Regional production, Infrastructure, GDRP, Industry and Tourism. The dataset has a total of 144 variables with MSE income as the response that has measures in 27 regencies in West Java province. In the following sections, we estimate the optimal selection parameter λ, predict the important variables and group variables, check the outliers and classic assumptions, determine the best model and mapp the significance of each important variable or group variable. The models that will be used are MGWR with LASSO and MGWR with Group LASSO. In the modelling section, we also use five different kernels for each model to obtain the best model of MSE income in West Java.

Important Variable Selection

The Mean-Squared Error (MSE) of LASSO and Group LASSO cross-validation process is shown in Figure 1. In LASSO CV process, 5 or 10 folds are commonly used in the process (Oyedele, 2023). In this study, we use 5 folds in the CV process. The results of the CV process can be seen in Figure 1 below. The minimum values of log lambda of LASSO and Group LASSO CV process, which have the smallest MSE values (Takano & Miyashiro, 2020) successively are 0,1147 and 0,0623.



(a)



(b)

Figure 1. MSE plot of CV LASSO (a) and CV Group LASSO (b)

As displayed in Table 2 and 3, the optimal LASSO and Group LASSO model with important variables are obtained after employing the best lambda that has the smallest MSE in the CV process. The important variables in LASSO model are Pulmonary tuberculosis cases (X_{62}), Flood cases (X_{103}), Beef production (X_{151}), Number of MSE's workers (X_{172}) and the Number of MSEs' expenditure (X_{173}) where only these variables that significantly affect the MSE's income. Meanwhile in Group LASSO, the important group of variables that were selected in the model are Fertility (G_7), Energy source (G_8), Disaster (G_{10}) also Industry and Tourism (G_{17}).

Table 2. Significant coefficient variables for LASSO

Variables	Coefficient
$\boldsymbol{X_{62}}$: Pulmonary tuberculosis cases	0,0721
$\boldsymbol{X_{103}}$: Flood cases	0,1030
$\boldsymbol{X_{151}}$: Beef production (kg)	0,0371
$\boldsymbol{X_{172}}$: Number of MSE's workers	0,1238
$\boldsymbol{X_{173}}$: Number of MSE's expenditure (thousand rupiah)	0,6442

Table 3. Significant coefficient variables for Group LASSO

Variables	Coefficient	Variables	Coefficient
X_{71}: Number of childbearing age couples	0,0797	X_{104}: Forest fires cases	0,0059
X_{72}: Number of family planning participants	0,0854	X_{105}: Tornado cases	0,0130
X_{81}: Percentage of PLN electricity users	0,0119	X_{171}: Number of MSE	0,0904
X_{82}: Percentage of gas users	0,0171	X_{172}: Number of MSE's workers	0,1787
X_{83}: Percentage of wood users	-0,0155	X_{173}: Number of MSE's expenditure (thousand rupiah)	0,2280
X_{84}: Percentage of access to adequate drinking water	0,0076	X_{174}: Number of Restaurants	0,0577
X_{101}: Earthquake cases	-0,0022	X_{175}: Number of staying local tourists	0,0384
X_{102}: Landslide cases	0,0126	X_{176}: Number of staying foreign tourists	0,0386
X_{103}: Flood cases	0,0144	X_{177}: Number of visiting local tourists	-0,0106
		X_{178}: Number of visiting foreign tourists	-0,0150

Classic Assumption

Before employed to the spatial modelling, the selection variable process results have to be checked on the classic assumption. The classic assumption test has a crucial role in statistical modelling to guarantee the result statistical validation (Ainiyah et al., 2016). It involves a normality test e.g. Kolmogorov-Smirnov normality test (Jurečková & Picek, 2007) and heterogeneity test e.g. Breusch-Pagan test (Halunga et al., 2017). This assumption can reduce the bias of the model and validate the results. As displayed in Table 4, model LASSO and Group LASSO have a normal distribution because they both have p-value > 0,05, so the H0 that said the model has a normal-like distribution can be accepted. Then, we can also see that the LASSO model has no constant error because the model has p-value < 0,05, which means a heteroscedasticity problem occurs in the model, meanwhile the Group LASSO model has constant error because it has p-value > 0,05. In addition, we can see that LASSO model has Moran's I index of 0,0148, which means the data characteristics on each location are clustered. Meanwhile, in Group LASSO model has a Moran's I index of -0,0846 which means the data characteristics on each location are dispersed.

Table 4. Classic assumption and Moran I test for LASSO and Group LASSO model

Tests	p-value	
	LASSO	**Group LASSO**
Kolmogorov-Smirnov normality test	0,4623	0,1154
Breusch-Pagan Heteroscedasticity test	0,0483	0,8425
	Moran's I index	
Moran's I test	0,0148	-0,0846

Another part of the classic assumption is the multicollinearity test, which explains about the collinearity level among the variables. The collinearity level on the multicollinearity test is measured by the Variance Inflation Factor (VIF), where variables that have more than 10 of VIF means those variables have a multicollinearity problem (Daoud, 2017). Table 5 and Table 6 are show the VIF of each important variables in LASSO and Group LASSO model which we can see all the variables in LASSO model has less than 10 VIF values, meanwhile in Group LASSO, still there are variables that has more than 10 VIF values because the Group LASSO approach reduce the residual sum of square error among the groups which makes it possible still there is high correlation between variables in the same group (Yunus et al., 2020).

Table 5. Multicollinearity test for LASSO model

Variables	VIF
X_{62}: Pulmonary tuberculosis cases	3,1623
X_{103}: Flood cases	2,3523
X_{151}: Beef production (kg)	2,4656
X_{172}: Number of MSE's workers	2,9514
X_{173}: Number of MSE's expenditure (thousand rupiah)	2,5907

Table 6. Multicollinearity test for Group LASSO model

Variables	VIF	Variables	VIF
X_{71}: Number of childbearing age couples	342,6608	X_{104}: Forest fires cases	3,7647
X_{72}: Number of family planning participants	411,0664	X_{105}: Tornado cases	19,1444
X_{81}: Percentage of PLN electricity users	6,3489	X_{171}: Number of MSE	46,6817
X_{82}: Percentage of gas users	11,4102	X_{172}: Number of MSE's workers	53,4117
X_{83}: Percentage of wood users	36,4436	X_{173}: Number of MSE's expenditure (thousand rupiah)	12,2694
X_{84}: Percentage of access to adequate drinking water	8,6875	X_{174}: Number of Restaurants	1,7907
X_{101}: Earthquake cases	9,0078	X_{175}: Number of staying local tourists	34,6214
X_{102}: Landslide cases	12,1726	X_{176}: Number of staying foreign tourists	30,8445
X_{103}: Flood cases	28,1466	X_{177}: Number of visiting local tourists	7,2476
		X_{178}: Number of visiting foreign tourists	3,0420

Model Comparison

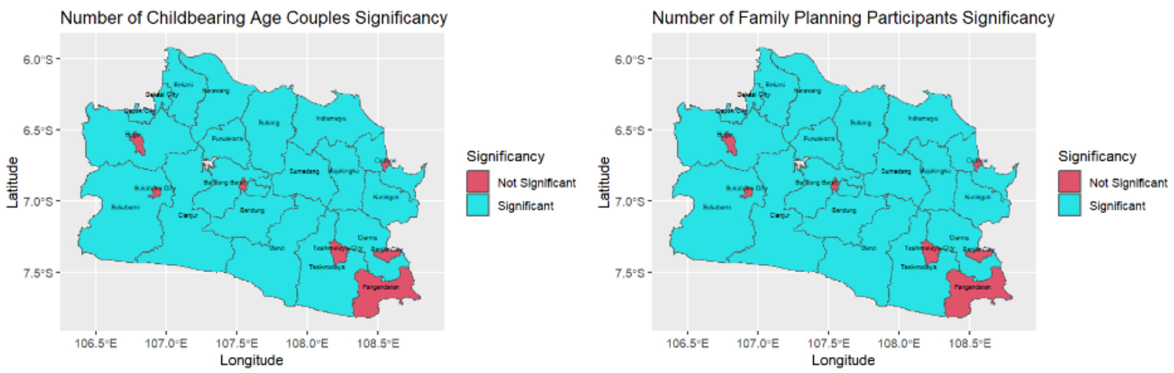
The accuracy parameters of each model are shown in Table 7. We can see that the best model is based on the model with the highest R^2 and the smallest AIC value is MGWR with Group LASSO in Bi-square kernel function. MGWR model has the highest R^2 value of 0,9999 or 99,99% which means only 0,01% of error that exists in the model. This model also has AIC value of 252,4375, which is the smallest AIC value among the other models, and the smaller AIC value of the model, the more accurate the model will be (Cavanaugh & Neath, 2019).

Table 7. Model Comparison

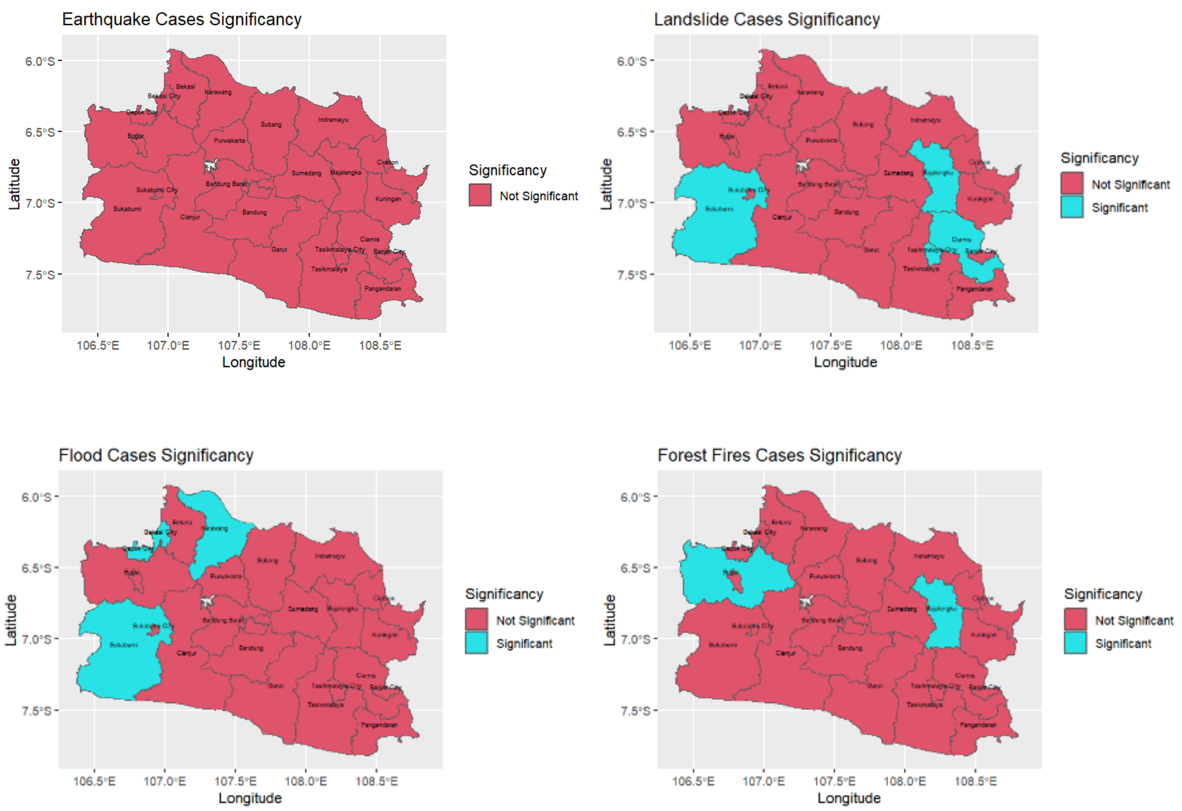
Model	MGWR LASSO		MGWR Group LASSO	
	R Squared	AIC	R Squared	AIC
Bi-square	0,9935	376,0386	0,9999	252,4375
Gaussian	0,9875	387,9883	0,9979	351,4153
Exponential	0,9899	384,4700	0,9985	343,6644
Tricube	0,9931	377,9702	0,9999	263,0751
Boxcar	0,9901	383,4834	0,9975	320,4471

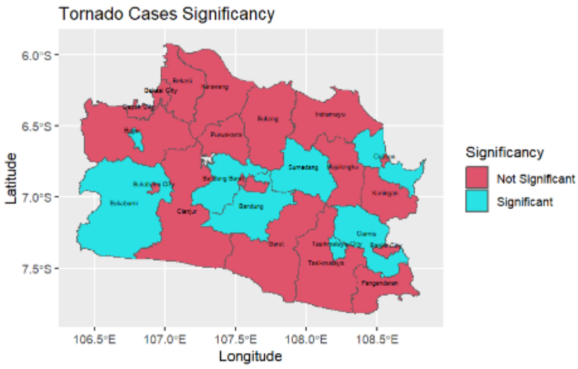
Best Model Mapping Result

The result of spatial modelling using MGWR with Group LASSO in Bi-square model can be seen in Table B.4. The table shows that few variables have positive effect on MSE for example, the Number of childbearing age couples (X_{71}) which means these variables can increase the MSE income in West Java. Meanwhile, there are also few variables which have negative effect on MSE income for example Flood cases (X_{103}) which means these variables can reduce MSE income in West Java. From Figure 2 to Figure 5, we can see that on average, only the Number of childbearing age couples (X_{71}), Number of family planning participants (X_{72}), Percentage of PLN electricity users (X_{81}), Percentage of gas users (X_{82}), Percentage of access to adequate drinking water (X_{84}), Number of MSE (X_{171}), Number of MSE workers (X_{172}) and Number of MSE's expenditure (X_{173}) which significantly affect the MSE income.

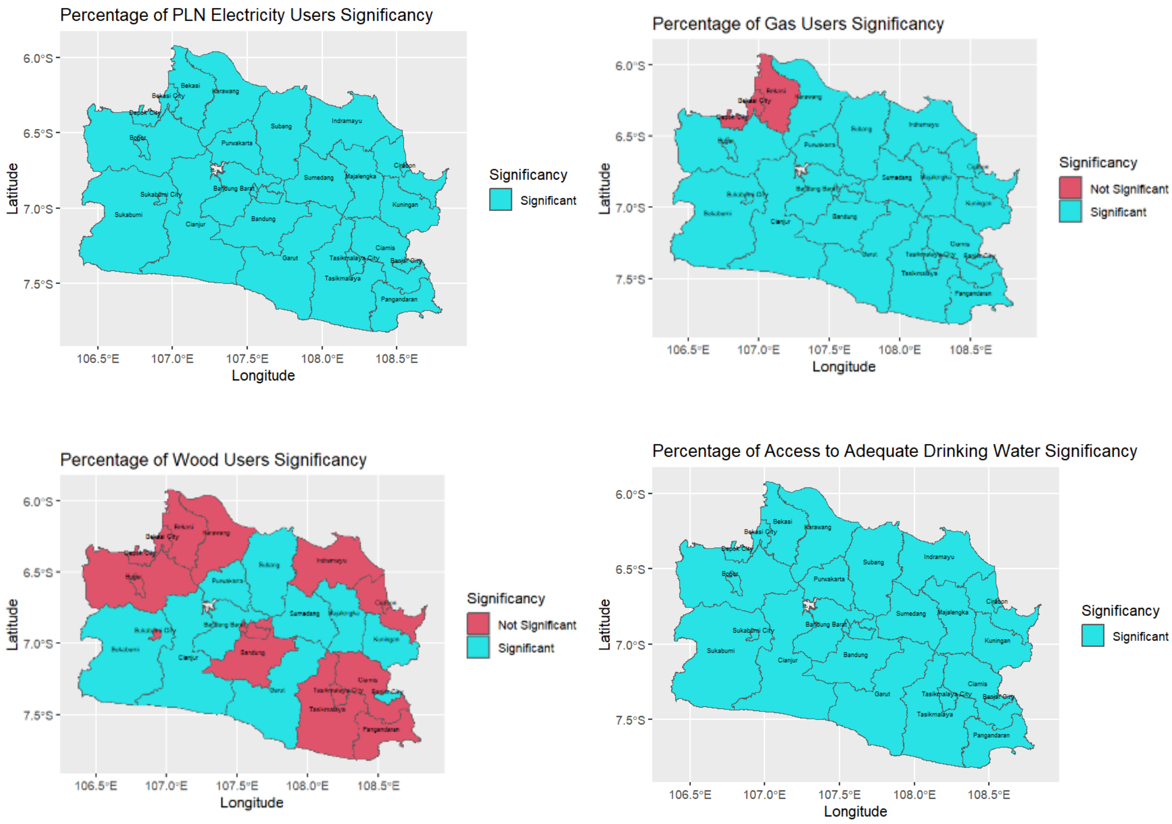


Figures 2. Fertility group variables significancy

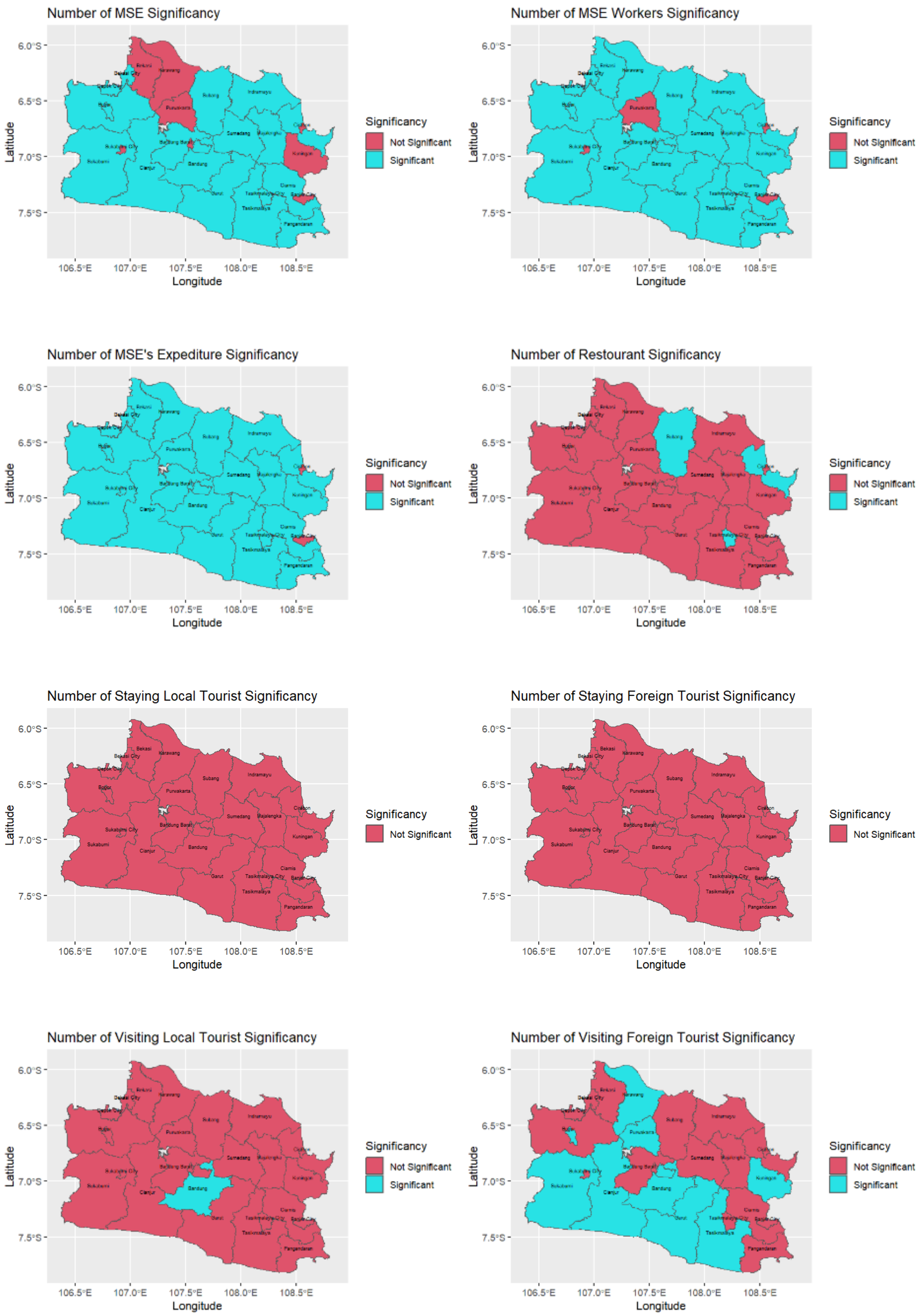




Figures 3. Disaster group variables significancy



Figures 4. Energy Source group variables significancy



Figures 5. Industry and Tourism group variables significancy

The spatial modelling result using MGWR with Group LASSO in Bi-square model as the best model has regression equation on the mean level as follows:

$$y_i = 122608{,}8334 + 8{,}5672X_{71} - 10{,}2736X_{72} - 1261{,}2520X_{81} - 28{,}7324X_{82} - 68{,}7266X_{83} + 53{,}7008X_{84} + 523{,}9789X_{101} + 6{,}7084X_{102} - 18{,}7635X_{103} + 234{,}3642X_{104} + 18{,}9748X_{105} - 0{,}4302X_{171} + 4{,}2617X_{172} + 1{,}2506X_{173} + 0{,}1228X_{174} + 0{,}0002X_{175} - 0{,}0027X_{176} - 0{,}0216X_{177} + 0{,}0107X_{178}$$

For example, the Number of childbearing age couples (X_{71}) has a mean value of 8,5672, it can be interpreted that this variable increases the MSE income by 8,5672 for every 1 increment. On the other hand, the Number of family planning participants (X_{72}) has a mean value of -10,2736, it can be interpreted that this variable decreases the MSE income by -10,2736 for every 1 increment (Maulida, 2013; Rahmawati et al., 2020; Basbay et al., 2016; Brundage et al., 2014; Alemayehu et al., 2023; Rianty & Rahayu, 2021; Jalaliah et al., 2022). We can also map the significantly affecting group variables as shown in Figure 6.

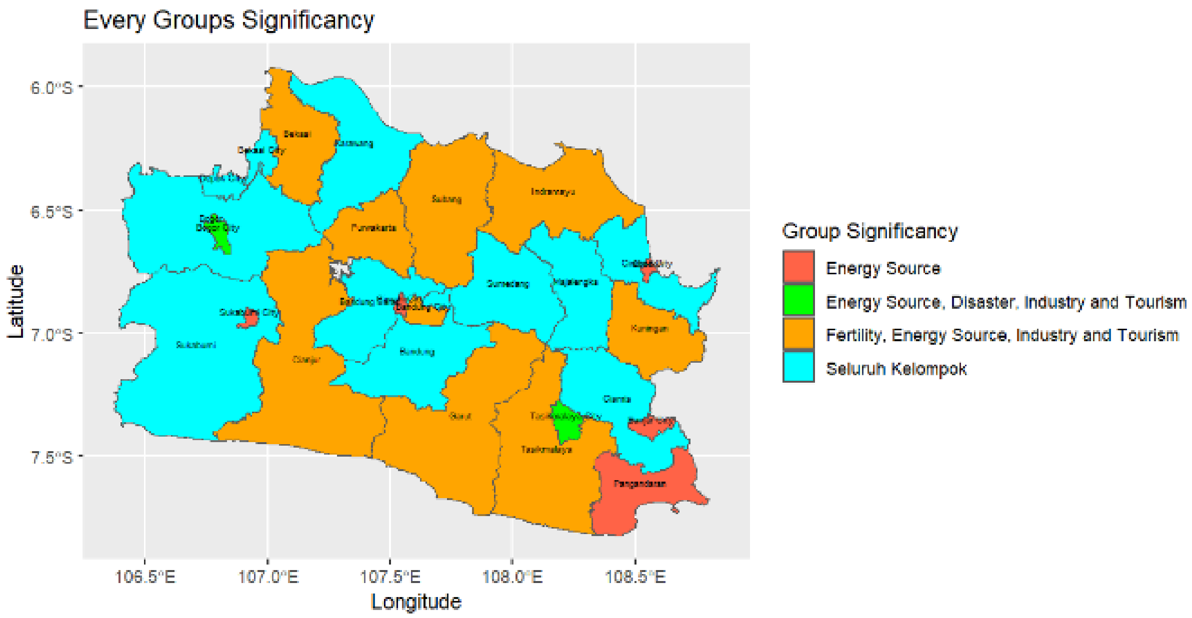


Figure 6. Group of Variables Significance

REFERENCES

Ainiyah, N., Deliar, A., & Virtriana, R. (2016). The Classical Assumption Test to Driving Factors of Land Cover Change in The Development Region of Northern Part of West java. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLI-B6*, 205–210. https://doi.org/10.5194/isprs-archives-XLI-B6-205-2016

Alemayehu, M. A., Agimas, M. C., Shewaye, D. A., Derseh, N. M., & Aragaw, F. M. (2023). Spatial distribution and determinants of limited access to improved drinking water service among households in Ethiopia based on the 2019 Ethiopian Mini Demographic and Health Survey: spatial and multilevel analyses. *Frontiers in Water*, *5*. https://doi.org/10.3389/frwa.2023.1166733

Al-Hasani, G., Asaduzzaman, M., & Soliman, A.-H. (2021). Geographically weighted Poisson regression models with different kernels: Application to road traffic accident data. *Communications in Statistics: Case Studies, Data Analysis and Applications*, *7*(2), 166–181. https://doi.org/10.1080/23737484.2020.1869628

Anshika, Singla, A., & Mallik, G. (2021). Determinants of financial literacy: Empirical evidence from micro and small enterprises in India. *Asia Pacific Management Review*, *26*(4), 248–255. https://doi.org/10.1016/j.apmrv.2021.03.001

Ayambila, S. N. (2023). Determinants of micro and small enterprises financial performance in the non-farm sector of Ghana: A quantile regression approach. *Cogent Economics & Finance*, *11*(2). https://doi.org/10.1080/23322039.2023.2225331

Basbay, M. M., Elgin, C., & Torul, O. (2016). Energy Consumption and the Size of the Informal Economy. *Economics*, *10*(1). https://doi.org/10.5018/economics-ejournal.ja.2016-14

Brundage, M. P., Chang, Q., Li, Y., Arinez, J., & Xiao, G. (2014). Utilizing energy opportunity windows and energy profit bottlenecks to reduce energy consumption per part for a serial production line. *2014 IEEE International Conference on Automation Science and Engineering (CASE)*, 461–466. https://doi.org/10.1109/CoASE.2014.6899366

Cavanaugh, J. E., & Neath, A. A. (2019). The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *WIREs Computational Statistics*, *11*(3). https://doi.org/10.1002/wics.1460

Comber, A., Harris, P., & Brunsdon, C. (2024). Multiscale spatially varying coefficient modelling using a Geographical Gaussian Process GAM. *International Journal of Geographical Information Science*, *38*(1), 27–47. https://doi.org/10.1080/13658816.2023.2270285

Daoud, J. I. (2017). Multicollinearity and Regression Analysis. *Journal of Physics: Conference Series*, *949*, 012009. https://doi.org/10.1088/1742-6596/949/1/012009

De Carvalho, F. de A. T., Lima Neto, E. de A., & Ferreira, M. R. P. (2017). A robust regression method based on exponential-type kernel functions. *Neurocomputing*, *234*, 58–74. https://doi.org/10.1016/j.neucom.2016.12.035

Díaz Rios, D. J., & González Castro, J. (2022). The commercial micro and small enterprise as a complex adaptive system. A management model. *SCIÉNDO*, *26*(2), 161–167. https://doi.org/10.17268/sciendo.2023.023

E. Colipano, T. (2022). Bussiness Performance of Micro-Enterprises in Northen Mindanao, Philippines: a Multinomial Logistic Regression Analysis. *International Journal of Applied Science and Research*, *05*(04), 105–118. https://doi.org/10.56293/IJASR.2022.5413

El-Habil, A. M. (2012). An Application on Multinomial Logistic Regression Model. *Pakistan Journal of Statistics and Operation Research*, *8*(2), 271. https://doi.org/10.18187/pjsor.v8i2.234

Esubalew, A. A., & Raghurama, A. (2020). The mediating effect of entrepreneurs' competency on the relationship between Bank finance and performance of micro, small, and medium enterprises (MSMEs). *European Research on Management and Business Economics*, *26*(2), 87–95. https://doi.org/10.1016/j.iedeen.2020.03.001

Fan, G.-F., Peng, L.-L., & Hong, W.-C. (2018). Short term load forecasting based on phase space reconstruction algorithm and bi-square kernel regression model. *Applied Energy*, *224*, 13–33. https://doi.org/10.1016/j.apenergy.2018.04.075

Fotheringham, A. S., Oshan, T. M., & Li, Z. (2023). *Multiscale Geographically Weighted Regression*. CRC Press. https://doi.org/10.1201/9781003435464

Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale Geographically Weighted Regression (MGWR). *Annals of the American Association of Geographers*, *107*(6), 1247–1265. https://doi.org/10.1080/24694452.2017.1352480

Fotheringham, A. S., Yu, H., Wolf, L. J., Oshan, T. M., & Li, Z. (2022). On the notion of 'bandwidth' in geographically weighted regression models of spatially varying processes. *International Journal of Geographical Information Science*, *36*(8), 1485–1502. https://doi.org/10.1080/13658816.2022.2034829

Halunga, A. G., Orme, C. D., & Yamagata, T. (2017). A heteroskedasticity robust Breusch–Pagan test for Contemporaneous correlation in dynamic panel data models. *Journal of Econometrics*, *198*(2), 209–230. https://doi.org/10.1016/j.jeconom.2016.12.00 5

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical Learning with Sparsity*. Chapman and Hall/CRC. https://doi.org/10.1201/b18401

Higazi, S. F., Abdel-Hady, D. H., & Ahmed Al-Oulfi, S. (2013). Application of Spatial Regression Models to Income Poverty Ratios in Middle Delta Contiguous Counties in Egypt. *Pakistan Journal of Statistics and Operation Research*, *9*(1), 93–110.

Huang, Y., Tibbe, T., Tang, A., & Montoya, A. (2024). Lasso and Group Lasso with Categorical Predictors: Impact of Coding Strategy on Variable Selection and Prediction. *Journal of Behavioral Data Science*, *3*(2). https://doi.org/10.35566/jbds/v3n2/montoya

Jalaliah, J., Wulandari, H. K., & Dumadi, D. (2022). Pengaruh Modal Kerja, Tenaga Kerja, dan Bahan Baku Terhadap Pendapatan UMKM Pabrik Tahu (Studi Empiris UMKM Tahu Kecamatan Banjarharjo Periode Tahun 2019-2021). *AURELIA: Jurnal Penelitian Dan Pengabdian Masyarakat Indonesia*, *1*(1), 68–78. https://doi.org/10.57235/aurelia.v1i1.32

Jurečková, J., & Picek, J. (2007). Shapiro–Wilk-type test of normality under nuisance regression and scale. *Computational Statistics & Data Analysis*, *51*(10), 5184–5191. https://doi.org/10.1016/j.csda.2006.08.026

Kappal, S. (2019). Data Normalization using Median & Median Absolute Deviation (MMAD) based Z-Score for Robust Predictions vs. Min-Max Normalization. *London Journal of Research in Science: Natural and Formal*, *19*(4), 39–44.

Ma, Z., & Huang, Z. (2023). A Bayesian Implementation of the Multiscale Geographically Weighted Regression Model with INLA. *Annals of the American Association of Geographers*, *113*(6), 1501–1515. https://doi.org/10.1080/24694452.2023.2187756

Maulida, Y. (2013). Pengaruh Tingkat Upah Terhadap Migrasi Masuk di Kota Pekanbaru. *Jurnal Ekonomi*, *21*(2), 1–12.

Nugroho, N. F. T. A., & Slamet, I. (2018). Geographically Weighted Regression Model with Kernel Bisquare and Tricube Weighted Function on Poverty Percentage Data in Central Java Province. *Journal of Physics: Conference Series*, *1025*, 012099. https://doi.org/10.1088/1742-6596/1025/1/012099

Olayemi, O. O., Okonji, P. S., & Ogbeiwi, M. O. (2022). Ease of Doing Business Reforms and Business Growth Among Selected Micro, Small And Medium Enterprises (MSMEs) in Present Day Nigeria: a Hierarchical Multiple Regression Modelling Approach. *Journal of Economics and Management Research*, *10*, 57–73. https://doi.org/10.22364/jemr.10.04

Oyedele, O. (2023). Determining the optimal number of folds to use in a K-fold cross-validation: A neural network classification experiment. *Research in Mathematics*, *10*(1). https://doi.org/10.1080/27684830.2023.2201015

Pillay, K. G., & Lin, S. P. (2023). Prediction of KLCI Index Through Economic LASSO Regression Model and Model Averaging. *Pakistan Journal of Statistics and Operation Research*, *19*(1), 103–113. https://doi.org/10.18187/pjsor.v19i1.4214

Rahmawati, F., Ananda, F. P., & Narmaditya, B. (2020). Socio-Economic Indicators and Income Inequality: Lesson from West Java in Indonesia. *Scientific Papers of the University of Pardubice, Series D: Faculty of Economics and Administration*, *28*(3). https://doi.org/10.46585/sp28031114

Rianty, M., & Rahayu, P. F. (2021). Pengaruh E-Commerce Terhadap Pendapatan UMKM Yang Bermitra Gojek Dalam Masa Pandemi Covid-19. *Akuntansi Dan Manajemen*, *16*(2), 153–167. https://doi.org/10.30630/jam.v16i2.159

Schneider, U., & Tardivel, P. J. (2020). The Geometry of Uniqueness and Model Selection of Penalized Estimators including SLOPE, LASSO and Basis Pursuit. *HAL Open Science*, *20*, 1–34. https://hal.science/hal-02548350v2

Sharif, S., & Kamal, S. (2018). Comparison of Significant Approaches of Penalized Spline Regression (P-splines). *Pakistan Journal of Statistics and Operation Research*, *14*(2), 289. https://doi.org/10.18187/pjsor.v14i2.1948

Takano, Y., & Miyashiro, R. (2020). Best subset selection via cross-validation criterion. *TOP*, *28*(2), 475–488. https://doi.org/10.1007/s11750-020-00538-1

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Yue, H., Duan, L., Lu, M., Huang, H., Zhang, X., & Liu, H. (2022). Modeling the Determinants of PM2.5 in China Considering the Localized Spatiotemporal Effects: A Multiscale Geographically Weighted Regression Method. *Atmosphere*, *13*(4), 627. https://doi.org/10.3390/atmos13040627

Yunus, M., Saefuddin, A., & Soleh, A. M. (2020). Pemodelan Statistical Downscaling dengan LASSO dan Group LASSO untuk Pendugaan Curah Hujan. *Indonesian Journal of Statistics and Its Applications*, *4*(4), 649–660. https://doi.org/10.29244/ijsa.v4i4.724

Zhong, S., Chen, D., Xu, Q., & Chen, T. (2013). Optimizing the Gaussian kernel function with the formulated kernel target alignment criterion for two-class pattern classification. *Pattern Recognition*, *46*(7), 2045–2054. https://doi.org/10.1016/j.patcog.2012.12.012