Innovative: International Multi-disciplinary Journal of Applied Technology
(ISSN 2995-486X) VOLUME 03 ISSUE 5, 2025

# A COMPARATIVE STUDY BETWEEN THE SPATIAL LINEAR REGRESSION MODEL AND THE WEIGHTED MODEL: WITH APPLICATION

**Jaufar Mousa Mohammed**
Department of Statistics Sciences, College of Administration & Economic, Kirkuk University, Iraq
Email: jaufar.mohammed@uokirkuk.edu.iq

**Abstract:**
The objective of the paper is to compare the SLRM with the WM and highlights the practical applications of both models in the spatial data analysis. Spatial linear regression is an important statistical method for spatial dependence data analysis, which can be used for spatial pattern detection and variation exploration of the field. In comparison, the weighted model is a way of estimating data by borrowing the weights of the relationships between points, for focused on the level of examining the differences among points (locations). Both models were implemented on spatial datasets to make predictions, this study only aims at comparing the accuracy of prediction for the two models on spatial data using statistical indicators mean absolute error (MAE) and root mean square error (RMSE). The findings suggest that there are pros and cons of both models, and the model choice depends on the properties of the data under study, as well as the intensity of the spatial relationship across observations. The results of this study suggest that the weighted model might be more applicable in regions with great fluctuation and inhomogeneous characteristics between sites, and the SLR model can be more applicable in the study of the continuously correlated spatial relationships.

**Keywords: Spatial Regression Model Geographically Weighted Regression, Spatial weights matrix**

## 1. Introduction

With the great progress of spatial data analysis methods, the study on the geographical interactions on social, economic and environmental media has gradually become one of the key axles for decision making in the era of big data. This is where spatial statistical models are relevant, because they allow the relationship between variables to be related to the geographical dimension, which extends the capability of the usual models that assume independent observations and constant relationships over space [1]. In this aspect the Spatial Linear Regression model or the Geographically Weighted Regression (GWR) model has been assumed to be some of the widely used techniques to address spatial variation. But they are very different in the way they think and are used. Spatial linear regression is in fact a generalisation of traditional linear models, for example how are we able to

incorporate (the idea of) spatial autocorrelation and so forth in slightly more general terms than specifying bins and have a fixed coefficient applied to all geographic units? However, the GWR model is an innovative approach which admits spatial non-stationary, by considering that regression coefficients are changing dynamically over geographical locations, to provide a more accurate estimation of local relationships [2]. Although GWR is superior in addressing spatial complexity, its underperformance under what conditions compared to the spatial linear regression model is questioned, especially in the two extreme cases, data size and distribution heterogeneity. The goal of this work is to perform a systematic comparison to evaluate the advantage of one model over the other, in terms of interpretability, statistical accuracy and computational efficiency, when considering real spatial data drawn from a spatial phenomenon (e.g., pollution or urban planning) [3]. The relevance of this comparison is to inform researchers and practitioners for choosing the appropriate model according to the data and research question, particularly given practical challenges such as the computational complexity of GWR and its difficulty of interpretation relative to the linear model.

### 1.1 **Research Objective**

The purpose of this work is to assess Spatial Linear Regression model and Geographically Weighted Regression (GWR) model on practical application. This involves comparing the prediction ability, examining the spatial interpretation of final outputs, and weighing the strengths and weaknesses of the models. The study also attempts to make guidelines about the choice of appropriate model for various spatial data applications.

### 1.2 **General Spatial Model (GSM)**

The general form of the spatial model includes spatial lag correlation and Auto correlated error, as shown in the following equation:

$$Y = \lambda WY + X\beta + WX\theta + u \quad , \quad |\lambda| < 1 \dots\dots\dots (1)$$
$$u = \rho Wu + e \quad , \quad |\rho| < 1$$

Where:

$Y$ is $(n \times 1)$ vector of dependent variables, $X$: A non-stochastic regression matrix, $W$: An $(n \times n)$ weight matrix, $e/X = i.i.d.N(0, \sigma_{en}^2 I_n)$ A vector of random errors, $\beta, \theta$: $(n \times 1)$ parameter vectors to be estimated, λ, ρ: Spatial regression parameters, $u$: Spatially correlated errors, $e$: A (P×1) vector of random errors [4], [5].

Model (1) represents the spatial lag of the dependent variables from the observations, and also includes the spatial lag of other variables, such as $WX$.

The second part of model (1) represents the spatial model for the random distribution, and it can be written as follows:

$$(I - \lambda W)Y = X\beta + WX\theta + u$$
$$Y = (I - \lambda W)^{-1}[X\beta + WX\theta + u] \dots\dots\dots (2)$$
$$u = (I - \rho W)^{-1}e$$

On the condition that the matrices are invertible, in fact, there are three weight matrices in model (1). Therefore, we can reformulate it as follows:

$$Y = \lambda WY + \psi\delta + u \quad , \quad |\lambda| < 1 \dots\dots\dots (3)$$
$$u = \rho Wu + e \quad , \quad |\rho| < 1$$

Where $\psi = [X\ WX]$ and $\delta = [\beta\ \theta]$. This is referred to as the Spatial Autoregressive Model, encompassing many other spatial econometric models, such as the spatial linear regression model.

### 1.3 The Spatial Linear Model (SLM)

which is formulated from the general model (1) after assuming that $\lambda = \rho = \theta = 0$, that is:

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

"In this model, $X$ is independent, and its parameters can be estimated using the ordinary least squares $(OLS)$ method.

### 1.4 Ordinary Least Square (OLS) for (SLM)

We can estimate the parameter of the spatial linear model by easily applying ordinary least squares (OLS) in (4) as follows:

$$Y = X\beta + e$$

$$e = Y - X\beta$$

$$e'e = (Y' - \beta'X')(Y - X\beta) = Y'Y - 2Y'X\beta + \beta'X'X\beta$$

We differentiate the last expression with respect to $\beta$ and set it equal to zero, resulting in:

$$\Longrightarrow \quad \beta'X'X = Y'X$$

$$\Longrightarrow \quad \hat{\beta} = (X'X)^{-1}Y'X \dots\dots\dots\dots (5)$$

And the variance is given by the following formula:

$$\sigma^2 = \frac{(Y - \hat{\beta}X)'(Y - \hat{\beta}X)}{n - k} \dots\dots\dots\dots\dots\dots (6\ \ )$$

## 2. Materials and Methods

In this study, two statistical modeling approaches were applied to spatial data on mortality across Iraq's governorates: the Spatial Linear Regression Model (SLRM) and the Geographically Weighted Regression (GWR) model. The dataset consisted of the number of deaths per 1,000 inhabitants due to chronic and non-communicable diseases (including cancer, diabetes, and respiratory diseases) for the year 2021, collected from national environmental statistics. The SLRM was estimated using the Ordinary Least Squares (OLS) method, assuming independent error terms and constant coefficients across all locations. Parameters were estimated, and spatial variance was calculated to assess the model fit and predictive ability.

The GWR model, on the other hand, allowed local variations by assigning spatially varying coefficients to each observation point, based on their geographic coordinates. A spatial weights matrix was constructed, considering adjacency and distance relationships between governorates, to capture spatial heterogeneity. Weighted Least Squares estimation was applied, producing localized parameter estimates and variances. To compare the two models, key performance metrics were calculated, including Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), which quantified predictive accuracy. Additionally, spatial patterns in parameter estimates were examined to assess the capacity of each model to capture local versus global spatial relationships.

## 3. Results

### 3.1 Spatial Linear Weighted Model

Also known as the **Geographically Weighted Regression (GWR)** model, it is used to analyze the relationship between a dependent variable and independent variables, allowing each geographic location to have its own regression coefficients [6]. In other words, the coefficients are estimated for each data point based on its geographic location in the data space. This model is written in the following form:

$$Y = X\beta_{(W)} + e \dots\dots\dots\dots (7)$$

Where $X$ is a vector of independent variables, $\beta_{(W)}$ is a vector of coefficients that depend on spatial weights, and $e$ is also a vector of errors.

### 3.2 Least Squares Method for Spatial Linear Weighted Model

In the spatial linear weighted model or the (GWR) model, we need to assign different weights to each spatial point, which leads to the modified form of weighted least squares [7]. Therefore, the derivation of the model (7) is as follows:

$$e = W(Y - \beta X)$$

$$e'e = (Y - \beta X)'W(Y - \beta X) = Y'YW - 2Y'WX\beta + X\beta WX'\beta$$

We differentiate the last expression with respect to β and set it equal to zero, resulting in

$$X\beta WX' = Y'WX$$

$$\hat{\beta}_{(W)} = (XWX')^{-1}Y'WX \dots\dots\dots\dots (8)$$

The variance of the estimated coefficients is given by:

$$\sigma^2\left(\hat{\beta}_{(W)}\right) = \sigma^2(XWX')^{-1} \dots\dots\dots\dots\dots (9)$$

Where $\hat{\beta}_{(W)}$ are the estimated weighted coefficients at the location $(u, v)$. As for $\sigma^2$, it is calculated from :

$$\sigma^2 = \frac{(Y - \hat{\beta}_{(W)}X)'(Y - \hat{\beta}_{(W)}X)}{n - k} \dots\dots\dots\dots\dots (10)$$

### 3.3 Spatial Weights Matrix

This is a square $(n \times n)$ matrix containing positive values, and it is not necessarily symmetric. It is constructed based on the adjacency relationships between observations. Each location is linked to other locations within the same row of the matrix, while the diagonal elements of the matrix are equal to zero [8]. The selection of the spatial weights matrix is critical in determining spatial effects, so a suitable weights matrix must be constructed. There are several methods to design this matrix $W$.

### 3.4 Binary Contiguity Weights Matrix

It is a square $(n \times n)$ matrix defined such that if $i$ and $j$ are adjacent, then $w_{ij} = 1$, and if $i$ and $j$ are not adjacent, then $w_{ij} = 0$, as shown in the following formula:

$$W = \left(w_{ij}\right)_{n\times n} = \begin{bmatrix} 1 & \text{if i neighbor j} \\ 0 & \text{otherwise} \end{bmatrix} \dots\dots\dots (11)$$

### 3.5 Row - Standardized Weights Matrix

This matrix is sometimes called the modified matrix, where the sum of each row is equal to one. Its calculation relies on the binary adjacency weight matrix, as shown in the following formula:

$$W^{std} = \left(w_{ij}\right) = \begin{bmatrix} \frac{w_{ij}}{\sum_i w_{ij}} & \text{if i neighbor j} \quad 0 < W^{std} \leq 1 \\ 0 & \text{otherwise} \end{bmatrix} \dots\dots\dots\dots (12)$$

### 3.6 Comparison Criteria

The model selection is a crucial task in Data Analysis, since it provides the best model among a number of candidate models. The Root-Mean-Square-Error ($RMSE$) criterion is applied in this context, and is defined as the root of the sum of squares divided by where is the estimated response of and is the true output of the regression $(n - k - 1)$. This criteria is derived on all the models and the model with minimum value of$RMSE$ is selected as the best [9], [11], [12]. This criterion is used in the present work, and its general form is given by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n - k - 1}} \dots\dots\dots\dots\dots\dots (13)$$

## 4. Discussion

### 4.1 Application

This study includes data on the number of deaths caused by certain chronic and non-communicable diseases across various governorates in Iraq for the year 2021, per 1,000 inhabitants. These data were collected from the Sustainable Development Goals (SDG) indicators with an environmental dimension, according to the 2022 Environmental Statistics. Where $(Y)$ represents the total number of deaths due to the mentioned chronic diseases, which is the dependent variable [13]. The independent variables are: $X_1$ The number of deaths caused by diabetes, $X_2$ The number of deaths caused by cancer (malignant tumors) ,$X_3$ The number of deaths caused by chronic respiratory diseases. As shown in Table (1), these data are considered spatial data consisting of 15 governorates (locations), each with coordinates $u(x)$ representing east-west and $v(x)$ representing north-south. The natural logarithm of these data was taken to ensure they follow a normal distribution.

**Table 1. Number of Deaths Due to Cancer, Diabetes, and Respiratory Diseases (2021)**

| Governorates (locutions) | $Y$ | $X_1$ | $X_2$ | $X_3$ | $v(x)$ | $u(x)$ |
|---|---|---|---|---|---|---|
| **1** | 3525 | 201 | 596 | 201 | 62 | 30 |
| **2** | 1371 | 25 | 333 | 15 | 54 | 38 |
| **3** | 1857 | 312 | 300 | 57 | 40 | 40 |
| **4** | 1208 | 148 | 203 | 19 | 38 | 30 |
| **5** | 12285 | 2286 | 2277 | 545 | 39 | 39 |
| **6** | 1745 | 299 | 381 | 75 | 28 | 37 |
| **7** | 1780 | 232 | 304 | 483 | 30 | 35 |
| **8** | 1250 | 158 | 256 | 95 | 30 | 48 |
| **9** | 2389 | 65 | 187 | 71 | 47 | 33 |
| **10** | 1201 | 251 | 352 | 582 | 25 | 36 |
| **11** | 882 | 169 | 216 | 40 | 27 | 42 |
| **12** | 2601 | 80 | 122 | 82 | 20 | 43 |
| **13** | 1037 | 182 | 241 | 774 | 20 | 50 |
| **14** | 2746 | 102 | 222 | 40 | 25 | 55 |
| **15** | 37215 | 167 | 749 | 78 | 15 | 60 |

### 4.2 Estimation of the Spatial Linear Model

After assuming $\lambda = \rho = \theta = 0$, the parameters $\beta$ of the model (4) were estimated using the least squares method with the formula (5). The spatial variance was then calculated using formula (6), and the results are shown in Table (2) below.

**Table 2. Estimate parameter spatial Linear model**

| $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\sigma^2$ | $RMSE$ |
|---|---|---|---|---|---|
| 2.0635 | -0.2282 | 1.2375 | -0.0640 | 0.5815 | 0.2299 |

That is, the spatial linear model is in the following form:

$$Y = 2.0635 - 0.2282X_1 + 1.2375X_2 - 0.064X_3$$

### 4.3 Estimation of the Weighted Spatial Linear Model Parameters

As mentioned earlier, in this model, each location is assigned specific coefficients, and there is a variance for each location [14], [15]. Therefore, by using formula (8) after finding the weights matrix

between each location to be estimated and the other locations, this process is repeated for all locations in the spatial data. As a result, we obtained 15 estimation equations, and the results are shown in Table 3 below.

**Table 3. Estimate parameter GWR model**

| Locations | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\sigma^2$ | $RMSE$ |
|---|---|---|---|---|---|---|
| **1** | 1.4533 | -0.8315 | 1.9409 | -0.0938 | 0.8645 | 0.2803 |
| **2** | 2.1198 | -1.4882 | 2.4522 | -0.1214 | 1.6707 | 0.3897 |
| **3** | 0.8695 | -0.1892 | 1.5309 | -0.1757 | 0.6784 | 0.2483 |
| **4** | 1.3159 | -0.2149 | 1.4774 | -0.1792 | 0.6560 | 0.2442 |
| **5** | 0.4545 | -0.0894 | 1.5325 | -0.1935 | 0.7248 | 0.2567 |
| **6** | 1.3946 | -0.0081 | 1.2185 | -0.1106 | 0.6416 | 0.2415 |
| **7** | 1.2729 | -0.0333 | 1.2810 | -0.1316 | 0.6464 | 0.2424 |
| **8** | 1.9282 | -0.0706 | 1.1199 | -0.0577 | 0.5980 | 0.2332 |
| **9** | 1.5811 | -0.6989 | 1.8308 | -0.1344 | 0.7965 | 0.2691 |
| **10** | 1.4487 | -0.0559 | 1.2010 | -0.0560 | 0.6224 | 0.2379 |
| **11** | 1.6635 | -0.0095 | 1.1444 | -0.0877 | 0.6180 | 0.2370 |
| **12** | 1.4821 | -0.0602 | 1.1530 | -0.0327 | 0.6138 | 0.2362 |
| **13** | 2.4177 | -0.0493 | 0.9207 | 0.0419 | 0.6419 | 0.2416 |
| **14** | 2.4825 | -0.0531 | 0.9307 | 0.0147 | 0.6344 | 0.2402 |
| **15** | 3.3427 | 0.0378 | 0.6755 | 0.0342 | 0.7185 | 0.2556 |

## 5. Conclusion.

**1)** The model indicates that deaths due to malignant tumors are the most influential variable in the total number of deaths.

**2)** There are negative effects of deaths due to diabetes and respiratory diseases, which calls for further analysis to understand the true relationship between these variables and overall mortality.

**3)** The variance and root mean square error suggest that the model provides acceptable predictions but may need improvement by adding additional spatial or demographic variables.

**4)** There is clear variation in the impact of variables across locations, indicating the importance of using the GWR model rather than a general linear model.

**5)** The negative and positive effects of diabetes and respiratory disease deaths vary by location, suggesting that regional environmental and health factors may play a role in explaining these differences.

**6)** Some regions (such as 1 and 2) have high RMSE values and variance, indicating that the model may be less accurate in those areas and may require additional variables to improve accuracy.

**7)** The GWR model shows significant spatial variation in the impact of variables on mortality, making it more accurate in interpretation and analysis compared to the general linear model.

**Recommendations**

1) Rely on GWR when there is a need to understand local effects, especially when there is clear variation in parameters between regions.
2) Improve model accuracy by using additional variables. Some locations in GWR have high RMSE, indicating that the model does not explain all factors affecting mortality.
3) It is recommended to add variables such as population density, healthcare level in each governorate, environmental pollution, local climate, and the prevalence rate of other chronic diseases.

4) If the goal is general analysis and identifying overall trends, the spatial linear model is preferred because it provides a single equation that interprets the relationship broadly.
5) If the goal is to make local decisions for each governorate, GWR is preferred as it shows how the effect of each variable differs across regions.
6) Due to the varying impact of variables across regions in GWR, health policies should be tailored to each governorate. For example, governorates where the effect of diabetes mortality is negative need to improve prevention and early treatment programs for diabetes. Governorates where the impact of malignant tumors is high should focus on improving early detection and cancer treatment programs. Governorates with varying effects from respiratory diseases should assess environmental factors such as pollution.
7) The two models can be compared with other spatial models, such as the Spatial Autoregressive Model (SAR) to study the effect of neighboring spatial relationships, and the Spatial Error Model (SEM) to analyze the effects of unobserved factors. Hybrid models combining GWR and machine learning techniques could also improve prediction accuracy.
8) The linear model showed a general fixed effect, but GWR revealed spatial differences, indicating that diseases may interact differently in each governorate. It is recommended to study the interactions between different chronic diseases and understand how they impact the overall mortality rate.
9) Adding new spatial variables and using more advanced spatial models could improve the accuracy of results.
10) Health policy recommendations should be based on local results for each governorate rather than using a uniform strategy for all regions.

**References**


[1] G. Arbia, A Primer for Spatial Econometrics with Applications in R. 2014.
[2] S. J. Cook, J. C. Hays, and R. J. Franzese, "Model Specification and Spatial Interdependence," in APSA Conference Paper, Sep. 2015.
[3] D. B. Omar and A. F. Tawfeeq, "Estimating Parameters of Some Spatial Regression Models With Experimental and Applied Study," Department of Statistics, College of Administration & Economics, University of Kirkuk, 2020.
[4] D. B. Omar, "Application of Spatial Error Model For Ordinary and Fuzzy Data With Comparison," Journal of Positive Sciences, vol. 2023, no. 21, 2023, [Online]. Available: dalia.badee@uokirkuk.edu.iq.
[5] M. M. Jaufar, M. M. Aziz, and A. F. Tawfeeq, "Fuzzy Estimations in Spatial Statistics," College of Computer Science and Mathematics, Department of Mathematics, University of Mosul, 2022.
[6] B. M. Kazar and M. Celik, Spatial Autoregression (SAR) Model Parameter Estimation Techniques. Springer Science & Business Media, 2012.
[7] A. Sulekan and S. S. S. Jamaludin, "Review on Geographically Weighted Regression (GWR) Approach in Spatial Analysis," Malays J. Fundam. Appl. Sci., vol. 16, no. 2, pp. 173–177, 2020.
[8] S. M. Soemartojo, R. D. Ghaisani, T. Siswantining, M. R. Shahab, and M. Ariyanto, "Parameter Estimation of Geographically Weighted Regression (GWR) Model Using Weighted Least Square and Its Application," in AIP Conf. Proc., vol. 2014, no. 1, Sep. 2018.
[9] T. E. Smith, "Spatial Data Analysis," University of Pennsylvania, School of Engineering and Applied Science, 2015, [Online]. Available: tesmith@seas.upenn.edu.

[10] Q. Zhou, C. Wang, and S. Fang, "Application of Geographically Weighted Regression (GWR) in the Analysis of the Cause of Haze Pollution in China," Atmos. Pollut. Res., vol. 10, no. 3, pp. 835–846, 2019.
[11] L. Anselin, "Spatial Econometrics: Methods and Models," Kluwer Academic Publishers, 1988.
[12] R. Bivand, E. Pebesma, and V. Gómez-Rubio, Applied Spatial Data Analysis with R, 2nd ed. Springer, 2013.
[13] H. LeSage and R. K. Pace, Introduction to Spatial Econometrics. CRC Press, 2009.
[14] M. J. Bowman and S. M. Levitin, "Spatial Analysis of Socioeconomic Data," Socioeconomic Planning Sciences, vol. 32, no. 4, pp. 285–298, 1998.
[15] A. Getis and J. K. Ord, "The Analysis of Spatial Association by Use of Distance Statistics," Geographical Analysis, vol. 24, no. 3, pp. 189–206, 1992.