Parameter: Journal of Statistics
Available at https://bestjournal.untad.ac.id/index.php/parameter
Vol. 5 No. 2 2025, 89 - 97
DOI: *https://doi.org/10.22487/27765660.2025.v5.i2.17900*

eISSN: 2776-5660

WATER QUALITY ANALYSIS IN THE RESIDENTIAL AREAS OF PONTIANAK CITY USING THE GEOGRAPHICALLY WEIGHTED LOGISTIC REGRESSION METHOD

Naomi Nessyana Debataraja[1], Dadan Kusnandar[2], Fika Dian Lestari[3], Joannes Fregis Philosovio Anugrahnu[4]

[123]*Departement of Statistics, Universitas Tanjungpura, Pontianak, Indonesia*

***e-mail**: [1]naominessyana@math.untan.ac.id, [2]dkusnand@ac.id, [3]fkdnlstr@gmail.com, [4]joannesfregis@student.untan.ac.id*

ABSTRACT

Water quality is a key indicator of a community's health and welfare, yet it has deteriorated significantly due to pollution caused by human activities. This study aimed to evaluate Geographically Weighted Logistic Regression's (GWLR) ability to handle spatial nonstationarity in the relationship between explanatory factors and water quality status in Pontianak City, and to compare its performance with logistic regression. Three modelling approaches were applied to classify water as polluted or non-polluted: (i) logistic regression with spatially invariant) parameters; (ii) GWLR with a fixed Gaussian kernel, producing spatially varying parameters using a fixed bandwidth; and (iii) GWLR with an adaptive Gaussian kernel, producing spatially varying parameters using an adaptive bandwidth. Model performance was compared using Akaike's Information Criterion (AIC) and classification accuracy. The GWLR model with a fixed Gaussian kernel produced an AIC of 22.52, whereas the logistic regression model produced a slightly lower AIC of 22.39; both models achieved a classification accuracy of 92.86%, with the adaptive-kernel GWLR showing comparable classification performance. These results indicate that, for the parameter settings considered, GWLR offered performance comparable to, but not substantially better than logistic regression for modelling the factors affecting water quality, despite its capacity to address spatial nonstationarity.

Keywords: *Gaussian kernel, GWLR, Spatial, Water pollution*

Cited: Debataraja, N. N., Kusnandar, D., Lestari, F. D., & Anugrahnu, J. F. P., (2025). Water Quality Analysis in the Residential Areas of Pontianak City Using the Geographically Weighted Logistic Regression Method. *Parameter: Journal of Statistics*, 5(2), 89–97. https://doi.org/10.22487/27765660.2025.v5.i2.17900

*Received: Nov 2025; Reviewed: Nov 2025; Published: Dec 2025

INTRODUCTION

Water quality is a critical environmental and public health issue, particularly in rapidly growing cities such as Pontianak. Reports from local environmental agencies have revealed rising levels of water pollution in rivers and shallow groundwater, caused by domestic waste, industrial effluents and inadequate sanitation infrastructure (Panggabean & Debataraja, 2025). These conditions are of particular concern because most Pontianak residents rely on these water sources for daily use, which makes identifying pollution determinants an urgent priority.

In many environmental studies, including water quality assessments, data often includes information on the geographical location where measurements were taken, known as spatial data (Debataraja & Kusnandar, 2023). Spatial data provides essential information on spatial patterns, gradients and regional variations influencing environmental phenomena (Hosmer et al., 2013). Logistic regression is a commonly used statistical method for analyzing the relationship between a dichotomous dependent variable such as polluted versus non-polluted water and one or more independent variables. However, traditional logistic regression assumes that the relationship between predictors and the response is constant across all locations. Environmental processes often vary spatially, and this spatial heterogeneity can cause models to misrepresent local conditions.

To address this limitation, Geographically Weighted Logistic Regression (GWLR) was developed as a spatial extension of logistic regression. GWLR incorporates location-specific parameter estimation by applying a geographical weighting function that assigns weights to nearby observations (Omrani et al., 2025). Unlike logistic regression, which assumes uniform coefficients, GWLR allows coefficients to vary across space, capturing spatial non-stationarity in the relationships between variables. The choice of weighting kernel also affects performance: the fixed Gaussian kernel uses a constant bandwidth across the study area, whereas the adaptive Gaussian kernel adjusts the bandwidth according to data density, offering greater flexibility in areas with sparse or clustered observations.

Previous studies on water quality in Pontianak have employed a variety of spatial and statistical methods, such as discriminant analysis, spatial mapping and regression-based approaches (Debataraja et al., 2019; Debataraja & Kusnandar, 2023; Kusnandar et al., 2019, 2020, 2022; Kusnandar, Debataraja, & Fitriani, 2021; Kusnandar, Debataraja, & Utari, 2021). While these studies emphasise the importance of spatial techniques for interpreting environmental data, they typically rely on models or descriptive spatial analysis. They do not explicitly test whether allowing regression coefficients to vary across space improves model performance compared to standard logistic regression. Furthermore, to the best of our knowledge, no previous study on Pontianak’s water quality has systematically evaluated geographically weighted logistic regression (GWLR) using different kernel specifications. In particular, the comparative performance of GWLR with a fixed (permanent) Gaussian kernel and GWLR with an adaptive Gaussian kernel has not been examined, even though these two kernels represent distinct strategies for handling spatial heterogeneity in areas with varying sampling densities. The present study addresses this methodological gap by applying and comparing logistic regression, GWLR with a fixed Gaussian kernel and GWLR with an adaptive Gaussian kernel, to identify the most appropriate modelling approach for analyzing the factors affecting water quality in Pontianak City.

In view of the escalating urgency of water contamination and the existence of spatial heterogeneity across Pontianak, a more comprehensive modelling approach is imperative. The objective of this study is to make a comparison between the performance of Logistic Regression (GWLR) and two GWLR specifications – GWLR with a fixed Gaussian kernel and GWLR with an adaptive Gaussian kernel – in order to ascertain which model is more effective in identifying the factors influencing water quality in Pontianak City. By explicitly evaluating the advantages and limitations of spatially varying parameter models, this study contributes to improving local environmental assessment, policymaking, and public health protection.

MATERIALS AND METHODS

This study used water sample data collected from 42 locations in Pontianak City using a stratified random sampling method (Debataraja, et.al., 2018). This sampling was used to ensure that water samples represented the spatial and environmental across Pontianak City, where water quality conditions differ by area. Dividing the study area into strata and randomly selecting locations within each stratum reduces sampling bias and prevents certain zones from being over or underrepresented. The dependent variable, Y, was the water quality status, which was defined using the Pollution Index (PI), as calculated in accordance with Regulation No. 32 of 2017 of the Republic of Indonesia's Minister of Health (Kusnandar et al., 2022). The PI was computed from water colour, total dissolved solids and turbidity, following the method described by (Kusnandar et al., 2019). The resulting PI values were classified as meeting or not meeting quality standards (i.e. not polluted or lightly polluted). In this study, the PI class meeting quality standards consisted of six location points, while the PI class not meeting quality standards consisted of 36 location points. These two classes formed the response variable Y, with $Y = 0$ for water meeting quality standards and $Y = 1$ for lightly polluted water. The explanatory variables X used in the modelling were dissolved oxygen (DO) as X_1, iron (Fe) content as X_2, and chemical oxygen demand (COD) as X_3. First, Logistic Regression model was fitted to examine the relationship between $Y$and (X_1, X_2, X_3), with the probability of a sample being lightly polluted modelled as (Backhaus et al., 2023; Hosmer et al., 2013)

$$\pi(x) = \frac{exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3)}{1 + exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3)} \quad (1)$$

To account for spatial heterogeneity, Geographically Weighted Logistic Regression (GWLR) was then applied as an extension of logistic regression that incorporates geographical coordinates into the model (Isazade et al., 2023). In GWLR, the longitude and latitude $(u_i, v_i)$of each sampling location were used to compute Euclidean distances between sites, and the regression coefficients were allowed to vary by location (Lessani & Li, 2024):

$$\pi(x_i) = \frac{exp\ \left(\beta_0(u_i, v_i) + \sum_{j=1}^{p} \beta_j(u_i, v_i) x_{ji}\right)}{1 + exp\left(\beta_0(u_i, v_i) + \sum_{j=0}^{p} \beta_j(u_i, v_i) x_{ji}\right)} \quad (2)$$

with the corresponding logit form (Lessani & Li, 2024):

$$g(x_i) = ln\ \left[\frac{\pi(x_i)}{1 - \pi(x_i)}\right] = \beta_0(u_i, v_i) + \sum_{j=1}^{p} \beta_j(u_i, v_i) x_{ji} \quad (3)$$

The GWLR parameters were estimated using maximum likelihood estimation via an iterative procedure (Fikri et al., 2019). A key component of GWLR is the spatial weighting scheme. In this study, this was based on the Gaussian kernel. Two types of Gaussian kernel weighting were used: First, a fixed Gaussian kernel, where a single bandwith is applied uniformly across the study area and the weights are constant and the second, an adaptive Gaussian kernel, where the bandwith $b_i$varies by location to adapt to local data density, with weights $w_{ik} = \exp\ (-d_{ik}^2/2b_i^2)$.

The optimal bandwidth values were obtained for both kernel types using the Cross-Validation (CV) method. The overall procedure in this study consisted of the following steps: Computing the Pollution Index and classifying samples into polluted and non-polluted categories (defining Y); assembling DO, Fe and COD as explanatory variables (defining X); fitting the Logistic Regression model; constructing the spatial dataset using the coordinates of each sample point; calculating Euclidean distances. Selecting bandwidths via CV; computing spatial weights with fixed and adaptive Gaussian kernels; estimating GWLR parameters; and testing the significance of the model coefficients. The performance of the logistic regression model and the two GWLR models (with fixed and adaptive Gaussian kernels) was then evaluated and compared using Akaike's information criterion (AIC), where lower values indicate a better model fit. Classification accuracy was also evaluated, defined as the percentage of observations correctly classified as polluted or non-polluted.

RESULTS AND DISCUSSION

The spatial distribution of the 42 water sampling locations in Pontianak City is shown in Figure 1. Water samples were collected in 2018 using a stratified random sampling design, and the pollution index (PI) at each site was calculated based on the Regulation of the Minister of Health of the Republic of Indonesia No. 32 of 2017. The numbers displayed inside each point in Figure 1 represent the sampling location IDs, while the point colors indicate the PI class: locations meeting the water quality standard (not polluted, n = 6) and locations not meeting the standard (lightly polluted, n = 36). The spatial data were processed and mapped using QGIS, and all coordinates reference system to ensure consistency in distance and area calculations for subsequent spatial modelling. A north arrow and a scale bar in kilometers are included in the map to indicate orientation and spatial extent relevant to the GWLR analysis.

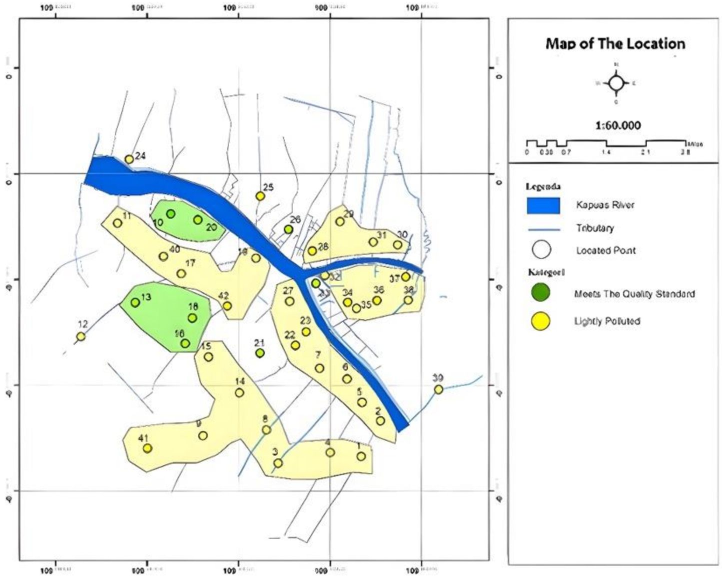

Figure 1. Pollution Index Value Classification Map

A logistic regression model was fitted with the R statistical package. Simultaneous testing in logistic regression uses the statistical significance of the *G* test. Based on the results of the analysis, it is known that the statistical value of the *G* test of 48.17552. It is greater than the value of $\chi^2_{(0.1;3)}$ = 6.251. This shows that at least one independent variable significantly influences the dependent variable. The results of partial testing are conducted to examine which of the independent variables significantly affect the dependent variable (Table 1).

Table 1. Estimation of the Logistic Regression Model Parameters

Parameter	**Estimate**	***Standard Error***	**Z**
β_0	-7.590	3.782	-2.006
β_1	-0.526	0.720	-0.731
β_2	2.861	1.255	2.279

Based on the results listed in Table 1, the logistic regression model for the factors affecting water quality in Pontianak is:

$$\pi(x) = \frac{exp\ (-7.590 - 0.526x_1 + 2.861x_2)}{1 + exp\ (-7.590 - 0.526x_1 + 2.861x_2)}$$

where x_1 denotes dissolved oxygen (DO) and x_2 denotes iron (Fe). The negative intercept -7.590 indicates that, when dissolved oxygen (DO) and iron (Fe) are at their reference levels, the baseline probability of a location being lightly polluted is very low. The coefficient for DO ($\beta_1 = -0.526$) implies that, holding Fe constant, a one-unit increase in DO decreases the log-odds of the water being lightly polluted by 0.526, corresponding to an odds ratio of exp $(-0.526) \approx 0.59$; in other words, each unit increase in DO reduces the odds of light pollution to about 59% of their previous value (a decrease of roughly 41%). In contrast, the Fe coefficient ($\beta_2 = 2.861$) shows that, for a one-unit increase in Fe with DO held constant, the log-odds of light pollution increase by 2.861, corresponding to an odds ratio of exp $(2.861) \approx 17.48$. This means that each unit increase in Fe multiplies the odds of light pollution by about 17.5, indicating that Fe has a much stronger influence on the classification of water quality than DO in this model. Meanwhile, the logit transformation of the logistic regression model is as follows:

$$g(x) = -7.590 - 0.526x_1 + 2.861x_2$$

The process of testing the parameters of the GWLR model with Fixed Gaussian Kernel weighting is used to determine the factors that influence the quality of river water in Pontianak City. Based on the results of the partial test analysis at the first location, the following results are obtained in Table 2.

Table 2. Testing Process Parameters of the GWLR Model with Fixed Gaussian Kernel Weighting on location (u_1, v_1)

Parameter	Estimate	*Standard Error*	*W*
β_0	-7.633	7.559	-1.009
β_1	-0.489	0.790	-0.619
β_2	2.835	0.945	3.000
β_3	0.092	0.017	5.475

Based on Table 2, the GWLR model with Fixed Gaussian Kernel weighting for the location (u_1, v_1) is as follows (Pardoe, 2021):

$$\pi(x) = \frac{exp\ (-7.633 + 2.835x_2 + 0.092x_3)}{1 + exp\ (-7.633 + 2.835x_2 + 0.092x_3)}$$

Meanwhile, the logit transformation model of the Fixed Gaussian Kernel weighting model is:

$$g(x) = -7.633 + 2.835x_2 + 0.092x_3$$

This local GWLR model for location (u_1, v_1)can be interpreted as follows. The negative intercept (-7.633) indicates that, when Fe and COD at this location are at their reference (baseline) levels, the underlying probability of the water being lightly polluted is very low. The coefficient for iron (2.835) shows that, at location $(u_1, v_1)$and holding COD constant, a one-unit increase in Fe increases the log-odds of the water being lightly polluted by 2.835, which corresponds to an odds ratio of exp $(2.835) \approx 17.03$. In other words, each unit increase in Fe at this location multiplies the odds of light pollution by about 17 times. Similarly, the COD coefficient (0.092) indicates that, with Fe held constant, a one-unit increase in COD increases the log-odds of light pollution by 0.092, corresponding to an odds ratio of exp $(0.092) \approx 1.10$; thus, the odds of light pollution increase by around 9–10% for each unit increase in COD at this site. These local parameter estimates suggest that, for the sampling location (u_1, v_1), Fe has a much stronger effect on the probability of light pollution than COD in the GWLR model.

Table 3. GWLR Logit Function at 42 Location Points with Fixed Gaussian Kernel Weighting

Locs	Logit Function of GWLR	Locs	Logit Function of GWLR
1	$g(x) = -7.633 + 2.835x_2 + 0.092x_3$	22	$g(x) = -7.579 + 2.847x_2 + 0.092x_3$
2	$g(x) = -7.622 + 2.843x_2 + 0.092x_3$	23	$g(x) = -7.577 + 2.849x_2 + 0.092x_3$
3	$g(x) = -7.624 + 2.824x_2 + 0.092x_3$	24	$g(x) = -7.506 + 2.849x_2 + 0.091x_3$
4	$g(x) = -7.665 + 2.868x_2 + 0.093x_3$	25	$g(x) = -7.519 + 2.867x_2 + 0.092x_3$
5	$g(x) = -7.613 + 2.846x_2 + 0.092x_3$	26	$g(x) = -7.537 + 2.864x_2 + 0.092x_3$
6	$g(x) = -7.602 + 2.847x_2 + 0.092x_3$	27	$g(x) = -7.493 + 2.884x_2 + 0.092x_3$
7	$g(x) = -7.594 + 2.847x_2 + 0.092x_3$	28	$g(x) = -7.548 + 2.863x_2 + 0.099x_3$
8	$g(x) = -7.606 + 2.832x_2 + 0.092x_3$	29	$g(x) = -7.545 + 2.869x_2 + 0.092x_3$
9	$g(x) = -7.595 + 2.826x_2 + 0.092x_3$	30	$g(x) = -7.557 + 2.874x_2 + 0.092x_3$
10	$g(x) = -7.586 + 2.821x_2 + 0.092x_3$	31	$g(x) = -7.555 + 2.872x_2 + 0.092x_3$
11	$g(x) = -7.505 + 2.848x_2 + 0.091x_3$	32	$g(x) = -7.555 + 2.866x_2 + 0.092x_3$
12	$g(x) = -7.546 + 2.824x_2 + 0.091x_3$	33	$g(x) = -7.555 + 2.858x_2 + 0.092x_3$
13	$g(x) = -7.541 + 2.837x_2 + 0.092x_3$	34	$g(x) = -7.575 + 2.857x_2 + 0.092x_3$
14	$g(x) = -7.591 + 2.831x_2 + 0.092x_3$	35	$g(x) = -7.579 + 2.856x_2 + 0.092x_3$
15	$g(x) = -7.572 + 2.836x_2 + 0.092x_3$	36	$g(x) = -7.580 + 2.859_2 + 0.092x_3$
16	$g(x) = -7.558 + 2.838x_2 + 0.092x_3$	37	$g(x) = -7.576 + 2.859x_2 + 0.092x_3$
17	$g(x) = -7.535 + 2.846x_2 + 0.092x_3$	38	$g(x) = -7.589 + 2.862x_2 + 0.092x_3$
18	$g(x) = -7.552 + 2.843x_2 + 0.092x_3$	39	$g(x) = -7.627 + 2.849x_2 + 0.092x_3$
19	$g(x) = -7.539 + 2.857x_2 + 0.092x_3$	40	$g(x) = -7.526 + 2.846x_2 + 0.092x_3$
20	$g(x) = -7.520 + 2.857x_2 + 0.092x_3$	41	$g(x) = -7.596 + 2.816x_2 + 0.092x_3$
21	$g(x) = -7.562 + 2.849x_2 + 0.092x_3$	42	$g(x) = -7.548 + 2.85x_2 + 0.091x_3$

Table 3 shows that, for all 42 locations, the local GWLR logit functions with fixed Gaussian kernel weighting have broadly similar coefficient patterns, but with small spatial variations in both the intercept and the slopes for $x_2$and x_3. In all locations, the coefficient of iron (x_2) is positive and relatively large (approximately 2.82–2.88), indicating that higher Fe content consistently increases the log-odds of the water being lightly polluted. This corresponds to an odds ratio of about exp $(2.84) \approx 17$, meaning that, at any given location, a one-unit increase in Fe multiplies the odds of light pollution by roughly 17 times, holding COD constant. The coefficient of COD (x_3) is also positive but much smaller in magnitude (around 0.091–0.099), implying that a one-unit increase in COD increases the odds of light pollution by about 9–10% $(\exp(0.092) \approx 1.10)$, conditional on Fe. The intercepts, which range roughly from $-7.66$to -7.50, indicate that, when both Fe and COD are at their reference levels, the baseline probability of light pollution remains low across all locations. Overall, these local parameter estimates suggest that, under the fixed Gaussian kernel GWLR model, iron content (x_2) has a much stronger influence on water quality status than COD (x_3), while the small differences in coefficients between locations reflect the presence of mild spatial nonstationarity in the effects of these variables.

The process of testing the parameters of the GWLR model with the Adaptive Gaussian Kernel weighting is used to determine the factors that influence the quality of river water in Pontianak City. Based on the results of the partial test analysis at the first location, the following results are presented in Table 4.

Table 4. Testing the Parameters of the GWLR Model with Adaptive Gaussian Kernel Weighting on location (u_1, v_1)

Parameter	**Estimate**	***Standard Error***	***W***
β_0	-7.694	7.559	-1.018
β_2	-0.439	0.790	-0.556
β_3	2.800	0.945	2.964

Based on Table 4, the GWLR model it is that obtained with the Adaptive Gaussian Kernel weighting for the location (u_1, v_1) is as follows:

$$\pi(x) = \frac{exp\ (-7.694 - 0.439x_2 + 2.800x_3)}{1 + exp\ (-7.694 - 0.439x_2 + 2.800x_3)}$$

Meanwhile, the logit transformation model of the Adaptive Gaussian Kernel weighting is:

$$g(x) = -7.694 - 0.493x_2 + 2.800x_3$$

At location (u_1, v_1), the negative intercept (-7.694) indicates that, when DO and Fe are at their reference levels, the baseline probability of the water being lightly polluted is very low. The coefficient for DO $(\beta_1 = -0.439)$ shows that, holding Fe constant, a one-unit increase in dissolved oxygen decreases the log-odds of light pollution by 0.439, corresponding to an odds ratio of $\exp(-0.439) \approx 0.64$. In other words, each unit increase in DO reduces the odds of light pollution to about 64% of their previous value (a decrease of roughly 36%). Conversely, the coefficient for Fe $(\beta_2 = 2.800)$ indicates that, for a one-unit increase in iron content with DO held constant, the log-odds of light pollution increase by 2.800, corresponding to an odds ratio of $\exp(2.800) \approx 16.45$. Thus, each unit increase in Fe multiplies the odds of light pollution by about 16.5, suggesting that iron has a much stronger impact on water quality status than dissolved oxygen at this location in the adaptive-kernel GWLR model.

The process of testing these parameters is repeated at each observation location or the observation location(u_{42}, v_{42}). Variables that significant are those with $|W| \geq Z_{0,05} = 1.64$. The GWLR logit weighting function of the Adaptive Gaussian Kernel for 42 observation locations is presented in Table 5.

Table 5. GWLR Logit Function at 42 Location Points with Adaptive Gaussian Kernel Weighting

Locs	**Logit Function of GWLR**	**Locs**	**Logit Function of GWLR**
1	$g(x) = -7.694 + 2.800x_2 + 0.093x_3$	22	$g(x) = -7.544 + 2.870x_2 + 0.091x_3$
2	$g(x) = -7.678 + 2.818x_2 + 0.092x_3$	23	$g(x) = -7.534 + 2.807x_2 + 0.091x_3$
3	$g(x) = -7{,}704 + 2.744x_2 + 0.092x_3$	24	$g(x) = -7.387 + 2.833x_2 + 0.090x_3$
4	$g(x) = -7.659 + 2.867x_2 + 0.093x_3$	25	$g(x) = -7.334 + 2.884x_2 + 0.090x_3$
5	$g(x) = -7.662 + 2.817x_2 + 0.092x_3$	26	$g(x) = -7.334 + 2.882x_2 + 0.090x_3$
6	$g(x) = -7{,}636 + 2.813x_2 + 0.092x_3$	27	$g(x) = -7.369 + 2.915x_2 + 0.091x_3$
7	$g(x) = -7.613 + 2.794x_2 + 0.091x_3$	28	$g(x) = -7.373 + 2.882x_2 + 0.090x_3$
8	$g(x) = -7.668 + 2.741x_2 + 0.091x_3$	29	$g(x) = -7.411 + 2.904x_2 + 0.091x_3$
9	$g(x) = -7.647 + 2.679x_2 + 0.091x_3$	30	$g(x) = -7.483 + 2.913x_2 + 0.091x_3$
10	$g(x) = -7.594 + 2.707x_2 + 0.090x_3$	31	$g(x) = -7.465 + 2.909x_2 + 0.092x_3$
11	$g(x) = -7.393 + 2.832x_2 + 0.090x_3$	32	$g(x) = -7.512 + 2.889x_2 + 0.092x_3$
12	$g(x) = -7.486 + 2.768x_2 + 0.090x_3$	33	$g(x) = -7.445 + 2.858x_2 + 0.091x_3$

Locs	Logit Function of GWLR	Locs	Logit Function of GWLR
13	$g(x) = -7{,}415 + 2.768x_2 + 0.089x_3$	34	$g(x) = -7.526 + 2.849x_2 + 0.091x_3$
14	$g(x) = -7{,}635 + 2{,}672x_2 + 0.090x_3$	35	$g(x) = -7.545 + 2.847x_2 + 0.091x_3$
15	$g(x) = -7{,}514 + 2{,}704x_2 + 0.089x_3$	36	$g(x) = -7.554 + 2.864x_2 + 0.092x_3$
16	$g(x) = -7{,}443 + 2{,}737x_2 + 0.089x_3$	37	$g(x) = -7.544 + 2.898x_2 + 0.092x_3$
17	$g(x) = -7{,}373 + 2{,}799x_2 + 0.089x_3$	38	$g(x) = -7.589 + 2.872x_2 + 0.092x_3$
18	$g(x) = -7.387 + 2.753x_2 + 0.089x_3$	39	$g(x) = -7.668 + 2.840x_2 + 0.0925x_3$
19	$g(x) = -7.335 + 2.839x_2 + 0.089x_3$	40	$g(x) = -7.377 + 2.811x_2 + 0.089x_3$
20	$g(x) = -7.354 + 2.849x_2 + 0.090x_3$	41	$g(x) = -7.630 + 2.705x_2 + 0.091x_3$
21	$g(x) = -7.411 + 2.773x_2 + 0.089x_3$	42	$g(x) = -7.351 + 2.798x_2 + 0.089x_3$

Table 5 shows that the relationship between the predictors $x_2$and x_3 and the log-odds of the event is not constant across all locations but varies spatially. The positive coefficients for $x_2$and x_3 at most locations suggest that these variables generally increase the likelihood of the event occurring as they increase, although the strength of this effect varies across different locations. This spatially varying effect highlights the importance of considering geographic context in modelling, as it allows for more accurate predictions and better understanding of the local drivers of the event being studied.

A comparison of logistic regression models and GWLR with the two weightings is done to find a better model in describing the factors that affect water quality in Pontianak. This comparison can be seen based on the AIC value and the percentage of classification accuracy (Table 6).

Table 6. Comparison of Model Suitability

Model	AIC	Classification Accuracy
Logistic Regression	22.39	92.86%
GWLR (*Fixed Gaussian Kernel)*	22.52	92.86%
GWLR (*Adaptive Gaussian Kernel)*	22.81	90.48%

A comparison of the logistic regression and GWLR models with fixed and adaptive Gaussian kernel weighting was carried out to identify which approach best describes the factors affecting water quality in Pontianak. As shown in Table 6, the smallest AIC value is obtained by the logistic regression model (22.39), with a classification accuracy of 92.86%, indicating that this model provides the best balance between goodness of fit and model complexity. The GWLR model with a fixed Gaussian kernel has a slightly larger AIC (22.52) but the same classification accuracy (92.86%), suggesting that allowing coefficients to vary in space does not yield a meaningful improvement in predictive performance for these data. The GWLR model with an adaptive Gaussian kernel performs even less favorable, with a higher AIC (22.81) and a lower classification accuracy (90.48%), indicating that this more flexible spatial weighting scheme does not translate into better model adequacy. Overall, these results imply that, despite the presence of spatial information, the added complexity of GWLR is not justified in this case, and standard logistic regression is sufficient and more efficient for modelling the factors influencing water quality in Pontianak City.

CONCLUSION

The application of Geographically Weighted Logistic Regression (GWLR) involved the utilisation of two distinct types of spatial parameter structures. The first type involved the generation of spatially varying parameters using a fixed Gaussian kernel, while the second type utilised an adaptive Gaussian kernel to generate spatially varying parameters. The primary objective of this study was to assess whether local parameter variation enhanced the performance of the model in comparison to the (spatially invariant) logistic regression model. As demonstrated in Table 6, the GWLR model with fixed Gaussian kernel weighting yielded an AIC of 22.52 and a classification accuracy of 92.86%. In contrast, the GWLR model with adaptive Gaussian kernel weighting resulted in a higher AIC of 22.81 and a lower accuracy of 90.48%. The findings suggest that the two types of GWLR parameter structures yield divergent levels of model adequacy, with the fixed-kernel GWLR demonstrating a closer alignment with the model, and the adaptive-kernel GWLR exhibiting inferior performance. It was determined that neither form of GWLR yielded an AIC that was lower than that of the logistic regression model (AIC = 22.39). This finding indicates that the spatially varying parameter estimates did not result in a measurable improvement over the parameters in this dataset. Consequently, the evidence from both types of GWLR parameterization suggests that local spatial variation in the coefficients was not substantial enough to improve model performance, and the logistic regression model remained the most efficient representation for the relationships between the predictors and water quality status in Pontianak.

REFERENCES

Backhaus, K., Erichson, B., Gensler, S., Weiber, R., & Weiber, T. (2023). *Multivariate Analysis An Application-Oriented introduction* (2nd ed). Springer Gabler.

Debataraja, N. N., & Kusnandar, D. (2023). *Pengantar Analisis Data Spasial*. UNTAN Press.

Debataraja, N. N., Kusnandar, D., Imro'ah, N., & Rachmadiar, M. (2019). Penerapan Metode Cokriging Untuk Mengestimasi Jumlah Zat Padat Terlarut Pada Air Di Permukiman Kota Pontianak. *Jurnal Matematika Sains Dan Teknologi*, *20*(2), 142–148. https://doi.org/10.33830/jmst.v20i2.208.2019

Fathurahman, M., Purhadi, Sutikno, & Ratnasari, V. (2016). Pemodelan Geographically Weighted Logistic Regression pada Indeks Pembangunan Kesehatan Masyarakat di Provinsi Papua. *In Prosiding Seminar Nasional MIPA*, 34–42.

Fikri, M., Debataraja, N. N., & Kusnandar, D. (2019). Penentuan Sebaran Spasial Pencemaran Air Di Kota Pontianak Menggunakan Analisis Diskriminan Dua Kelompok. *Media Statistika*, *12*(2), 226. https://doi.org/10.14710/medstat.12.2.226-235

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationship*. Wiley.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (1st ed.). Wiley. https://doi.org/10.1002/9781118548387

Isazade, V., Qasimi, A. B., & Dong, P. (2023). Integration of Moran's I, geographically weighted regression (GWR), and ordinary least square (OLS) models in spatiotemporal modeling of COVID-19 outbreak in Qom and Mazandaran, Iran. *Modeling Earth Systems and Environment*, *9*, 3923–3937. https://doi.org/10.1007/s40808-023-01729-y

Kusnandar, D., Debataraja, N. N., & Dewi, P. R. (2019). Classification of water quality in Pontianak city using multivariate statistical techniques. *Applied Mathematical Sciences*, 1069–1075. https://doi.org/10.12988/ams.2019.99130

Kusnandar, D., Debataraja, N. N., & Fitriani, S. (2021). Pemodelan Sebaran Total Dissolved Solid Menggunakan Metode Mixed Geographically Weighted Regression. *Jorunal of Statistical Application and Computational Statistics*, *13*(1), 9–16. https://doi.org/10.34123/jurnalasks.v13i1.257

Kusnandar, D., Debataraja, N. N., & Nusantara, R. W. (2022). An Application of Geographically Weighted Regression for Assessing Water Polution in Pontianak, Indonesia. *Cauchy J. Mat Murni Dan Apl*, *7*(2), 185–194.

Kusnandar, D., Debataraja, N. N., Rizki, S. W., & Saputri, E. (2020). Water quality mapping in Pontianak City using multiple discriminant analysis. *AIP Conference Proceedings*, *2268*(1), 020006. https://doi.org/10.1063/5.0016809

Kusnandar, D., Debataraja, N. N., & Utari, S. (2021). Pemodelan Tingkat Kualitas Air di Kota Pontianak dengan Menggunakan Multivariate Geographically Weighted Regression. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *15*(3), 493–502. https://doi.org/10.30598/barekengvol15iss3pp493-502

Lessani, M. N., & Li, Z. (2024). SGWR: Similarity and Geographically Weighted Regression. *International Journal of Geographical Information Science*, *38*(1232–1255). https://doi.org/10.1080/13658816.2024.2294076

Lestari, F. D., Kusnandar, D., & Debataraja, N. N. (2020). Estimasi Parameter Model Geographically Weighted. *Buletin Ilmiah Matematika Statistika Dan Terapannya (Bimaster)*, *9*(1), 159–164.

Minister of Environment. (2003). *Peraturan Menteri Lingkungan Hidup Nomor 115 Tahun 2003 Tentang Pedoman Penentuan Status Mata Air*.

Minister of Health of the Republic of Indonesia. (2017). *Regulation of the Minister of Health of the*

Republic of Indonesia Number 32 of 2017 concerning Environmental Health Quality Standards and Water Health Requirements for Sanitation Hygiene Needs, Swimming Pools, Solus Per Aqua and Public Baths.

Omrani, F., Shad, R., & Ziaee, S. A. (2025). A Multiscale geographically weighted regression approach to emphasize the effects of traffic characteristics on vehicular emissions. *Science of the Total Environment*, *25*, 100315. https://doi.org/10.1016/j.aeaoa.2025.100315

Panggabean, M., & Debataraja, N. N. (2025). Quantitative Analysis of Land Transportation Facilities and Infrastructure (SDG 9) With Labor Absorption Related to GDRP (SDG 8) in West Kalimantan Province, Indonesia. *Journal of Lifestyle & SDG's Review*, *5*(7), e07252. https://doi.org/10.47172/2965-730X.SDGsReview.v5.n07.pe07252

Pardoe, I. (2021). *Applied Regression Modeling*. John Wiley & Sons. Inc.