Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2025, 2025:130
https://doi.org/10.28919/cmbn/9588
ISSN: 2052-2541

# GEOGRAPHICALLY WEIGHTED REGRESSION WITH BURR XII DISTRIBUTION FOR SPATIAL ANALYSIS OF DIARRHEA INCIDENCE IN SURAKARTA CITY

RIZWAN ARISANDI[1,*], PURHADI[2], PUSPITA KARTIKASARI[3], LUTFIATUL JANNAH[3], DIMAS ELANG SETYOKO[1]

[1]Department of Computer Science, Universitas Bina Nusantara, Jakarta 11480, Indonesia

[2]Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia

[3]Department of Statistics, Universitas Diponegoro, Semarang 50275, Indonesia



**Abstract:** Diarrhea is one of the leading causes of mortality in Indonesia. The persistence of this disease is closely related to spatial factors, such as environmental conditions and existing infrastructure. In Surakarta City, diarrhea was recorded as the most prevalent disease in 2024, with a 58% increase in cases compared to the previous year. The aim of this study is to apply the Geographically Weighted Regression (GWR) approach with the Burr XII distribution to model the spatial distribution of diarrhea cases in Surakarta. The study encompasses 54 urban villages and utilizes seven predictor variables, such as population density, elevation, distance to the nearest hospital, slope, proximity to rivers, distance to waste disposal sites, and rainfall. The Burr XII distribution was applied to handle the skewness frequently observed in health-related datasets such as diarrhea incidence, while GWR was used to generate location-specific parameter estimates. The developed model revealed clear spatial heterogeneity in how diarrhea cases are influenced by the predictor variables. Several factors, including population density, hospital distance, elevation, and river proximity, showed varying positive and negative effects depending on the geographical area. The results underline the necessity for targeted and well-coordinated public health measures to mitigate diarrhea risk, especially in communities with inadequate healthcare services and poor waste management systems.

**Keywords:** geographically weighted regression; burr xii distribution; diarrhea; spatial analysis.

**2020 AMS Subject Classification:** 62P10.

---

*Corresponding author
E-mail address: rizwan.arisandi@binus.ac.id

Received September 05, 2025

## 1. INTRODUCTION

The World Health Organization (WHO) defines diarrhea as a disease characterized by passing stool at least three times within 24 hours, or at a frequency higher than what is normal for the affected individual [1]. Diarrhea is a common symptom of gastrointestinal tract infection caused by bacteria, viruses, and parasites [2]. Its persistence is supported by relatively ready modes of transmission, namely through the intake of food and water that have been contaminated, as well as through person-to-person contact arising from inadequate hygiene practices [3]. At the global level, diarrhea causes approximately 443,832 deaths among children under five years old and 50,851 deaths among children aged nine years, out of an estimated 1.7 billion diarrhea cases in children each year [4].

According to the Central Statistics Agency of Surakarta City, diarrhea was the most frequently reported disease in 2024, with 11,434 recorded cases, representing an increase of 58 percent relative to the preceding year, which recorded only 7,209 cases [5]. The high number of diarrhea cases in Surakarta is strongly influenced by environmental and infrastructural factors within the community [6]. These factors will be examined using the Geographically Weighted Regression (GWR) method to assess their spatial influence on the distribution of diarrhea cases in Surakarta City.

The spatial distribution of diarrhea cases shows clear differences across distinct geographic areas. Diarrhea cases, especially among children, have been studied in previous research using spatial clustering methods [7][8][9]. However, a large proportion of these studies used global models or clustering methods that cannot fully reflect variations at the local level. Therefore, Geographically Weighted Regression (GWR) has been utilized across multiple disciplines such as epidemiology, disaster analysis, and environmental studies to address spatial heterogeneity [10][11][12].

Although the GWR method offers numerous benefits, its application in health-related research, particularly for diarrheal diseases, is still scarce. Conventional models often assume spatial homogeneity, potentially overlooking localized relationships that are essential for accurate disease mapping and targeted interventions. To overcome this limitation, this study applies the GWR method with the Burr XII distribution, which offers greater flexibility in modeling skewed and heavy-tailed data, as commonly found in health datasets. Thus, this study aims to model the spatial variation of diarrhea incidence in Surakarta using the Burr XII-based GWR approach and identify the local determinants that significantly influence the disease pattern.

# 2. Data and Methods

## 2.1 Data and Descriptive Statistics of Variables

This study utilises secondary data obtained from the Surakarta City Health Office, the Central Bureau of Statistics (BPS), and several publicly accessible geospatial data sources, including OpenStreetMap, the OpenTopography portal, and CHIRPS monthly precipitation data provided by the Climate Hazards Center, University of California. The unit of analysis consists of 54 urban villages (kelurahan) in Surakarta City. The data in this study can be accessed at the link:

1. https://www.openstreetmap.org/
2. https://surakartakota.bps.go.id/id/statistics-table/2/MzUwIzI=/jumlah-kasus-penyakit-menurut-kecamatan-dan-jenis-penyakit.html
3. https://surakartakota.bps.go.id/id/statistics-table/2/NTkwIzI=/kepadatan-penduduk-per-km2-menurut-kelurahan--jiwa-.html
4. https://surakartakota.bps.go.id/id/statistics-table/2/NTg4IzI=/jumlah-penduduk-menurut-kelurahan--jiwa-.html
5. https://surakartakota.bps.go.id/id/statistics-table/2/NTgwIzI=/luas-daerah-menurut-kelurahan.html
6. https://portal.opentopography.org/
7. https://data.chc.ucsb.edu/products/CHIRPS-2.0/indonesia_monthly/bils/

The research variables used in this study are described in Table 1.

**Table 1.** Research Variables

| Notations | Variable | Definition | Unit |
|---|---|---|---|
| Y | Number of diarrhea cases | Total number of diarrhea cases recorded in each urban village | cases |
| $X_1$ | Population density | Number of people per square kilometer in each urban village | people/km$^2$ |
| $X_2$ | Distance to the nearest hospital | Average distance from residential areas to the nearest hospital | kilometers (km) |
| $X_3$ | Rainfall | Average annual rainfall | millimeters (mm/year) |
| $X_4$ | Distance to waste disposal site | Average distance to the nearest official waste disposal site (TPS) | kilometers (km) |
| $X_5$ | Elevation | Average height above sea level | meters (m) |
| $X_6$ | Slope | Average stepness of the landscape | percent (%) |
| $X_7$ | Distance from the nearest river | Average proximity to the nearest river | kilometers (km) |

In addition, the data is summarized using descriptive statistics, including mean, standard deviation, minimum, and maximum values, to give an idea of the variation among urban villages. These summarized results are presented in the Results section.

### 2.2 Data Standardization

Standardization is an initial step used to convert variables with different measurement scales into a uniform scale without changing the differences in data ranges [13]. This procedure is crucial in applying GWR because variables with greater magnitudes can cause disproportionate impacts on the model. Standardization is carried out to guarantee that predictor variables with varying units are comparable in this study [14]. Each variable is adjusted to have an average of 0 and a standard deviation of 1. The formula for standardization is presented below:

$$z = \frac{x-\mu}{\sigma} \tag{1}$$

### 2.3 Spatial Distribution of Diarrhea Cases

The spatial distribution of diarrhea cases in 54 urban villages in Surakarta City was visualized using thematic maps. Data on administrative boundaries of urban villages were obtained from publicly available geospatial sources, namely OpenStreetMap. The mapping process was performed with R software using tmap package. The number of diarrhea cases was displayed as a choropleth map, with a classification of five categories based on the natural breaks (Jenks) method to highlight spatial variations. This stage aims to provide an initial picture of disease spread patterns and identify areas with potentially higher concentrations of cases, which further serves as the basis for spatial regression analysis.

### 2.4 Geographically Weighted Regression (GWR) Burr XII Modeling

Spatial regression is a model used to account for spatial factors, especially spatial autocorrelation and spatial variability. Spatial autocorrelation refers to the impact of nearby observations on a specific data point, while spatial variability describes differences in the relationships among variables across various regional areas [15]. Classical regression models presuppose independence and spatial uniformity among data points, assumptions that are frequently breached in spatial datasets due to autocorrelation resulting from the closeness of locations. To address these limitations, spatial models such as the Spatial Autoregressive (SAR) and Spatial Error Model (SEM) have been developed. The SAR model incorporates spatial dependence in the response variable, while the SEM captures it through the error structure [16].

However, both SAR and SEM assume constant relationships across space. In practice, especially when dealing with socio-economic or environmental data, these relationships may vary from one location to another. To accommodate this condition, Geographically Weighted Regression (GWR) was introduced as a locally adaptive method that allows regression coefficients to vary spatially, providing more flexible and spatially detailed inferences [17]. GWR extends the classical linear regression model by estimating local coefficients at each observation point, based on its geographic coordinates $(u_i, v_i)$ [18]. The general form of the GWR model is [19]:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \tag{2}$$

Where:

$y_i$ : the response variable at location $i$

$(u_i, v_i)$ : the geographic coordinates of the $i$-th location

$\beta_k(u_i, v_i)$ : $k$-th local regression coefficient

$x_{ik}$ : $k$-th explanatory variable

$\varepsilon_i$ : the error term, assumed to be independently and identically distributed.

To estimate the local coefficients $\beta_k(u_i, v_i)$, a spatial weighting function is applied to give higher weight to nearby observations. This study employs a Gaussian kernel function as the spatial weighting function [20]:

$$w_{ij} = \exp\left(\frac{d_{ij}^2}{h^2}\right) \tag{3}$$

Where $d_{ij}$ is the Euclidean distance between observations $i$ dan $j$, and $h$ is the bandwidth controlling the spatial extent of the kernel. The optimal bandwidth is determined using the Akaike Information Criterion (AIC) to achieve the best model fit. Specifically, the corrected AIC (AICc) used in GWR is calculated as [21]:

$$AIC_c(h) = -2\log L + 2\,tr(S) + \frac{2\,tr(S)(tr(S)+1)}{n - tr(S) - 1} \tag{4}$$

Where log $L$ is the log-likelihood of the Burr XII model, $tr(S)$ is the trace of the hat matrix representing the effective number of parameters, and $n$ is the number of observations. The value of $h$ that minimizes $AIC_c(h)$ is selected as the optimal bandwidth to balance model fit and complexity.

However, traditional GWR assumes that the residuals follow a normal distribution, which may not be suitable when the response variable is positively skewed or heavy-tailed. In this study, to accommodate such distributional characteristics, we adopted the Burr Type XII distribution as the underlying distribution of the response variable $Y$. The probability density function (pdf) of Burr

XII distribution, derived by differentiating its cumulative distribution function, is given as follows [22]:

$$f(y;c,d,\theta) = \frac{cd\, y^{c-1}}{\theta^{c}\left(1+\left(\frac{y}{\theta}\right)^{c}\right)^{d+1}}\ , y > 0,\ \ c,d,\theta > 0 \tag{5}$$

In the context of GWR, the scale parameter $\theta$ is modeled as a function of spatially varying covariates through a log-link function:

$$\theta_i = exp\left[\beta_0(u_i,v_i) + \sum_{k=1}^{p} \beta_k(u_i,v_i)\, x_{ki}\right] \tag{6}$$

This formulation ensures that the scale parameter remains positive and allows flexible modeling of skewed or heavy-tailed response variables. Thus, the model combines the flexibility of the Burr XII distribution in handling skewed data with the local adaptation of GWR to capture spatial non-stationarity in the regression structure. Parameter estimation is carried out using the Maximum Likelihood Estimation (MLE) method by minimizing the negative log-likelihood of the Burr XII distribution.

### 2.5 Spatial Variation of Local Coefficients

One of the main advantages of the Geographically Weighted Regression (GWR) framework is its ability to capture spatial heterogeneity in the relationship between predictor and response variables through the estimation of locally distinct regression coefficients at each location [23]. Unlike the global regression that produces one parameter for each predictor variable, GWR produces location-specific coefficients that can detect non-stationary effects [24].

In this study, local coefficients for each variable were extracted from the GWR-Burr XII model. To facilitate interpretation, the coefficient estimation results were visualized in the form of maps for 54 urban villages in Surakarta City. Color gradations are used to illustrate the size of the influence and the direction of the relationship of each variable. Positive values indicate that an increase in the predictor variable is associated with an increase in diarrhea cases, while negative values indicate a protective effect or a decrease in risk. Local coefficient mapping is a common approach in spatial epidemiology as it allows researchers to identify vulnerability clusters and highlight areas where determinants of a disease show significant differences [25]. This stage is important for translating statistical results into practical recommendations for public health planning and more spatially targeted interventions.

## 3. MAIN RESULTS

### 3.1 Descriptive Statistics of Variables

Table 2 provides summary statistics for all variables used in the modelling process. These statistics offer insight into the central tendency and variability of each variable, which is essential for understanding the data characteristics before applying the spatial regression model.

**Table 2.** Descriptive Statistics of Study Variables

| Variable | Description | Mean | Standard deviation | Min | Max |
|---|---|---|---|---|---|
| Y | Diarrhea cases | 133.43 | 159.49 | 3.00 | 887.00 |
| $X_1$ | Population density (people/$km^2$) | 13,945.43 | 4,883.15 | 7,343.93 | 27,730.43 |
| $X_2$ | Distance to the nearest hospital (km) | 0.50 | 0.26 | 0.15 | 1.26 |
| $X_3$ | Rainfall (mm/year) | 166.92 | 0.97 | 165.38 | 169.39 |
| $X_4$ | Distance to waste disposal site (km) | 1.85 | 1.22 | 0.23 | 5.49 |
| $X_5$ | Elevation (m) | 96.83 | 5.39 | 90.14 | 114.97 |
| $X_6$ | Slope (%) | 89.13 | 0.60 | 87.24 | 89.98 |
| $X_7$ | Distance to nearest river (km) | 0.00023 | 0.00025 | 0.00001 | 0.00089 |

The average number of diarrhea cases across 54 urban villages was 133.43 cases, with a standard deviation of 159.49, indicating considerable variation in disease incidence between regions. The minimum number of cases was 3, while the maximum reached 887 cases in a single village. Population density ($X_1$) showed a wide range from 7,344 to 27,730 people/km², with a mean of 13,945.43. This suggests notable differences in residential concentration across the study area. Distance to the nearest hospital ($X_2$) had an average of 0.50 km, ranging from 0.15 km to 1.26 km, reflecting disparities in healthcare accessibility. Rainfall ($X_3$) values were relatively uniform, with a mean of 166.92 mm/year and a small standard deviation of 0.97.

Elevation ($X_5$) ranged from 90.14 m to 114.97 m, with a mean of 96.83 m. Slope ($X_6$) had an average of 89.13%, with low variation across villages, indicating generally flat topography in the region. Distance to the nearest waste disposal site ($X_4$) varied from 0.23 km to 5.49 km, with a mean of 1.85 km. Meanwhile, distance to the nearest river ($X_7$) exhibited extremely small values for all observations, ranging from approximately 0.00001 km to 0.00089 km (around 0.01 to 0.89 meters). This indicates that all urban villages are located in very close proximity to rivers, with minimal spatial variation across the area.

### 3.2 Data Standardization

The standardization process was performed on all predictor variables to have a mean of 0 and standard deviation of 1. The results of standardization are shown in Table 3, which shows that before standardization the variables had different measurement scales. After standardization, all variables have a mean close to 0 and a standard deviation close to 1. This ensure that no variable dominates the model simply because of differences in measurement scale.

**Table 3.** Data Standardization

| Variable | Mean Before | SD Before | Mean After | SD After |
|---|---|---|---|---|
| $X_1$ | 13945.43 | 4483.15 | 0 | 1 |
| $X_2$ | 0.50 | 0.26 | 0 | 1 |
| $X_3$ | 166.92 | 0.97 | 0 | 1 |
| $X_4$ | 1.85 | 1.22 | 0 | 1 |
| $X_5$ | 96.83 | 5.39 | 0 | 1 |
| $X_6$ | 89.13 | 0.60 | 0 | 1 |
| $X_7$ | 0.00 | 0.00 | 0 | 1 |

### 3.3 Spatial Distribution of Diarrhea Cases

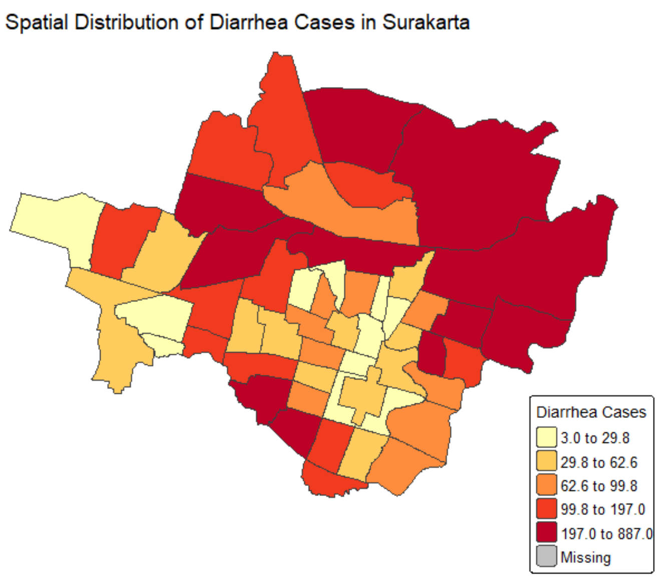


Figure 1. Spatial distribution of diarrhea cases in surakarta

Based on the spatial distribution of diarrhea cases across 54 sub-districts (kelurahan) in Surakarta, the number of reported cases varies considerably throughout the city. The highest number of cases is observed in the northern and northeastern parts, particularly in areas such as Mojosongo, Kadipiro, and Jebres, with reported cases ranging from 197 to 887. Meanwhile, the southern and western regions, including kelurahan such as Pajang, Kratonan, and Panularan, show moderate to high incidence levels.

Conversely, the lowest number of diarrhea cases, ranging from 3.0 to 29.8, is concentrated in several centrally located kelurahan, such as Kampung Baru and Gajahan. These areas exhibit notably lower incidence compared to other parts of the city. Overall, the map indicates an uneven spatial distribution of diarrhea cases in Surakarta, with a tendency for higher clustering in the peripheral areas and lower counts in the central urban zones. This pattern may reflect differences in population density, sanitation infrastructure, and environmental conditions across regions.

## 3.4 Estimation of GWR Burr XII Model

The estimation process of the GWR-Burr XII model involved fitting a local regression model at each spatial unit using maximum likelihood estimation (MLE), weighted by spatial proximity through a Gaussian kernel. The optimal bandwidth used in the model was 1028.362, selected based on the minimum total Akaike Information Criterion (AIC), which yielded a total AIC value of 6228.377. This bandwidth reflects the spatial scale at which the influence of neighboring observations remains informative for estimating local parameters.

The model produced a set of location-specific regression coefficients ($\beta_1$ to $\beta_7$) for the seven independent variables, as well as two Burr XII distribution parameters ($c$ and $d$). Table 4 summarizes the descriptive statistics of these estimated parameters across the 54 urban villages in Surakarta City.

**Table 4.** Summary Statistics of Local Parameter Estimates in GWR Burr XII Model

| Parameter | Mean | Standard deviation | Min | Max |
|---|---|---|---|---|
| $b_1$ (Population density) | 0.4251 | 0.5942 | -0.9547 | 3.1013 |
| $b_2$ (Distance to the nearest hospital) | 0.0155 | 0.4408 | -1.4485 | 1.1491 |
| $b_3$ (Rainfall) | 0.0256 | 1.1689 | -1.9435 | 4.5908 |
| $b_4$ (Distance to waste disposal site) | 0.2419 | 1.2060 | -2.7984 | 2.7995 |
| $b_5$ (Elevation) | 0.1881 | 0.8188 | -3.5232 | 2.7653 |
| $b_6$ (Slope) | -0.2562 | 0.3156 | -1.5111 | 0.8283 |
| $b_7$ (Distance to nearest river) | 0.4745 | 0.4271 | -0.2420 | 2.5955 |
| $c$ (Burr XII shape parameter) | 11.5489 | 24.2738 | - | - |
| $d$ (Burr XII shape parameter) | 19.9262 | 22.3621 | - | - |

The summary of local parameter estimates confirms substantial spatial variation in the relationship between diarrhea incidence and its predictors. The coefficient for population density ($b_1$) ranges from $-0.95$ to 3.10 with a standard deviation of 0.59, indicating that the direction and strength of association vary across locations. Similarly, the coefficient for distance to the nearest hospital ($b_2$) shows a broad range from −1.45 to 1.15 and a standard deviation of 0.44, suggesting heterogeneous effects of health service accessibility across neighborhoods.

High variability is also observed in rainfall ($b_3$) and distance to waste disposal sites ($b_4$), with standard deviations of 1.17 and 1.21, respectively, reflecting strong spatial heterogeneity in environmental influences. Elevation ($b_5$) also displays notable variation (standard deviation of 0.82) with coefficients spanning from −3.52 to 2.77. In contrast, slope ($b_6$) exhibits relatively more stable effect spatially, with a narrow coefficient range (−1.51 to 0.83) and a lower standard deviation of 0.32. The coefficient for distance to rivers ($b_7$) is moderately variable, ranging from −0.24 to 2.60 with a standard deviation of 0.43.

The Burr XII shape parameters further highlight the model's flexibility in capturing distributional characteristics of the response variable. The average value of parameter c is 11.55 with a standard deviation of 24.27, while parameter d has a mean of 19.93 and a standard deviation of 22.36. These values indicate the model's ability to accommodate skewness and heavy tails in the diarrhea incidence data. In summary, the GWR-Burr XII model demonstrates strong performance in capturing both spatial variation and non-normality in the response distribution, making it a suitable approach for spatial epidemiological modeling of diarrhea cases.

To highlight the spatial variability captured by the model, Table 5 provides the log link equations of the GWR Burr XII model for each sub-district in Surakarta City.

**Table 5.** Log-link equations of the GWR-Burr XII model for each sub-district in Surakarta City

| Sub-district | $\boldsymbol{\beta_0}$ | $\boldsymbol{X_1}$ | $\boldsymbol{X_2}$ | $\boldsymbol{X_3}$ | $\boldsymbol{X_4}$ | $\boldsymbol{X_5}$ | $\boldsymbol{X_6}$ | $\boldsymbol{X_7}$ | c | d |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 4.052 | 0.374 | 0.214 | -1.146 | 0.47 | 0.303 | -0.171 | 0.052 | 2.524 | 1.431 |
| 2 | 3.201 | 0.066 | 0.175 | 0.726 | 0.253 | -0.196 | -0.533 | 0.311 | 36.410 | 0.036 |
| 3 | 5.119 | -0.899 | 0.692 | 0.402 | -0.243 | -0.788 | 0.038 | 0.324 | 6.052 | 1.814 |
| 4 | 7.692 | 1.072 | -0.234 | 0.812 | -1.963 | 1.771 | -0.171 | 0.617 | 2.134 | 34.857 |
| 5 | 5.387 | 0.624 | 0.157 | 4.003 | 1.15 | 0.005 | -0.033 | -0.102 | 4.536 | 0.686 |
| 6 | 4.379 | 0.386 | 0.204 | -0.058 | 0.51 | 0.31 | -0.161 | 0.02 | 2.699 | 1.279 |
| 7 | 5.198 | 0.204 | 0.04 | -0.595 | 1.555 | 0.053 | -0.58 | 0.624 | 3.087 | 0.816 |
| 8 | 3.843 | 0.399 | -0.162 | -0.545 | 1.612 | -0.114 | -0.093 | 0.723 | 11.290 | 0.080 |
| 9 | 5.764 | 0.167 | 0.006 | -0.865 | 2.721 | 0.137 | -0.6 | 0.448 | 6.349 | 0.347 |
| 10 | 6.483 | 1.518 | -0.809 | 0.599 | -1.417 | -0.664 | -0.662 | 1.175 | 5.694 | 1.692 |
| 11 | 6.468 | 0.271 | 0.144 | -0.125 | -0.085 | 0.565 | -0.22 | 0.233 | 1.815 | 28.751 |
| 12 | 5.817 | 0.279 | 0.64 | -1.355 | 1.979 | 1.465 | -0.255 | 0.085 | 8.461 | 1.222 |
| 13 | 3.704 | 1.237 | -0.452 | -0.609 | 1.302 | 0.586 | -0.404 | 0.558 | 43.209 | 0.033 |
| 14 | 3.272 | 0.512 | 0.262 | -1.943 | 0.943 | -0.236 | -0.055 | -0.096 | 3.977 | 1.098 |
| 15 | 4.618 | 0.702 | 0.263 | 2.823 | 1.149 | -0.306 | -0.014 | -0.154 | 6.955 | 0.605 |

| Sub-district | $\beta_0$ | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | c | d |
|---|---|---|---|---|---|---|---|---|---|---|
| 16 | 3.187 | -0.093 | 0.198 | 0.791 | 0.238 | -0.288 | -0.467 | 0.341 | 138.584 | 0.038 |
| 17 | 7.17 | 0.113 | 0.229 | -0.193 | -0.348 | 0.205 | -0.498 | 0.616 | 1.477 | 51.401 |
| 18 | 9.174 | 3.101 | -1.177 | 0.681 | -0.622 | -3.523 | 0.828 | 2.595 | 49.111 | 1.364 |
| 19 | 6.851 | 0.19 | 0.208 | -0.237 | -0.262 | 0.423 | -0.362 | 0.374 | 1.579 | 44.397 |
| 20 | 6.794 | 0.156 | 0.28 | -0.393 | -0.316 | 0.367 | -0.468 | 0.552 | 1.663 | 47.863 |
| 21 | 6.95 | 0.219 | 0.139 | -0.149 | -0.154 | 0.529 | -0.255 | 0.332 | 1.669 | 49.192 |
| 22 | 3.654 | 0.351 | -0.341 | -0.109 | 1.673 | -0.666 | -0.046 | 0.39 | 55.746 | 0.015 |
| 23 | 4.51 | 0.429 | -0.254 | -0.857 | 2.112 | -0.154 | -0.131 | 0.572 | 3.635 | 0.326 |
| 24 | 7.277 | 0.136 | 0.167 | -0.158 | -0.222 | 0.235 | -0.373 | 0.531 | 1.471 | 56.323 |
| 25 | 7.508 | 1.409 | -0.696 | 0.717 | -2.055 | 1.263 | -0.149 | 0.788 | 2.861 | 31.489 |
| 26 | 6.873 | 0.082 | 0.143 | -0.103 | -0.041 | -0.103 | -0.384 | 0.675 | 1.382 | 21.779 |
| 27 | 7.132 | 0.124 | 0.099 | -0.103 | 0.018 | 0.014 | -0.273 | 0.576 | 1.452 | 37.016 |
| 28 | 5.639 | 0.31 | 0.162 | -0.012 | 0.099 | 0.426 | -0.176 | 0.131 | 2.045 | 8.908 |
| 29 | 7.598 | 1.564 | -0.704 | 0.676 | -2.265 | 1.465 | -0.15 | 0.873 | 3.205 | 32.618 |
| 30 | 7.328 | 0.355 | -0.148 | 0.396 | 0.025 | -0.419 | -0.104 | 0.551 | 1.702 | 55.761 |
| 31 | 7.162 | 0.227 | 0.008 | 0.146 | 0.299 | -0.295 | -0.093 | 0.523 | 1.678 | 48.808 |
| 32 | 3.778 | 0.435 | 0.438 | -1.764 | 0.59 | 0.052 | -0.077 | -0.08 | 3.964 | 2.678 |
| 33 | 6.158 | -0.955 | 1.149 | -0.143 | -0.373 | 0.044 | -1.511 | 1.074 | 5.831 | 4.200 |
| 34 | 3.251 | 0.176 | 0.171 | 0.666 | 0.254 | -0.15 | -0.541 | 0.336 | 77.162 | 0.014 |
| 35 | 3.989 | 0.885 | -1.448 | -0.166 | -0.386 | -0.663 | -0.347 | 1.027 | 60.820 | 0.117 |
| 36 | 6.704 | 0.372 | 0.083 | 0.388 | -0.172 | 0.63 | -0.103 | 0.321 | 2.216 | 43.803 |
| 37 | 4.483 | 0.287 | 0.294 | -0.968 | 0.172 | 0.28 | -0.225 | 0.219 | 2.202 | 2.741 |
| 38 | 7.123 | 0.376 | 0.024 | 0.275 | -0.161 | 0.381 | -0.101 | 0.46 | 2.004 | 70.344 |
| 39 | 5.595 | 0.383 | 0.567 | -1.9 | 2.799 | 1.173 | 0.29 | -0.242 | 8.414 | 0.802 |
| 40 | 7.187 | 0.164 | 0.049 | -0.019 | 0.188 | -0.098 | -0.173 | 0.53 | 1.524 | 41.809 |
| 41 | 5.285 | 0.235 | -0.074 | -0.92 | 2.289 | -0.004 | -0.527 | 0.52 | 3.637 | 0.473 |
| 42 | 7.652 | 0.917 | -0.291 | 0.59 | -1.229 | 0.781 | -0.149 | 0.623 | 1.802 | 39.993 |
| 43 | 6.377 | 0.181 | 0.403 | -0.473 | -0.259 | 0.691 | -0.477 | 0.556 | 2.159 | 34.453 |
| 44 | 4.276 | 0.353 | 0.407 | -1.248 | 0.343 | 0.192 | -0.127 | 0.158 | 3.025 | 3.201 |
| 45 | 4.748 | 0.337 | 0.212 | 0.16 | 0.316 | 0.257 | -0.161 | 0.049 | 2.524 | 2.248 |
| 46 | 7.238 | 0.053 | 0.222 | -0.116 | -0.303 | 0 | -0.538 | 0.755 | 1.373 | 38.162 |
| 47 | 5.652 | 0.134 | 0.011 | -0.379 | 1.903 | 0.121 | -0.795 | 0.707 | 4.825 | 0.580 |
| 48 | 6.224 | 1.296 | -0.592 | 0.644 | -2.798 | 2.765 | -0.236 | 0.585 | 4.000 | 1.532 |
| 49 | 6.907 | 0.254 | 0.083 | 0.05 | -0.059 | 0.47 | -0.143 | 0.394 | 1.890 | 51.428 |
| 50 | 6.954 | 0.073 | 0.279 | -0.367 | -0.281 | 0.303 | -0.671 | 0.696 | 1.583 | 35.772 |
| 51 | 6.876 | 0.129 | -0.299 | 0.499 | -0.624 | -0.153 | -0.119 | 0.305 | 2.270 | 48.843 |
| 52 | 4.378 | 0.641 | -0.439 | -1.144 | 2.64 | -0.204 | 0.151 | 0.792 | 11.999 | 0.080 |
| 53 | 7.257 | 0.184 | 0.091 | -0.09 | -0.017 | 0.297 | -0.206 | 0.444 | 1.611 | 62.484 |
| 54 | 7.91 | 0.461 | 0.042 | 4.591 | 0.118 | 0.617 | -0.032 | 0.154 | 2.349 | 28.214 |

Among all sub-districts, Mojosongo recorded the highest case count, while Laweyan reported the lowest. Their respective GWR Burr XII models, expressed in log link form, are as follows:

**Laweyan** (lowest case count, 3 cases)

$log(\eta) = 7.598 + 1.564X_1 - 0.704X_2 + 0.676X_3 - 2.265X_4 + 1.465X_5 - 0.150X_6 + 0.873X_7$ with $c = 3.2054,\ d = 32.6182$

In Laweyan, the coefficient for population density (X1) is positive (1.564), indicating that higher density is associated with an increase in the scale parameter (θ), which corresponds to a lower probability of diarrhea incidence. The coefficient for distance to the nearest hospital ($X_2$) is negative (-0.704), suggesting that closer proximity to medical facilities reduces risk. Distance to waste disposal site ($X_4$) is also negative (-2.265), implying that better accessibility to waste management contributes to lower disease occurrence. These results suggest that in low-incidence areas such as Laweyan, access to key facilities and services plays a protective role.

**Mojosongo** (highest case count, 887 cases)

$log(\eta) = 6158 - 0.955X_1 + 1.149X_2 - 0.143X_3 - 0.373X_4 + 0.044X_5 - 1.511X_6 + 1.074X_7$ with $c =\ 5.8312\ \ d = 4.2002$

In Mojosongo, population density ($X_1$) shows a strong negative coefficient (-0.955), indicating that higher density is associated with a lower θ and thus a higher probability of diarrhea incidence. The distance to waste disposal site ($X_4$) has a negative coefficient (-0.373), suggesting that households located farther from waste sites tend to face an increased risk of diarrhea, likely due to limited access to proper disposal facilities. The percentage of households without proper sanitation ($X_5$) is slightly positive (0.044), highlighting that poor sanitation conditions remain a dominant risk factor in this area.

Overall, this comparison illustrates the spatial heterogeneity in how environmental and infrastructural factors influence diarrhea incidence. Laweyan's low risk profile is shaped by infrastructure accessibility, while Mojosongo's high risk profile reflects vulnerabilities related to density and sanitation deficits.

## 3.5 Spatial Variation of Local Coefficients

The local coefficient maps derived from the GWR-Burr XII model reveal spatial heterogeneity in the influence of each explanatory variable on diarrhea incidence. These maps visualize the distribution of coefficient estimates across sub-districts, with color gradations indicating the magnitude and direction of each variable's effects.

$b_1$ (Population density)

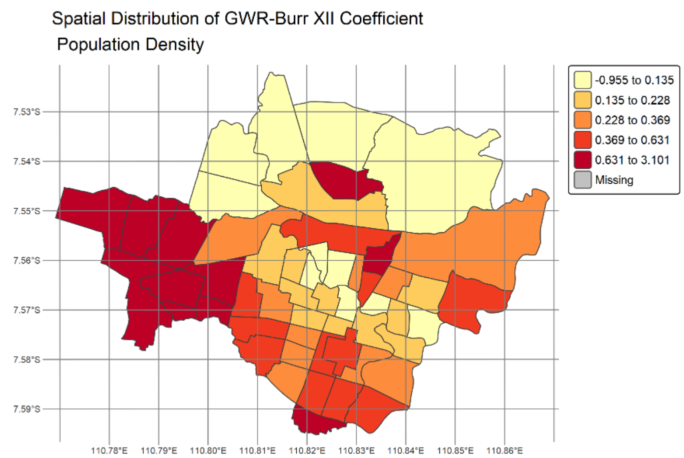


Figure 2. Spatial distribution of local coefficients for population density

The coefficient for population density ranges from −0.955 to 1.101, with values showing clear spatial variation across the city. Positive coefficients are predominantly concentrated in the central, southern, and southeastern subdistricts, indicating that higher population density in these areas tends to elevate the risk of diarrhea incidence. Conversely, negative coefficients are more evident in the northern and northwestern parts of the city, suggesting that in these locations, higher density may be linked to a reduced risk—possibly due to better infrastructure and environmental management that mitigate the effects of crowding. Overall, the spatial heterogeneity highlights that the influence of population density on diarrhea incidence is not uniform but context-dependent across different subdistricts.

$b_2$ (Distance to the nearest hospital)

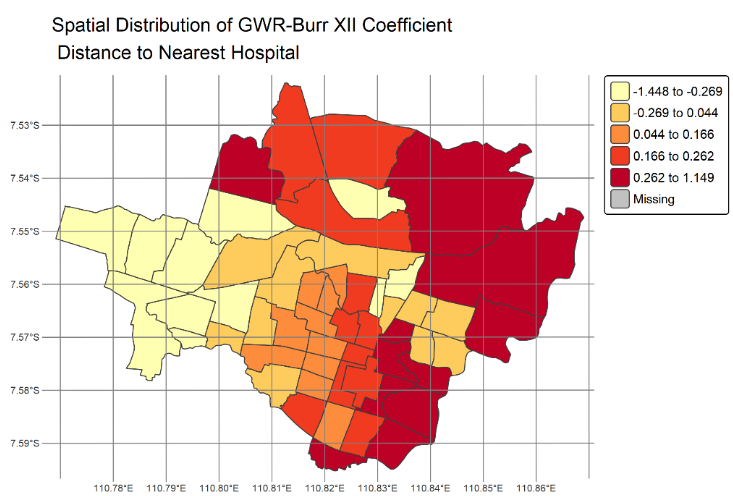


Figure 3. Spatial distribution of local coefficients for distance to the nearest hospital

The coefficient for distance to the nearest hospital ranges from −1.448 to 1.149, with values showing notable spatial variation. Strong positive effects are concentrated in the eastern, southeastern, and parts of the central subdistricts, suggesting that greater distance from hospitals in these areas is associated with an increased risk of diarrhea incidence, likely due to limited and delayed access to medical care. In contrast, negative coefficients are observed in the northern and western subdistricts, indicating that shorter distances to healthcare facilities may play a protective role by enabling faster treatment and better access to health services. This spatial variation underscores the importance of equitable healthcare accessibility in mitigating diarrhea risk across the city.

$b_3$ (Rainfall)

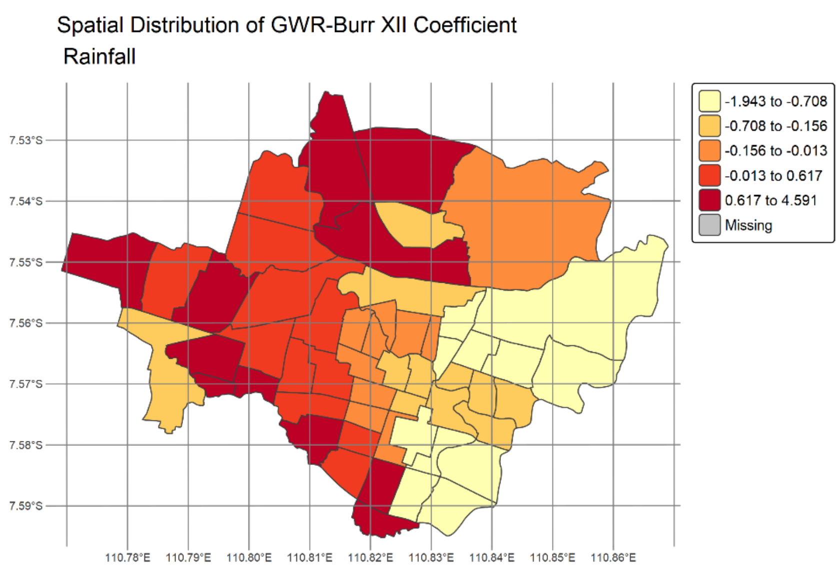


**Figure 4.** Spatial distribution of local coefficients for rainfall

The coefficient for rainfall ranges from −1.943 to 4.591, showing relatively wide spatial variability compared to other variables. Positive effects dominate in the western, central, and southern subdistricts, where higher rainfall is associated with increased diarrhea incidence. This may be linked to flooding, runoff, or water source contamination that escalates transmission risk. In contrast, negative coefficients are concentrated in the eastern and northeastern areas, suggesting that rainfall there may have a neutral or even protective effect, possibly due to better drainage systems or lower exposure to contaminated water. These findings indicate that the impact of rainfall on diarrhea incidence is spatially heterogeneous and strongly shaped by local environmental and infrastructural conditions.

$b_4$ (Distance to waste disposal site)

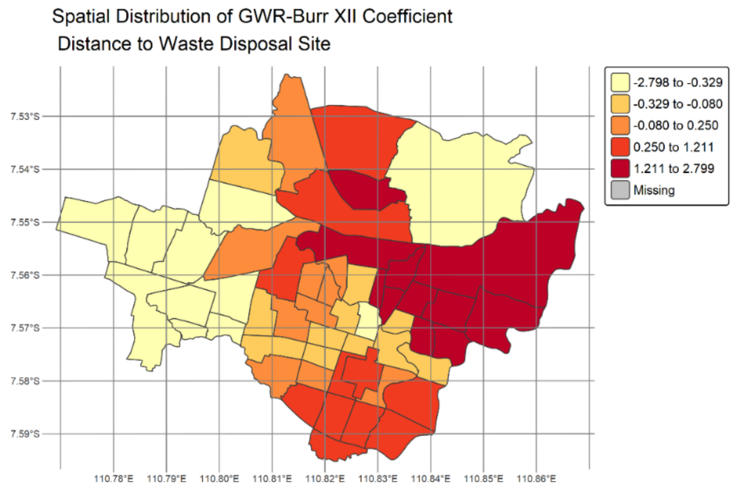


**Figure 5.** Spatial distribution of local coefficients for distance to waste disposal site

The coefficient for distance to waste disposal sites ranges from −2.798 to 2.799, indicating substantial spatial heterogeneity. Strong positive effects are concentrated in the eastern, southeastern, and parts of the southern subdistricts, suggesting that households located farther from formal waste disposal sites experience higher diarrhea risk, possibly due to reliance on informal or unsafe dumping practices. Conversely, negative coefficients are observed in the northern and some western areas, implying that in these regions, greater distance from waste facilities may be linked to reduced risk, potentially reflecting better overall waste infrastructure or alternative disposal mechanisms. These contrasting patterns highlight the importance of localized waste management strategies in shaping health outcomes.

$b_5$ (Elevation)

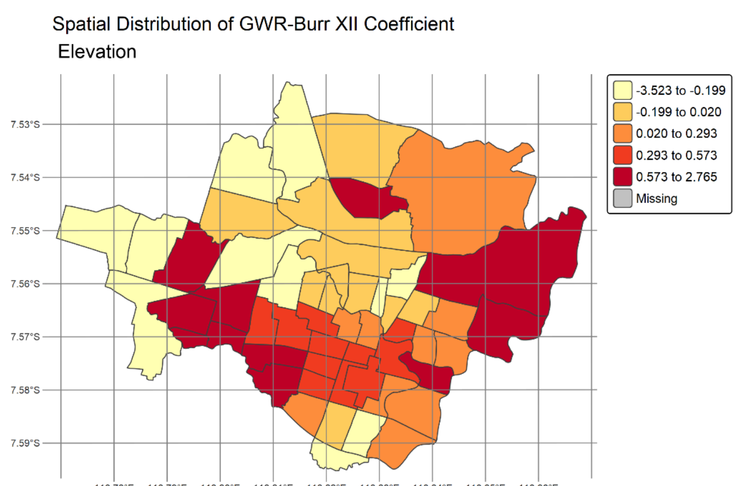


**Figure 6.** Spatial distribution of local coefficients for elevation

The coefficient for elevation ranges from −3.523 to 2.765, showing considerable variability across space. Positive effects are concentrated in the eastern, southeastern, and parts of the southern subdistricts, indicating that higher elevation in these areas is associated with increased diarrhea incidence, potentially due to limited access to piped water, poor sanitation facilities, or infrastructural constraints in upland settlements. In contrast, negative coefficients dominate the northern and western subdistricts, suggesting that in these areas, higher elevation may reduce risk—possibly because of better drainage, lower flood exposure, or improved environmental conditions. These spatial differences highlight that the relationship between elevation and diarrhea incidence is context-specific and mediated by local infrastructure and environmental factors.

$b_6$ (Slope)

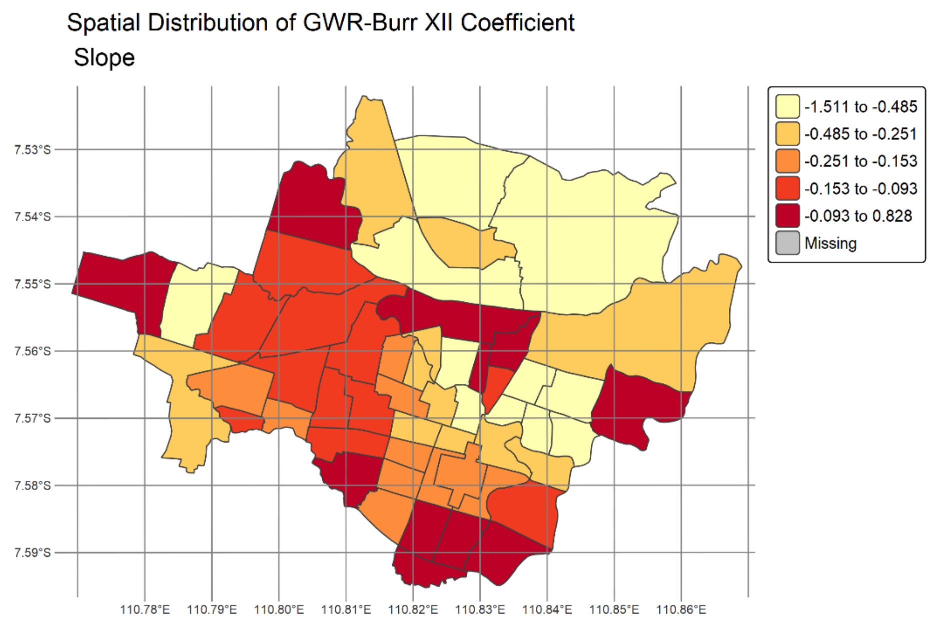


**Figure 7.** Spatial distribution of local coefficients for slope

The coefficient for slope ranges from −1.511 to 0.828, reflecting moderate spatial variability. Positive effects are concentrated in the southern, central, and parts of the eastern subdistricts, where steeper slopes may influence runoff and drainage patterns that facilitate the spread of contaminants, thereby increasing diarrhea risk. Conversely, negative coefficients are observed in the northern and some western areas, suggesting that in flatter terrains, slope has less influence on disease transmission, possibly due to more stable water flow and reduced runoff-related contamination. This spatial heterogeneity highlights how topographic variation can mediate environmental pathways of diarrhea risk across the city.

$b_7$ (Distance to nearest river)

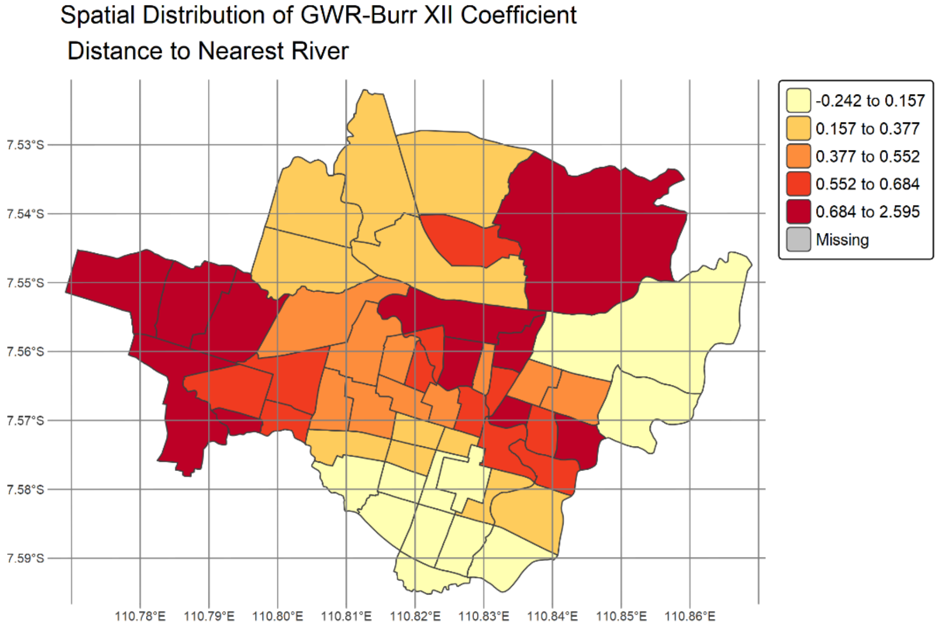


**Figure 8.** Spatial distribution of local coefficients for distance to nearest river

The coefficient for distance to the nearest river ranges from −0.242 to 2.595, with clear spatial variation across subdistricts. Strong positive effects are concentrated in the central and northeastern areas, indicating that closer proximity to rivers in these regions is associated with elevated diarrhea risk, likely due to contamination of water sources or higher exposure to flooding events. Meanwhile, negative coefficients are observed in some southern and western subdistricts, suggesting that distance from rivers there may play a protective role, potentially reflecting reduced dependence on river water or more effective flood control and sanitation infrastructure. These patterns emphasize the dual role of rivers as both vital resources and potential pathways for disease transmission.

## 4. CONCLUSION

This study implemented a Geographically Weighted Regression (GWR) model with the Burr XII distribution to analyze the spatial variation of diarrhea incidence across 54 urban villages in Surakarta. Seven predictor variables were considered, including population density, elevation, distance to the nearest hospital, slope, proximity to rivers, distance to waste disposal sites, and rainfall. The Burr XII distribution addressed the skewed nature of health data, while GWR captured local variations in relationships.

Results show that population density and hospital access have different effects in different areas; some neighborhoods are at higher risk with dense conditions and limited access to services, while others show a protective effect. Rainfall, elevation, slope, and proximity to rivers also showed diverse patterns, confirming that topography and hydrology have important contributions to disease spread. On the other hand, distance to dumping sites appears to be the dominant factor: eastern and southern areas tend to be more vulnerable when far from formal disposal facilities, while western and northern areas display the opposite trend.

The implication of these findings is the need for health policies that are tailored to local conditions. Improvements to the waste management system should be prioritized in areas with low access to TPS, while in areas with limited health services, equitable distribution of facilities and transportation is needed. In flood-prone areas or those near rivers, strengthening water and sanitation management is crucial to reduce the risk of disease exposure. In the future, research can be expanded by adding spatial variables and community behavior factors, as well as exploring other more flexible statistical distributions to describe the complexity of health data more comprehensively.

## AUTHORS' CONTRIBUTIONS

Conceptualization, methodology, formal analysis: Rizwan Arisandi and Purhadi; software, visualization: Rizwan Arisandi and Dimas Elang Setyoko; validation, Puspita Kartikasari and Lutfiatul Jannah; investigation, resources, writing—original draft preparation, supervision, writing—review and editing: Rizwan Arisandi and Puspita Kartikasari; data curation, Lutfiatul Jannah.

## CONFLICT OF INTEREST

The authors declare that there is no conflict of interest

## REFERENCES

[1] WHO, Diarrhoeal Disease, 2024. https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease.

[2] T. Winarni, Faktor-Faktor Yang Berhubungan Dengan Kejadian Diare Pada Balita Di Wilayah Kerja Puskesmas Pangkalan Balai Kabupaten Banyuasin Tahun 2021, Thesis, Sekolah Tinggi Ilmu Kesehatan, 2021.

[3] N. Noviyanti, S. Fatmawati, Studi Korelasi Personal Hygiene Ibu Dengan Kejadian Diare Pada Balita Di Kota Surakarta, Aisyiyah Surak. J. Nurs. 5 (2024), 132-139. https://doi.org/10.30787/asjn.v5i2.1564.

[4] S. Sulastri, A. Santoso, N.C. Kumalasari, M.A. Ramadhani, R.L. Vifta. Penyuluhan Edukasi Terkait Pencegahan dan Penatalaksanaan Diare Pada Anak Bagi Warga Desa Nogosari, J. Inov. Pengabdi. Masy. 4 (2024), 37-44.

[5] Badan Pusat Statistika Kota Surakarta, Surakarta Dalam Angka (Surakarta City in Figures) 2025, Surakarta: Badan Pusat Statistik Kota Surakarta, 2025.

[6] E. Rokhayati, Y. Hani, et al. Hubungan Sanitasi Lingkungan Dengan Kejadian Diare Pada Balita di Kecamatan Jebres Surakarta, Avicenna: J. Health Res. 7 (2024), 85-91. https://doi.org/10.36419/avicenna.v7i1.1032.

[7] K. Ghosh, A.S. Chakraborty, S. SenGupta, Identifying Spatial Clustering of Diarrhoea among Children Under 5 Years Across 707 Districts in India: A Cross Sectional Study, BMC Pediatr. 23 (2023), 272. https://doi.org/10.1186/s12887-023-04073-3.

[8] D.Y. Fa'idah, A.M. Hudzaifa, R.S. Pontoh, Clustering of Childhood Diarrhea Diseases Using Gaussian Mixture Model, Commun. Math. Biol. Neurosci. 2024 (2024), 10. https://doi.org/10.28919/cmbn/8365.

[9] D.Y. Fa'idah, G. Darmawan, B. Tantular, R.S. Pontoh, A.M. Hudzaifa, H. Sain, Spatial Clusters and Determinants of the High Incidence of Diarrhea in Children, Commun. Math. Biol. Neurosci. 2024 (2024), 80. https://doi.org/10.28919/cmbn/8669.

[10] X. Wu, J. Zhang, Exploration of Spatial-Temporal Varying Impacts on COVID-19 Cumulative Case in Texas Using Geographically Weighted Regression (GWR), Environ. Sci. Pollut. Res. 28 (2021), 43732-43746. https://doi.org/10.1007/s11356-021-13653-8.

[11] C. Shang, Q. Chen, K. Wang, Y. Zhang, G. Zheng, et al., Research on Spatiotemporal Heterogeneity of the Impact of Earthquakes on Global Copper Ore Supply Based on Geographically Weighted Regression, Sustainability 16 (2024), 1487. https://doi.org/10.3390/su16041487.

[12] S.A. Ania, N. Ilyas, E.T. Herdiani, Robust Spatial Temporal Analysis with Improved Geographically and Temporally Weighted Regression Model of Dengue Incidence Rate, Commun. Math. Biol. Neurosci. 2024 (2024), 22. https://doi.org/10.28919/cmbn/8376.

[13] A. Athifaturrofifah, R. Goejantoro, D. Yuniarti, Perbandingan Pengelompokan K-Means dan K-Medoids Pada Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas, Eksponensial 10 (2020), 143-152.

[14] F.L. Gewers, G.R. Ferreira, H.F.D. Arruda, F.N. Silva, C.H. Comin, et al., Principal Component Analysis, ACM Comput. Surv. 54 (2021), 1-34. https://doi.org/10.1145/3447755.

[15] H. Yasin, A.R. Hakim, B. Warsito, Development Life Expectancy Model in Central Java Using Robust Spatial Regression with M-Estimators, Commun. Math. Biol. Neurosci. 2020 (2020), 69. https://doi.org/10.28919/cmbn/4984.

[16] N.A. Jelita, S.S. Handajani, I. Susanto, Modeling Human Development Index of East Java Using Spatial Autoregressive and Spatial Error Ensemble, PYTHAGORAS J. Pendidik. Mat. 19 (2024), 165-178. https://doi.org/10.21831/pythagoras.v19i2.78621.

[17] E. Marhamah, I.G.N.M. Jaya, Modeling Positive COVID-19 Cases in Bandung City by Means Geographically

Weighted Regression, Commun. Math. Biol. Neurosci. 2020 (2020), 77. https://doi.org/10.28919/cmbn/4991.

[18] Sediono, M.F.F. Mardianto, S.M. Ulyah, et al. The Modelling of Earthquake Magnitude in the Southern Part of Java Island Using Geographically Weighted Regression, Commun. Math. Biol. Neurosci. 2022 (2022), 13. https://doi.org/10.28919/cmbn/6753.

[19] W. Zhao, J. Zhong, J. Lv, Influencing Factors and Spatial Differentiation of Rental Housing in a Smart City: A GWR Model Analysis, Measurement: Sensors 33 (2024), 101126. https://doi.org/10.1016/j.measen.2024.101126.

[20] R. Putra, S.W. Tyas, M.G. Fadhlurrahman, Geographically Weighted Regression with the Best Kernel Function on Open Unemployment Rate Data in East Java Province, Enthusiastic: Int. J. Stat. Data Sci. 2 (2022), 26-36.

[21] Z. Li, A.S. Fotheringham, T.M. Oshan, L.J. Wolf, Measuring Bandwidth Uncertainty in Multiscale Geographically Weighted Regression Using Akaike Weights, Ann. Am. Assoc. Geogr. 110 (2020), 1500-1520. https://doi.org/10.1080/24694452.2019.1704680.

[22] R. Arisandi, Survival Analysis with Log-Burr XII Regression on Diabetes, AIP Conf. Proc. 3165 (2024), 040004. https://doi.org/10.1063/5.0215817.

[23] S. Wang, K. Gao, L. Zhang, B. Yu, S.M. Easa, Geographically Weighted Machine Learning for Modeling Spatial Heterogeneity in Traffic Crash Frequency and Determinants in US, Accid. Anal. Prev. 199 (2024), 107528. https://doi.org/10.1016/j.aap.2024.107528.

[24] D.W. Cullen, V. Guida, Use of Geographically Weighted Regression to Investigate Spatial Non-Stationary Environmental Effects on the Distributions of Black Sea Bass (Centropristis Striata) and Scup (Stenotomus Chrysops) in the Mid-Atlantic Bight, USA, Fish. Res. 234 (2021), 105795. https://doi.org/10.1016/j.fishres.2020.105795.

[25] F.S. Muhammad, S.M. Shahabudin, M.B.A. Talib, Measuring Spatial Inequalities in Maternal and Child Mortalities in Pakistan: Evidence from Geographically Weighted Regression, BMC Public Health 24 (2024), 2229. https://doi.org/10.1186/s12889-024-19682-5.