Modeling Mortality Data Using Multi-Scale Geographically Weighted Regression

Oxy Susanto[1], Dian Lestari[1,2,*], and Fida Fathiyah Addini[1,2]
[1]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok, Indonesia
[2]Advanced Risk, Actuarial, and Financial Analytics Laboratory (ARAFA Lab), Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok, Indonesia
*Corresponding author: dian.lestari@sci.ui.ac.id

Received: 1 June 2026
Revised: 17 August 2026
Accepted: 19 August 2026

ABSTRACT – **Insurance is a contractual agreement in which the insurer provides financial protection against uncertain risks in exchange for a premium paid by the insured. In the context of life insurance, accurate estimation of mortality rates is essential, as it directly influences premium determination. This study aims to develop a Multi-scale Geographically Weighted Regression (MGWR) model and compare its performance with Multiple Linear Regression (MLR) and Geographically Weighted Regression (GWR) in the spatial analysis of mortality data. While MLR assumes a constant relationship between predictor and response variables across all locations, it fails to account for spatial heterogeneity. GWR addresses this limitation by constructing local regression models using geographical weights. However, GWR applies a single bandwidth for all parameters, which may reduce estimation accuracy. MGWR overcomes this limitation by assigning distinct bandwidths to each parameter through a back-fitting algorithm, resulting in greater model flexibility and precision. The proposed model is applied to the logarithm of mortality counts data in Japan, and the analysis indicates that MGWR with a fixed bandwidth and tricube kernel function yields the best performance, achieving the lowest corrected Akaike Information Criterion (AICc) and the highest adjusted coefficient of determination, outperforming both GWR and MLR. Moreover, MGWR produces lower and more evenly distributed residuals, demonstrating its capability to capture complex spatial variations. These results highlight MGWR's superiority in estimating mortality rates, thereby enabling more accurate, region-specific life insurance premium calculations.**
Keywords – **back-fitting, bandwidth, Bonferroni correction, weighted least square**

I. INTRODUCTION

According to Article 246 of the Indonesian Commercial Code (Kitab Undang-Undang Hukum Dagang), insurance is a contract where an insurer provides compensation to an insured for losses or damages resulting from an uncertain event, receiving a premium in return. The determination of this premium is a critical aspect of the insurance industry, as it must be competitive while also being sufficient to cover potential losses. In life insurance, premiums are heavily influenced by the expected future payout of death claims, for which mortality rates are a key component. Inaccurate estimations of mortality rates can significantly impact premiums, as these rates are involved in repeated and cumulative calculations over the coverage period. Therefore, insurance companies require comprehensive data analysis to ensure fair and profitable premium settings.

Linear regression is a widely used analytical tool in various fields, including insurance, due to its simplicity and effectiveness in explaining the relationship between variables [1]. For example, it has been used to assess the influence of environmental factors on mortality rates [2] and to model personal health insurance costs based on factors like age and BMI [3]. However, a major limitation of linear regression is its assumption that relationships between variables are uniform across all regions. These relationships often vary due to local characteristics such as geography, socioeconomic conditions, and policies, leading to spatial heterogeneity that linear regression cannot capture. In the context of mortality rate estimation, the effect of factors like medical facility quality can differ between regions based on local conditions [4], making a more localized modeling approach essential for accurate premium determination.

To address this limitation, Geographically Weighted Regression (GWR) was developed. Unlike a single global linear regression model, GWR creates a separate local regression model for each observation point. It follows Tobler's First Law of Geography [5], "everything is related to everything else, but near things are more related than distant things." To achieve this, GWR uses a weighted least square (WLS) method, which assigns greater weight to observations closer to the point being analysed, with the weighting determined by a kernel function and a bandwidth parameter [6]. This method allows GWR to capture spatial heterogeneity and produce a set of unique parameter estimates for each location, providing a more accurate understanding of the relationship between variables across different areas [7].

Despite its advantages over linear regression, GWR has a key limitation: it assumes all parameters in the model have the same bandwidth. The spatial scale of influence can vary for different factors. For instance, socioeconomic factors might have a broad impact [8] while environmental factors might only affect a small local area [9]. Forcing a single bandwidth on all parameters reduces the model's accuracy and fails to capture the true complexity of spatial relationships. This can also increase the risk of local multicollinearity, making parameter estimates unstable and model interpretation more difficult [1].

To overcome GWR's limitations, Multi-Scale Geographically Weighted Regression (MGWR) was developed. MGWR is a more flexible and accurate method because it allows each parameter to have its own unique bandwidth, enabling the analysis of variable relationships at their appropriate spatial scales [10]. This unique feature also helps mitigate the risk of local multicollinearity that is inherent in GWR. Since MGWR cannot use the WLS method directly due to varying bandwidths, it employs a back-fitting algorithm, an iterative process that updates each parameter's estimate one by one until all converge [7]. This provides more stable and accurate estimates by accommodating spatial variations flexibly.

The flexibility and accuracy of MGWR have led to its use in various fields, including health and disaster management. For example, it has been used to analyse spatial variations in COVID-19 hospitalization and mortality rates [11] and to capture multi-scale variations in factors related to obesity rates [10]. In the insurance industry, the ability of MGWR to model data while accounting for geographical variations makes it valuable not only for premium calculation but also for strategic risk management. It can be used as an internal tool to identify high-risk areas and understand the factors contributing to high mortality rates, providing an important risk-based alternative to socially sensitive premium differentiation policies. This paper systematically discusses the MGWR modeling workflow and applies it to mortality data in Japan to analyse the impact of spatial variation.

II. LITERATURE REVIEW

2.1 Multiple Linear Regression

Multiple Linear Regression (MLR) is a statistical method used to explain the relationship between a response variable y_i and multiple predictor variables $(x_{i1}, x_{i2}, \dots, x_{ik})$. The model assumes a linear relationship and is expressed as [1]:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i \tag{1}$$

where y_i is the response for the i-th observation, β_0 is the intercept, β_k are regression coefficients, x_{ik} are predictor values, and $\epsilon_i \sim N(0, \sigma^2)$ are independent errors. In matrix form, the model is written as:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon} \tag{2}$$

where $\boldsymbol{y}$ is an $n \times 1$ vector of responses, $\boldsymbol{X}$ is an $n \times p$ matrix of predictors (including intercept), $\boldsymbol{\beta}$ is a $p \times 1$ vector of coefficients, and $\boldsymbol{\epsilon}$ is an $n \times 1$ error vector.

The goal of parameter estimation is to find the vector $\widehat{\boldsymbol{\beta}}$ that best describes the relationship between the response vector $\boldsymbol{y}$ and the predictor matrix $\boldsymbol{X}$. This is achieved using the Ordinary Least Squares (OLS) method, which minimizes the Sum of Squared Errors (SSE):

$$SSE = \boldsymbol{y}^T\boldsymbol{y} - 2\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{y} + \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X\beta} \tag{3}$$

Taking the derivative with respect to β and setting it to zero gives:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \tag{4}$$

After estimating the regression parameters, simultaneous (F-test) and partial (t-test) significance tests are used to assess model fit and identify insignificant predictors. For these tests to be valid, key assumptions must be met; normality of residuals, constant variance (homoscedasticity), independence of residuals, no multicollinearity, and linearity between predictors and the response. Violations may affect the reliability of the model.

MLR is a foundational global model that provides a systematic framework for analysing the relationship between variables. However, it assumes spatial homogeneity, the same relationship holds across all locations, which limits its effectiveness for spatial data. This motivates the use of spatial models such as Geographically Weighted Regression (GWR) and Multi-Scale GWR (MGWR) to account for local variation. Spatial heterogeneity describes how relationships between variables differ across locations due to factors like environment or socioeconomic conditions. Since traditional regression assumes uniform relationships, it may fail to capture these differences. Heteroscedasticity in residuals can indicate such spatial variation. To address this, GWR and MGWR estimate location-specific coefficients, using nearby data points and standardized predictors to better reflect local dynamics.

2.2 Geographically Weighted Regression

Geographically Weighted Regression (GWR) is a regression technique that allows the coefficients to vary by location, making it an extension of traditional regression. However, a limitation of GWR is its assumption that the same bandwidth is applied to all parameters. The GWR model is expressed as [12]:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{m} \beta_k(u_i, v_i) x_{ik} + \epsilon_i \tag{5}$$

where y_i is the response at location i, x_{ik} is the k-th predictor at location i, (u_i, v_i) are the coordinates for location i, $\beta_0(u_i, v_i)$ is the local intercept, $\beta_k(u_i, v_i)$ is the local coefficient for predictor k, and ϵ_i is the error term at location i.

When it comes to GWR, the bandwidth is a crucial parameter that dictates the range of influence from nearby locations during the estimation of local parameters. A fixed bandwidth maintains a constant distance, which is ideal for data points that are evenly spaced. Conversely, an adaptive bandwidth uses a fixed number of nearest neighbours, making it a better choice for data that's unevenly distributed, as it helps reduce estimation bias. The most common way to measure the distance between two points in GWR is with the Euclidean distance, which is simply the straight-line distance [13]. Based on the Pythagorean theorem, it's calculated as follows:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{6}$$

where x and y represent the longitude and latitude coordinates of locations i and j, respectively.

A kernel function is a mathematical tool used to assign spatial weights w_{ij} based on the distance d_{ij} between locations i and j, and the bandwidth bw [14]. It determines how much influence a neighboring location has on the parameter estimation at a target location. This study compares four types of kernel functions: Gaussian, exponential, bisquare, and tricube, each with unique characteristics that influence the analysis results.

In GWR, the spatial weight matrix $\boldsymbol{W}(u_i, v_i)$ is an $n \times n$ diagonal matrix used to estimate local parameters $\boldsymbol{\beta}(u_i, v_i)$ at each location. It reflects how much influence surrounding locations have on the target location (u_i, v_i). Each diagonal element $w_j(u_i, v_i)$ represents the weight of location j on location i, calculated using a chosen kernel function and bandwidth [15]:

$$\boldsymbol{W}(u_i, v_i) = \begin{bmatrix} w_1(u_i, v_i) & 0 & \cdots & 0 \\ 0 & w_2(u_i, v_i) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n(u_i, v_i) \end{bmatrix} \tag{7}$$

Unlike MLR which uses OLS for parameter estimation, GWR applies Weighted Least Squares (WLS). The goal is to minimize the weighted sum of squared errors (WSSE):

$$WSSE = \boldsymbol{\epsilon}^T \boldsymbol{W}(u_i, v_i)\boldsymbol{\epsilon} = (\boldsymbol{y} - \boldsymbol{X\beta})^T \boldsymbol{W}(u_i, v_i)(\boldsymbol{y} - \boldsymbol{X\beta}) \tag{8}$$

Expanding and minimizing the expression with respect to $\boldsymbol{\beta}(u_i, v_i)$, we get the GWR estimator:

$$\widehat{\boldsymbol{\beta}}(u_i, v_i) = [\mathbf{X}^T \boldsymbol{W}(u_i, v_i)\mathbf{X}]^{-1}\mathbf{X}^T \boldsymbol{W}(u_i, v_i)\boldsymbol{y} \tag{9}$$

Bandwidth optimization aims to determine the best bandwidth value so that the GWR model captures spatial patterns accurately without overfitting or underfitting. To find the optimal bandwidth, the corrected Akaike Information Criterion (AICc) is used as the objective function, as it balances model fit and complexity while adjusting for small sample bias. In GWR, AICc is bandwidth-dependent and thus well-suited for this optimization. The AICc is defined as:

$$AICc = 2n\log(\hat{\sigma}) + n\log(2\pi) + n\left(\frac{n + \mathrm{tr}(\boldsymbol{S})}{n - 2 - \mathrm{tr}(\boldsymbol{S})}\right) \tag{10}$$

where $\hat{\sigma}$ is the residual standard deviation, n is number of observations, and $\mathrm{tr}(\boldsymbol{S})$ is the trace of the hat matrix $\boldsymbol{S}$, defined such that $\widehat{\boldsymbol{y}} = \boldsymbol{Sy}$. The optimal bandwidth is selected by minimizing the AICc value using the Golden Section Search, a numerical method efficient for optimizing single-variable functions [6, 16].

The Golden Section Search (GSS) is an efficient numerical algorithm used to find the minimum or maximum of a single-variable function within a specified interval. What makes it particularly useful is that it doesn't require the function's derivative, making it suitable for non-differentiable functions or when an analytical derivative is too complex to find. At its core, the method leverages the golden ratio, an irrational number approximately equal to 1.618 or its inverse, 0.618. In the context of GWR, GSS is a crucial tool for finding the optimal bandwidth by minimizing a function that measures model fit, such as the AICc. For a minimization problem, GSS begins with an initial interval and a tolerance level. Two interior points are strategically placed using the golden ratio. The function (in this case, the AICc value) is then evaluated at these points. Based on which value is smaller, the search interval is narrowed. This process is repeated until the interval is smaller than the specified tolerance, at which point the midpoint of the final interval is taken as the approximate optimum. Because it systematically reduces the search space with minimal function evaluations, GSS is computationally frugal and widely used for optimizing parameters in various numerical experiments [17]. Model selection is based on two metrics: adjusted $R^2$and AICc (Corrected Akaike Information Criterion). Adjusted $R^2$is used to independently evaluate model performance, while AICc is used to compare the relative quality between models.

III. METHODOLOGY

3.1 Multi-Scale Geographically Weighted Regression

Moving beyond the constraints of traditional GWR, Multi-Scale Geographically Weighted Regression (MGWR) was developed as a more advanced and nuanced approach to spatial analysis. It solves the critical problem of GWR's single-bandwidth assumption by letting each predictor variable have its own distinct spatial scale of influence, represented by a unique bandwidth. This key innovation allows MGWR to capture the true complexity of a dataset, leading to more flexible and accurate models. The general form of the MGWR model is:

$$y_i = \sum_{k=0}^{m} \beta_{bwk}(u_i, v_i)x_{ik} + \epsilon_i \tag{11}$$

where y_i is the response at location i, $\beta_{bwk}(u_i, v_i)$ is the local coefficient for predictor k estimated with its own bandwidth, x_{ik} is the predictor value, and ϵ_i is the error term.

MGWR constructs a separate weight matrix for each predictor, since each one has its own bandwidth. As a result, MGWR generates a unique weight matrix for every combination of predictor and observation location, making the estimation process more complex. The weight matrix for predictor k at location (u_i, v_i) is defined as:

$$W^{(k)}(u_i, v_i) = \begin{bmatrix} w_1^{(k)}(u_i, v_i) & 0 & \cdots & 0 \\ 0 & w_2^{(k)}(u_i, v_i) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n^{(k)}(u_i, v_i) \end{bmatrix} \tag{12}$$

where $W^{(k)}(u_i, v_i)$ represents the spatial weight for predictor k between observation j and target location i.

3.2 Back-fitting Algorithm

While the standard GWR model relies on a straightforward WLS method for parameter estimation, the complexity of MGWR requires a more sophisticated solution. Since each predictor in MGWR has its own unique bandwidth and weight matrix, a single unified WLS calculation isn't possible. To overcome this, Fotheringham et al. [10] proposed an elegant back-fitting algorithm, drawing inspiration from a technique used in Generalized Additive Models [18]. This iterative process begins by initializing estimates (often from a global GWR model) and then cycles through each predictor one at a time. For each predictor, a partial residual is calculated by removing the effects of all other variables. This residual is then used in a GWR model to update the local coefficients and bandwidth for that specific predictor. The entire process is repeated until the model reaches convergence, ensuring each variable's spatial scale is optimally determined.

Each regression term $\beta_{bwk}(u_i, v_i)x_{ik}$ in MGWR is reformulated as an additive component f_{ik}, resulting in a model structure:

$$y_i = \sum_{k=0}^{m} f_{ik} + \epsilon_i \tag{13}$$

To initialize, estimates from a GWR model are used as starting values for each f_k. Then, for each predictor x_n, a partial residual is computed:

$$y - \left(\sum_{k=0, k\neq n}^{m} f_k \right) \tag{14}$$

Convergence is determined by a metric called the Score of Change - Sum of Squared Errors (SOC-SSE), which is calculated as:

$$\text{SOC-}SSE = \frac{|SSE_i - SSE_{i-1}|}{SSE_i} \tag{15}$$

The model is considered converged when the SOC-SSE falls below a predetermined threshold, such as 10^{-5} [7]. This iterative approach ensures that MGWR can produce variable-specific bandwidths, offering a more flexible and spatially nuanced model.

3.3 Statistical Inference and Model Assessment

MGWR's multi-scale nature requires specialized methods for statistical inference and model assessment. While local t-tests can be used to assess the significance of each regression coefficient at a given location, performing numerous tests increases the risk of false positives (Type I errors). To address this, a modified Bonferroni correction is applied. This method adjusts the significance level based on the Effective Number of Parameters (ENP) for each predictor. The ENP is derived from the trace of the hat matrix for each parameter and provides a more conservative and balanced approach to inference. The corrected significance level for a parameter is given by:

$$\alpha_k{}^* = \frac{\alpha}{ENP_k} \tag{16}$$

with $ENP_k = \mathrm{tr}(S_k)$.

To determine if the spatial variations in coefficients are statistically meaningful or just random noise, a Monte Carlo test is employed. By comparing the observed coefficient variability to that from randomized simulations, this test distinguishes between local parameters (those with significant spatial variation) and global parameters (those with no significant spatial pattern). This is crucial for understanding whether the spatial patterns in the model truly reflect geographic structure.

Multicollinearity, or high correlation between predictor variables, is generally less of a concern in MGWR than in global models. The use of variable-specific bandwidths localizes the analysis, which naturally helps to reduce correlations among predictors within each specific location. While MGWR doesn't have a dedicated algorithm for addressing local multicollinearity, it is typically managed during the variable selection stage of the global model [7].

Like GWR, MGWR uses AICc for model comparison. However, since MGWR lacks a single hat matrix, the total effective number of parameters must be calculated by summing the individual ENPs for each predictor:

$$ENP_{total} = \sum_{k=0}^{m} \mathrm{tr}(S_k) \tag{17}$$

The MGWR AICc is then given by [16]:

$$AICc = 2n\log(\hat{\sigma}) + n\log(2\pi) + n\left(\frac{n + ENP_{total}}{n - 2 - ENP_{total}}\right) \tag{18}$$

Additionally, Adjusted R-squared is used to assess model quality. In MGWR, it is adapted to use the ENP_{total} instead of the raw count of predictors, ensuring that both AICc and Adjusted R-squared accurately reflect the model's multi-scale complexity.

$$Adjusted\text{−}R^2 = 1 - \left(\frac{(1 - R^2)(n - 1)}{n - ENP_{total} - 1}\right) \tag{19}$$

IV. RESULTS AND DISCUSSIONS

This study uses the Tokyo Mortality Rate dataset, covering deaths among the working-age population (25–64) across 262 areas within 70 km of Chiyoda, Tokyo, in 1990. The dataset, free of missing values or outliers, was sourced from Arizona State University and previously used for spatial mortality analysis [15]. The variables used in this study are listed in Table 1.

After performing descriptive, all predictor variables were standardized to eliminate scale differences, allowing for meaningful comparison of regression coefficients. The response variable, db2564 (deaths among the working-age population), was log-transformed to reduce skewness and approximate normality, improving model assumptions. A correlation matrix was then used to assess linear relationships between variables and detect potential multicollinearity, supporting the development of a more accurate regression model.

Table 1 Description of variables.

Variable	Type	Description
X_CENTROID	Numeric (Continuous)	X-coordinate of the area's centroid, used to determine geographic location in Cartesian coordinates.
Y_CENTROID	Numeric (Continuous)	Y-coordinate of the area's centroid, used to determine geographic location in Cartesian coordinates.
db2564	Numeric (Discrete)	Observed number of deaths among the working-age population (25–64) in each area.
eb2564	Numeric (Continuous)	Expected number of deaths in the 25–64 age group, calculated based on known mortality rates and local population size.
OCC_TEC	Numeric (Continuous)	Proportion of professional workers in an area, reflecting socioeconomic status.
OWNH	Numeric (Continuous)	Proportion of owner-occupied housing, indicating economic stability and social well-being.
POP65	Numeric (Continuous)	Proportion of elderly residents (65+), reflecting demographic structure and health service needs.
UNEMP	Numeric (Continuous)	Unemployment rate, representing the percentage of job seekers without employment.

Based on Figure 1, there is generally no strong correlation among the predictor variables, indicating that multicollinearity is unlikely. This is further supported by the Variance Inflation Factor (VIF) values for eb2564, POP65, UNEMP, and OWNH are 1.577, 1.408, 1.527, and 3.876, respectively. All VIF values are well below the commonly accepted threshold of 10, and even below the more conservative threshold of 5. These results indicate that multicollinearity is not a serious concern in the initial model based on the global VIF assessment. The response variable (db2564) shows moderately strong correlations with the predictors, supporting the use of multiple linear regression (MLR) as an initial global model. To maintain model parsimony, OCC_TEC was excluded from the subsequent analysis because it consistently lacked statistical significance across the preliminary MLR, GWR, and MGWR models. Its removal was considered in relation to the possibility of omitted-variable bias. Although statistical insignificance alone cannot completely rule out such bias, the consistent lack of evidence that OCC_TEC contributed to explaining the response variable provided the basis for its exclusion. The remaining predictors were therefore retained for the subsequent spatial modeling analysis.

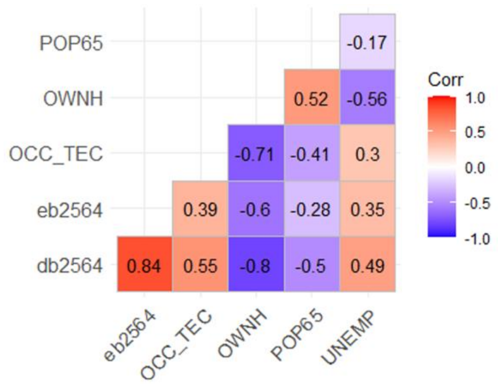

Figure 1 Correlation matrix.

MLR was conducted first to confirm global linear relationships before applying spatial models, making it a suitable starting point for the analysis. The estimated MLR equation is as follows:

$$\hat{y} = 4.52 + 0.67x_1 - 0.41x_2 - 0.18x_3 + 0.10x_4 \quad (20)$$

where $\hat{y}$ is the estimated logarithm of deaths among individuals aged 25–64 (log(db2564)), x_1 is eb2564 variabel, x_2 is OWNH variable, x_3 is POP65 variable, and x_4 is UNEMP variable.

The next step in MLR is to test assumptions to ensure model validity. When met, regression coefficients can be interpreted reliably. Hypothesis testing is also conducted using the F-test to assess overall model significance and the t-test to evaluate the contribution of each predictor. These steps confirm the model's adequacy before advancing to spatial methods.

Table 2 MLR Assumption Test.

Assumption	Method	p-value	Status
Normality	Kolmogorov-Smirrnov	0.4399	Satisfied
Residual Autocorrelation	Durbin-Watson	1e-10***	Not Satisfied
Homoscedasticity	Breusch-Pagan	0.01007*	Not Satisfied

As shown in Table 2, the failure to satisfy the assumptions of homoscedasticity and independent errors (autocorrelation) indicates that the conventional multiple linear regression (MLR) framework is not fully appropriate as the primary modeling approach for this dataset. Relying solely on MLR estimates under these violations may lead to misleading statistical inferences. Therefore, modeling spatial dependence becomes a methodological necessity, motivating the use of geographically weighted regression (GWR) and multi-scale geographically weighted regression (MGWR). The regression results show that all predictor variables, eb2564, OWNH, POP65, and UNEMP, are statistically significant, each with p-values well below the 0.05 threshold. Furthermore, the F-test confirms the overall significance of the model, with a p-value of < 2.2e-16.

Table 3 MLR, GWR, and MGWR Comparison.

	MLR	GWR	MGWR
AICc	341.0518	236.6671	171.5484
Adjusted R^2	85.39%	91.72%	93.78%

Spatial modeling using MLR, GWR, and MGWR was applied to capture how relationships between variables vary across regions. To find the best model, different bandwidth types (fixed and adaptive) and kernel functions (Gaussian, exponential, bisquare, tricube) were compared, as each influences spatial weighting and coefficient interpretation. Table 3 compares the performance of MLR, GWR, and MGWR using AICc and adjusted R^2. MGWR outperforms both MLR and GWR, with the lowest AICc (171.55) and highest adjusted R^2 (93.78%), indicating superior model fit and explanatory power. Furthermore, Table 4 show that MGWR consistently outperforms GWR in terms of AICc and adjusted R^2. Although MGWR is more complex, it remains more accurate and flexible for capturing multi-scale spatial heterogeneity. The only exception is the adaptive exponential kernel, where GWR slightly outperforms MGWR due to the latter's complexity not being matched by a proportional gain in model fit.

Table 4 AICc and Adjusted R^2 of GWR and MGWR.

Bandwidth		Gaussian	Exponential	Bisquare	Tricube
Fixed	AICc GWR	240.8173	261.5132	239.1816	240.1283
	AICc MGWR	**184.6761**	**236.8634**	**173.2163**	**171.5484**
	Adjusted R^2 GWR	91.63%	90.80%	91.36%	91.23%
	Adjusted R^2 MGWR	**93.80%**	**92.53%**	**93.88%**	**93.78%**
Adaptive	AICc GWR	254.3361	**277.5795**	238.8798	236.6671
	AICc MGWR	**243.1192**	278.1938	**203.6524**	**198.4481**
	Adjusted R^2 GWR	90.23%	89.11%	91.67%	91.72%
	Adjusted R^2 MGWR	**91.15%**	**89.43%**	**93.12%**	**93.18%**
Bandwidth		Gaussian	Exponential	Bisquare	Tricube
Fixed	AICc GWR	240.8173	261.5132	239.1816	240.1283
	AICc MGWR	**184.6761**	**236.8634**	**173.2163**	**171.5484**
	Adjusted R^2 GWR	91.63%	90.80%	91.36%	91.23%
	Adjusted R^2 MGWR	**93.80%**	**92.53%**	**93.88%**	**93.78%**

Based on the AICc values, the GWR model using adaptive bandwidth with the tricube kernel had the lowest AICc, indicating the best model fit. This is further supported because the same model achieved the highest adjusted R^2. Therefore, the GWR model with adaptive bandwidth and tricube kernel was selected as the optimal model for capturing the spatial patterns in the data. This model uses a bandwidth of 55 observations, meaning that for estimating the regression coefficients at each observation point, the 55 nearest neighbouring observations are taken into account.

Based on the AICc values, the MGWR model with fixed bandwidth and tricube kernel showed the best performance. Although the highest adjusted R^2 (93.88%) was achieved using the bisquare kernel, model selection focused on AICc minimization, in line with MGWR modeling principles. Therefore, the fixed bandwidth tricube model was chosen as the optimal MGWR specification. The selected MGWR model applies a fixed bandwidth, with different spatial scales (in distance units) assigned to each parameter: Intercept (28.944), eb2564 (16.310), OWNH (134.652), POP65 (52.547), and UNEMP (48.845). This confirms the model's ability to capture varying spatial influences for each variable.

An anomaly was observed when utilizing the adaptive exponential kernel, where the standard GWR model yielded a lower AICc than the MGWR model. This discrepancy can be explained by examining model complexity through the Effective Number of Parameters (ENP). The standard GWR model applied a single bandwidth of 17, resulting in an ENP of 54.87. Conversely, the MGWR model optimized much smaller bandwidths for the majority of parameters, increasing the ENP to 72.31. For this specific kernel, the increased model complexity in MGWR was not accompanied by a proportional improvement in likelihood, thereby resulting in a less favorable AICc score. Thus, the slightly better performance of GWR under this particular configuration does not necessarily indicate that GWR is generally superior to MGWR, but rather reflects the trade-off between model flexibility and complexity for this kernel.

To further assess the robustness of the spatial models, local multicollinearity was evaluated. For the standard GWR model, local Variance Inflation Factors (VIFs) were mapped across the study area, and none of the local VIF values exceeded the critical threshold of 10. This result indicates that severe local multicollinearity was not evident in the GWR specification. For MGWR, local multicollinearity was considered when interpreting the variable-specific bandwidths, since different spatial scales may produce different levels of local correlation among predictors. However, the use of variable-specific bandwidths does not by itself eliminate local multicollinearity. Therefore, the absence of severe global multicollinearity at the initial MLR stage, together with the local VIF assessment for GWR, was used as supporting evidence for the reliability of the spatial models. The potential for local correlation was also considered when interpreting the MGWR coefficient estimates.

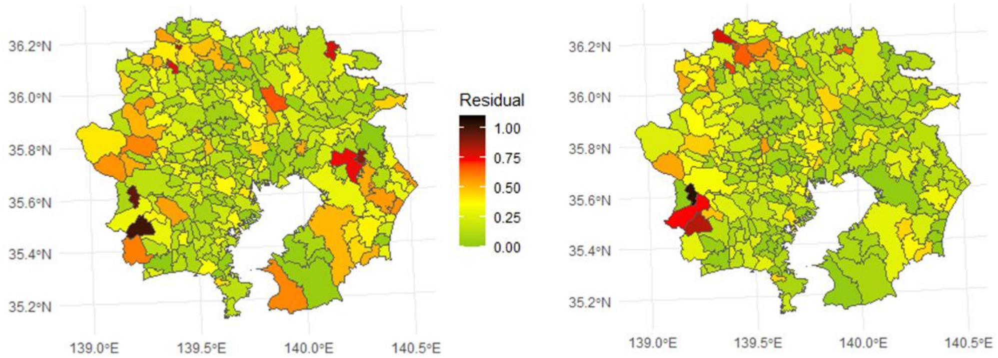

Figure 2 GWR (left) and MGWR (right) Residuals.

Figure 2 illustrates that most residuals fall in the low-to-moderately low categories, indicating good model performance with relatively small prediction errors. While peripheral prediction errors exist in both models, potentially reflecting spatial edge effects, the residual plots demonstrate that MGWR yields substantially fewer and smaller errors than standard GWR. This overall reduction in residuals indicates that MGWR provides an improved fit despite the presence of extreme local variations at the boundaries. Meanwhile, Figure 3 visualize the spatial variation of the estimated MGWR coefficients. The coefficient for eb2564 is stronger in the northern region, while OWNH shows mostly negative values, especially in the east. The coefficient for POP65 tends to be lower in peripheral areas, while UNEMP shows a spatial contrast, with positive effects in the south and negative effects in the north. The Monte Carlo test results assess whether coefficient variability across space is statistically significant. Although only the intercept and eb2564 show significant spatial variation (p-value < 0.05), indicating a need for local modeling, MGWR was applied to all variables, including predictors such as OWNH, POP65, and UNEMP that did not exhibit statistically significant spatial variation. This approach allows the MGWR backfitting algorithm to determine the appropriate spatial scale for each predictor. Predictors with insignificant spatial variation may be assigned large or maximum bandwidths and therefore operate effectively as global parameters, rather than requiring a manual separation of variables into local and global components.

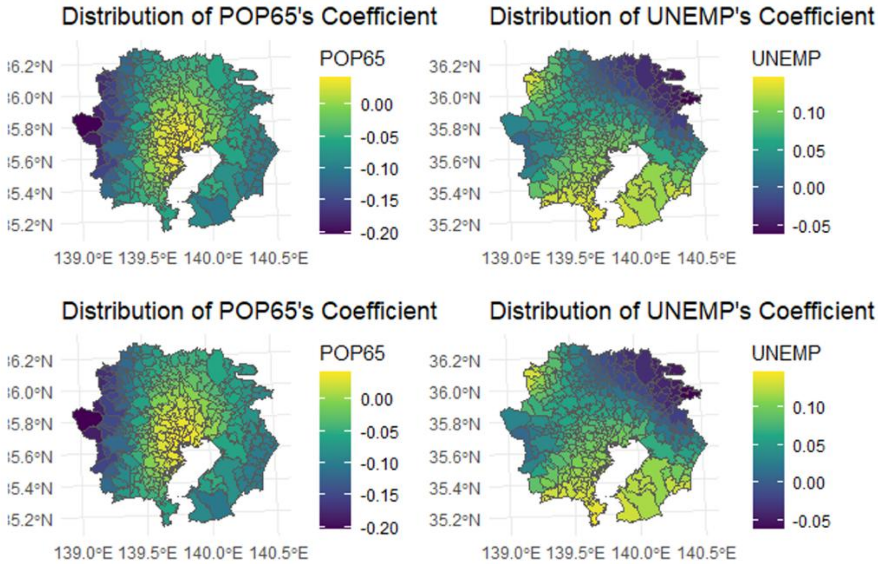

Figure 3 MGWR Coefficients.

V. CONCLUSIONS AND SUGGESTIONS

This study went beyond a simple application, focusing on the theoretical development of the Multi-Scale Geographically Weighted Regression (MGWR) model. Our methodological approach aimed to more accurately capture spatial heterogeneity, a key limitation of the standard GWR model. We constructed the MGWR model using a sophisticated back-fitting algorithm, which uniquely allows each regression coefficient to be estimated at its optimal spatial scale by assigning variable-specific bandwidths. This innovation not only boosts model accuracy but also significantly lowers the risk of local multicollinearity by aligning the scale of analysis with the distinct spatial characteristics of each variable.

The effectiveness of our approach was confirmed through rigorous model validation. The MGWR model's performance was superior to GWR in modeling the number of deaths among the productive-age population in Tokyo, Japan. This was evident from its lower AICc and higher adjusted $\boldsymbol{R}^2$ values, which indicate a better fit to the data. Furthermore, MGWR produced more evenly distributed and smaller residuals, confirming its superior ability to capture complex spatial dynamics. The results underscore MGWR's potential for providing more accurate, location-specific insights across various fields of spatial analysis, including its promising applications within the insurance industry.

Moving forward, we recommend two key areas for future research. First, extending the MGWR framework to a more generalized form that allow the analysis of response variables with non-normal distributions, such as count data like mortality or insurance claims. Second, incorporating a temporal dimension. This model would optimize bandwidths across both space and time, offering a more nuanced way for the insurance industry to analyse and adjust premiums based on dynamic, localized risk trends.

ACKNOWLEDGEMET

This research is funded by the Matching Fund Research Grant Scheme between FMIPA UI and FMIPA UNTAN, 2025, under research contract number: PKS-18/UN2.F3.D/PPM.00.02/2025.

REFERENCES

[1] Montgomery D C, Peck E A, and Vining G G 2012 *Introduction to linear regression analysis* (New York: Wiley)

[2] Samillan V J, Flores-León D, Rojas E, and Zutta B R (Environmental and climatic impact on the infection and mortality of SARS-CoV-2 in Peru) 2021 *J. Basic Clin. Physiol. Pharmacol.* **32** 121-127. https://doi.org/10.1515/jbcpp-2021-0007

[3] Wang X and Tuo S (Predictive modeling of personal medical insurance costs: Analyzing key factors and interactions) 2024 *J. Comput. Electron. Inf. Manage.* **15**(2) 45-56. https://doi.org/10.54097/nqsw4120

[4] Nolte E and McKee C M (Measuring the health of nations: Updating an earlier analysis) 2008 *Health Aff.* **27**(1) 58-71. https://doi.org/10.1377/hlthaff.27.1.58

[5] Tobler W R (A computer movie simulating urban growth in the Detroit region) 1970 *Econ. Geogr.* **46**(Sup1) 234-240. https://doi.org/10.2307/143141

[6] Fotheringham A S, Brunsdon C, and Charlton M 2002 *Geographically weighted regression: The analysis of spatially varying relationships* (New York: Wiley)

[7] Fotheringham A S, Oshan T M, and Li Z 2023 *Multiscale geographically weighted regression: Theory and practice* (Boca Raton: CRC Press)

[8] Dillon E (Unemployment trends and their macroeconomic implications: A global perspective) 2024 *J. Econ. Econ. Educ. Res.* **25**(S4) 112-120.

[9] Zhang Y, Hannigan M, and Lv Q 2021 *Proceedings of the 29th International Conference on Advances in Geographic Information Systems.*

[10] Fotheringham A S, Yang W, and Kang W (Multiscale geographically weighted regression (MGWR)) 2017 *Ann. Am. Assoc. Geogr.* **107**(6) 1247-1261. https://doi.org/10.1080/24694452.2017.1352480

[11] Thompson R L, Johnson G D, Gordon R, and Nakaya T (Spatially varying associations between community-level predictors and COVID-19 outcomes in NYC) 2024 *Eur. J. Public Health.* **34**(Sup3) ckae144.1425. https://academic.oup.com/eurpub/article/34/Supplement_3/ckae144.1425/7843878

[12] Fotheringham A S, Charlton M E, and Brunsdon C (Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis) 1998 *Environ. Plan. A.* **30**(11) 1905-1927. https://doi.org/10.1068/a301905

[13] Anton H and Rorres C 2013 *Elementary linear algebra: Applications version 11th ed*. (New York: John Wiley & Sons).

[14] Fischer M M and Getis A 2009 *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications* (Heidelberg: Springer Berlin) https://doi.org/10.1007/978-3-642-03647-7

[15] Nakaya T, Fotheringham A S, Brunsdon C, and Charlton M (Geographically weighted Poisson regression for disease association mapping) 2005 *Stat. Med.* **24**(17) 2963-2981. https://doi.org/10.1002/sim.2129

[16] Zhang J, Yang Y, and Ding J (Information criteria for model selection) 2023 *WIREs Comput. Stat.* **15**(5) e1607. https://doi.org/10.1002/wics.1607

[17] Chapra S C and Canale R P 2010 *Numerical methods for engineers* (New York: McGraw-Hill)

[18] Hastie T and Tibshirani R (Generalized additive models) 1986 *Stat. Sci.* **1**(3) 297-318. https://doi.org/10.1214/ss/1177013604