CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 11 (1) (2026), Pages 239-254
p-ISSN: 2086-0382; e-ISSN: 2477-3344

# Spatial Variation of HDI in East Java: A Tricube-Based Geographically Weighted Regression–Flower Pollination Algorithm Modeling Approach

Friansyah Gani*, Henny Pramoedyo, and Achmad Efendi

*Department of Statistics, Faculty of Mathematics and Natural Science, Brawijaya University, Indonesia*

**Abstract**

Understanding spatial disparities in human development is essential for designing equitable development policies. This study examines the spatial variation of the Human Development Index (HDI) in East Java Province using an integrated Geographically Weighted Regression–Flower Pollination Algorithm (GWR–FPA) optimized with a Tricube kernel. The integration of GWR and FPA enables simultaneous spatial weighting and bandwidth optimization using the corrected Akaike Information Criterion (AICc) as the objective function. For standard GWR, the bandwidth was selected using Cross-Validation (CV) to minimize prediction error, while for the GWR–FPA model, bandwidth optimization was performed using the Flower Pollination Algorithm (FPA) with the corrected Akaike Information Criterion (AICc) as the objective function. Three predictors were analyzed: population size ($X_1$), literacy rate ($X_2$), and mean years of schooling ($X_3$). Statistical diagnostics indicated significant spatial autocorrelation and heteroskedasticity in the OLS residuals, justifying the use of a spatial modeling framework. The GWR estimates revealed strong spatial non-stationarity: $X_1$ showed no significant local effect, whereas educational factors ($X_2$ and $X_3$) were significant in all 38 districts and cities. The FPA optimization enhanced bandwidth selection, resulting in improved model fit. Model comparison based on AIC and AICc showed that the GWR–FPA–Tricube model achieved the lowest values (AIC = 135.8821; AICc = 137.0045), outperforming both global OLS and standard GWR. The results highlight the dominant contribution of education-related components to the spatial decomposition of HDI variation across East Java. The optimized model provides a more accurate spatial representation of local development disparities, supporting targeted policy interventions and illustrating the effectiveness of integrating metaheuristic optimization within spatial regression.

**Keywords:** GWR; FPA; Tricube; HDI.

Copyright © 2025 by Authors, Published by CAUCHY Group. This is an open access article under the CC BY-SA License (https://creativecommons.org/licenses/by-sa/4.0)

## 1. Introduction

Human development disparities across regions require analytical approaches capable of capturing spatial inequality to support effective policy formulation [1], [2]. In Indonesia, the Human Development Index (HDI) remains a key indicator of well-being, reflecting achievements in health, education, and income. East Java, as one of the most populous provinces, exhibits notable spatial variation in HDI across its districts and cities, emphasizing the need for spatially aware

***Corresponding author**. E-mail: friansyahgani21@student.ub.ac.id

Submitted: December 03, 2025 Reviewed: December 19, 2025 Accepted: January 13, 2026
DOI: https://doi.org/10.18860/cauchy.v11i1.38007

modeling techniques [3], [4]. Such spatial variation poses challenges for conventional analysis and highlights the importance of localized approaches. Therefore, methods that can explicitly account for spatial heterogeneity are necessary to inform effective policy decisions.

Conventional regression models such as Ordinary Least Squares (OLS) assume constant relationships between variables over space [5]. This assumption is often unrealistic in geographically diverse regions where socio-economic conditions differ significantly [6]. As a result, global models may overlook localized dynamics, yielding less accurate interpretations of the spatial structure and component-wise variation of HDI [7]. Consequently, there is a need for methods that incorporate spatially varying effects to better capture local variations. These methods can provide more nuanced insights into regional development disparities.

Geographically Weighted Regression (GWR) provides a framework to account for spatial non-stationarity by allowing regression coefficients to vary across locations [8], [9]. Using the Tricube kernel, GWR assigns higher weights to nearby observations while progressively reducing the influence of more distant ones, making it suitable for socio-economic phenomena with smooth spatial gradients [10], [11]. However, GWR's performance strongly depends on optimal bandwidth selection, which may be difficult to obtain using conventional numerical methods. This limitation motivates the integration of metaheuristic optimization techniques. By optimizing bandwidths, GWR can more accurately capture local variations and improve model performance.

To address this limitation, metaheuristic algorithms such as the Flower Pollination Algorithm (FPA) have been employed to optimize GWR bandwidths and kernel parameters. FPA utilizes global search via Lévy flights and local refinement through neighborhood-based pollination, enabling it to escape local minima and produce more stable solutions [12], [13]. Integrating FPA with GWR under a Tricube kernel results in a hybrid model that improves both spatial adaptivity and predictive accuracy. This approach allows the model to capture complex spatial patterns more reliably. Therefore, GWR–FPA provides a robust framework for spatially varying regression analysis.

It should be noted that some explanatory variables used in this study, particularly literacy rate and mean years of schooling, are structural components of the Human Development Index. Accordingly, the modeling framework is not intended to establish causal relationships, but rather to examine the spatial decomposition and relative contribution of HDI-related components across regions. This conceptual limitation is explicitly acknowledged, and all interpretations are restricted to explanatory and descriptive insights rather than causal inference. Nevertheless, the findings can inform evidence-based planning and highlight regional differences in development components. Therefore, this study focuses on descriptive and explanatory insights without making causal claims.

This study applies the GWR–FPA Tricube approach to model spatial variations in HDI across East Java. The research aims to (1) examine the spatial heterogeneity in the contribution of HDI-related components, (2) evaluate the performance of OLS, standard GWR, and optimized GWR–FPA using AIC-based model selection, and (3) generate a more accurate spatial representation of HDI disparities. The findings are expected to support region-specific development planning and contribute to the advancement of hybrid spatial–metaheuristic modeling in socio-economic analysis. By combining spatial modeling with metaheuristic optimization, this study provides a practical tool for regional policy analysis. Therefore, this research not only analyzes HDI disparities but also demonstrates the value of hybrid modeling approaches in socio-economic studies.

It should be emphasized that the present study focuses on educational components of the Human Development Index (HDI), specifically literacy rate and mean years of schooling, as key predictors of HDI spatial variation. While HDI is traditionally calculated from health, education, and income indices, health- and income-related indicators were excluded from the model for two main reasons. First, these factors tend to exhibit less spatial variability at the district level in East Java compared to education, which is more sensitive to local policy and socio-economic conditions. Second, the objective of this study is to examine the relative contribution and spatial decomposition of education-related components, rather than modeling all HDI determinants comprehensively. By concentrating on education, the analysis can more clearly reveal the dominant spatial patterns and local effects influencing HDI across districts. Therefore, despite the exclusion of health and economic factors, this research remains relevant for understanding regional disparities in human development and for informing targeted educational policies.

## 2. Methods

This research employs secondary data obtained from the Central Statistics Agency (*Badan Pusat Statistik*, BPS) of East Java Province. The dataset consists of one response variable, namely the Human Development Index (HDI), and three predictor variables ($X_1$, $X_2$, $X_3$), which represent socioeconomic and demographic factors influencing HDI in 2024. Specifically, $X_1$ denotes the total population in each district or city, $X_2$ represents the literacy rate as the percentage of the population aged 15 years and above who are literate, and $X_3$ indicates the mean years of schooling, reflecting the average number of years of formal education completed by the population. The analysis covers all regencies and cities in East Java Province, totaling 38 administrative units.

**Table 1:** Research Variables

| Variable | Symbol | Unit | Source (Year) |
|---|---|---|---|
| Human Development Index | $Y$ | Index | BPS (2024) |
| Population Size | $X_1$ | Persons | BPS (2024) |
| Literacy Rate | $X_2$ | % | BPS (2024) |
| Mean Years of Schooling | $X_3$ | Years | BPS (2024) |


**Figure 1:** Study Area (East Java Province)

The analytical method used in this study is Geographically Weighted Regression optimized using the Flower Pollination Algorithm (GWR–FPA). The research procedure is as follows:

1. **Data preparation**: input the 2024 HDI dataset of East Java Province, construct geographical distance and spatial weight matrices (Tricube kernel), and standardize predictors ($X_1$, $X_2$, $X_3$).

2. **Preliminary diagnostics**: test spatial autocorrelation (Moran's I) and heteroskedasticity (Breusch–Pagan) on OLS residuals. Apply GWR if spatial dependence or heterogeneity is detected.
3. **Model and FPA setup**:
   (a) Define GWR kernel and bandwidth search range.
   (b) Set FPA parameters: population size $N$, scaling factor $\gamma$, Lévy parameter $\lambda$, switch probability $p$, max iterations.
   (c) Use corrected AIC (AICc) as fitness function to optimize bandwidth.
4. **Optimization process**: update candidate bandwidths via FPA. Stop when convergence or max iterations are reached.
5. **Final model and interpretation**: estimate GWR with optimal bandwidth, evaluate using AICc, and map local coefficients $\beta_k(u_i, v_i)$ across 38 districts to examine spatial variation in HDI determinants.

## 2.1. Spatial Test

### *2.1.1. Multicollinearity Diagnostic*

To assess whether the independent variables exhibit multicollinearity, the *Variance Inflation Factor* (VIF) was employed [14]. VIF measures the degree to which the variance of a regression coefficient is inflated due to linear correlation among predictors. For a given predictor $X_k$, the VIF is defined as:

$$\mathrm{VIF}(X_k) = \frac{1}{1 - R_k^2}, \tag{1}$$

where $R_k^2$ represents the coefficient of determination obtained by regressing $X_k$ on all other independent variables. A VIF value greater than 10 indicates severe multicollinearity, while values below this threshold suggest that multicollinearity is not a concern. This test ensures that the predictors included in the model are not excessively correlated with one another.

### *2.1.2. Spatial Autocorrelation Test Using Moran's I*

Spatial autocorrelation in the regression residuals was examined using *Moran's I*, a widely used global spatial statistic that determines whether values are spatially clustered, dispersed, or randomly distributed [15]. Moran's I is formulated as:

$$I = \frac{n}{W} \cdot \frac{\sum_i \sum_j w_{ij}(e_i - \bar{e})(e_j - \bar{e})}{\sum_i (e_i - \bar{e})^2}, \tag{2}$$

where $n$ is the number of spatial units, $w_{ij}$ is the spatial weight between locations $i$ and $j$, $e_i$ is the residual at location $i$, $\bar{e}$ is the mean residual, and $W = \sum_i \sum_j w_{ij}$. A hypothesis test is performed using the standardized statistic:

$$Z_I = \frac{I - \mathbb{E}[I]}{\sqrt{\mathrm{Var}(I)}}, \tag{3}$$

where a $p$-value greater than 0.05 indicates the absence of spatial autocorrelation. This test ensures that the independence assumption of regression is not violated due to spatial effects.

### *2.1.3. Breusch–Pagan Heteroscedasticity Test*

To evaluate whether the variance of the residuals is constant, the *Breusch–Pagan (BP)* test was applied. This test detects heteroskedasticity by regressing the squared residuals on the independent variables [16]. The BP statistic is defined as:

$$\mathrm{BP} = \sum_{i=1}^{n} \hat{u}_i^2 z_i^\top \left(Z^\top Z\right)^{-1} z_i, \tag{4}$$

where $\hat{u}_i$ denotes the residuals of the regression model, and $Z$ represents the matrix of independent variables. Under the null hypothesis of homoskedasticity, the BP statistic follows a chi-square distribution with degrees of freedom equal to the number of predictors. A $p$-value greater than 0.05 leads to the acceptance of the null hypothesis, indicating that the model does not suffer from heteroskedasticity.

## 2.2. Flower Pollination Algorithm

The Flower Pollination Algorithm (FPA) represents a metaheuristic optimization approach inspired by the biological process of flower pollination [17]. As in nature, the reproductive success of a flower species depends on pollination [18]. In this method, the algorithm distinguishes between two mechanisms: global pollination and local pollination [19].The global pollination mechanism of the Flower Pollination Algorithm (FPA) is an update rule used to generate new candidate solutions through Lévy flight-based exploration. The update equation is defined as follows:

$$\theta_i^{t+1} = \theta_i^t + \gamma L(\lambda) \left(\theta^* - \theta_i^t\right) \tag{5}$$

The local pollination mechanism represents a neighborhood-based update rule that exploits local solution differences to refine candidate solutions. The local update equation is given by:

$$\theta_i{}^{t+1} = \theta_i{}^t + \epsilon \left(\theta_j{}^t - \theta_k{}^t\right) \tag{6}$$

The Flower Pollination Algorithm (FPA) updates candidate solutions (flowers) using four main mechanisms. **Global Pollination** allows exploration of the search space by moving solutions towards the global best with Lévy-distributed steps. **Local Pollination** enables exploitation by adjusting solutions based on the difference between two randomly selected candidates. **Selection Rule** ensures that a candidate solution is updated only if it improves the objective function. Finally, **Best Solution Update** stores the best solution found so far, guiding the algorithm towards the optimal solution.

In this study, the Flower Pollination Algorithm (FPA) optimizes the bandwidth parameter of the Geographically Weighted Regression (GWR) model by minimizing the corrected Akaike Information Criterion (AICc). The objective (fitness) function is defined as:

$$f(\theta) = \mathrm{AICc}(\theta), \tag{7}$$

where $\theta$ denotes the GWR bandwidth parameter.

## 2.3. Geographically Weighted Regression (GWR)

Geographically Weighted Regression (GWR) is a local regression technique used to model spatially varying relationships between a response variable and a set of predictor variables [20]. Unlike global linear regression, which assumes that model parameters are constant across all locations, GWR allows regression coefficients to vary spatially, enabling the model to capture local heterogeneity [21].

Let $y_i$ denote the response variable at location $i$, and let $(u_i, v_i)$ represent the geographical coordinates of that location. The GWR model is expressed as:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ki} + \varepsilon_i, \tag{8}$$

where $\beta_k(u_i, v_i)$ are location-specific regression coefficients and $\varepsilon_i$ is the random error term.

The local parameter estimates are obtained using weighted least squares. The estimator of the local regression coefficients at location $(u_i, v_i)$ is given by:

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = \left(\mathbf{X}^\top \mathbf{W}(u_i, v_i)\mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{W}(u_i, v_i)\mathbf{y}, \tag{9}$$

where $\mathbf{X}$ is the matrix of predictor variables with dimension $n \times p$, where $n$ is the number of locations and $p$ is the number of predictors; $\mathbf{y}$ is the response vector of dimension $n \times 1$; and $\mathbf{W}(u_i, v_i)$ is a diagonal spatial weighting matrix of dimension $n \times n$, with the weights $w_{ij}$ on its diagonal.

The bandwidth $b$ in Geographically Weighted Regression (GWR) determines the spatial extent over which observations influence the local parameter estimates. If not optimized, the bandwidth is usually selected based on the data, either by using a fixed distance (e.g., a set number of nearest neighbors or a predefined geographic distance) or by following a rule-of-thumb method. However, to obtain accurate and reliable GWR estimates, it is recommended to optimize the bandwidth by minimizing a criterion such as the Akaike Information Criterion corrected (AICc) or the Cross-Validation (CV) score.

In cases where $b$ has not been optimized, it is obtained from either a default value or a predefined rule applied to the dataset, rather than from a specific equation.

### *2.3.1. Bandwidth Selection Using FPA and Tricube Kernel Function*

In this study, the bandwidth parameter of the GWR model is optimized using the Flower Pollination Algorithm (FPA) with the objective of minimizing the corrected Akaike Information Criterion (AICc). The optimal bandwidth $b^*$ is obtained as:

$$b^* = \arg\min_{b} \ \mathrm{AICc}(b) \tag{10}$$

Once the optimal bandwidth is determined, the Tricube kernel function is used to construct spatial weights:

$$w_{ij}(b) = \begin{cases} \left(1 - \left(\dfrac{d_{ij}}{b}\right)^3\right)^3, & \text{if } d_{ij} < b, \\ 0, & \text{if } d_{ij} \geq b, \end{cases} \tag{11}$$

where $d_{ij}$ denotes the Euclidean distance between locations $i$ and $j$.

This procedure applies exclusively to the GWR–FPA model. For comparison, the standard GWR model uses a pre-specified bandwidth determined from data characteristics, ensuring a clear separation of bandwidth selection procedures between the two models.

### *2.3.2. Model Selection Using AIC and AICc*

Model selection is conducted using the Akaike Information Criterion (AIC) and its corrected form AICc [22]. These criteria evaluate model quality by balancing goodness-of-fit and model complexity [23]. The mathematical formulations are:

$$AIC = -2\ln(\hat{L}) + 2k \tag{12}$$

$$AICc = AIC + \frac{2k(k+1)}{n-k-1} \tag{13}$$

where $\hat{L}$ is the log-likelihood, $k$ is the number of estimated parameters, and $n$ is the sample size.

The criteria for selecting the best model are summarized in Table 2.

**Table 2:** Model Selection Criteria Based on AIC/AICc

| **Criterion** | **Description** |
|---|---|
| Lower AIC | Indicates better model fit |
| Lower AICc | Preferred for small sample sizes |
| Penalty Term | Avoids overfitting by penalizing complex models |
| Best Model Rule | Model with the smallest AICc is selected |

Based on this criterion, the model with the minimum AICc value is selected as the best-performing model among all competing specifications.

## 3. Results and Discussion

**Lemma 3.1** (Derivation of the Objective Function of GWR for FPA Optimization)**.** *Let $\theta$ denote the bandwidth parameter of the Geographically Weighted Regression (GWR) model. For each spatial location $i$, the local regression model is defined as:*

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ki} + \varepsilon_i. \tag{14}$$

*The weighted least squares estimator of the local coefficients is:*

$$\hat{\boldsymbol{\beta}}(u_i, v_i; \theta) = (X^\top W(u_i, v_i; \theta) X)^{-1} X^\top W(u_i, v_i; \theta)\mathbf{y}. \tag{15}$$

*The fitted values of the GWR model can be written using the smoothing matrix:*

$$\hat{\mathbf{y}}(\theta) = S(\theta)\mathbf{y}, \qquad S(\theta) = X(X^\top W(\theta) X)^{-1} X^\top W(\theta). \tag{16}$$

*The residual vector is:*

$$\mathbf{e}(\theta) = (I - S(\theta))\mathbf{y}. \tag{17}$$

*The estimated error variance is given by:*

$$\hat{\sigma}^2(\theta) = \frac{\mathbf{e}(\theta)^\top \mathbf{e}(\theta)}{n - tr(S(\theta))}. \tag{18}$$

*Assuming normal errors, the log-likelihood becomes:*

$$\ln L(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\hat{\sigma}^2(\theta)\right) - \frac{1}{2\hat{\sigma}^2(\theta)}\mathbf{e}(\theta)^\top \mathbf{e}(\theta). \tag{19}$$

*The corrected Akaike Information Criterion for GWR is:*

$$AICc(\theta) = -2\ln L(\theta) + 2\,tr(S(\theta)) + \frac{2\,tr(S(\theta))\,(tr(S(\theta)) + 1)}{n - tr(S(\theta)) - 1}. \tag{20}$$

*Thus, the objective function to be minimized by the Flower Pollination Algorithm (FPA) is:*

$$f(\theta) = AICc(\theta). \tag{21}$$

It should be noted that in Geographically Weighted Regression (GWR), the weighting matrix is location-specific. For each spatial location $i$, a local weighting matrix $W_i(b)$ is constructed based on distances from location $i$. The overall smoothing matrix $S(b)$ is therefore formed by stacking the local hat vectors corresponding to each $W_i(b)$, rather than arising from a single global weighting matrix. The trace of $S(b)$ represents the effective number of parameters in the GWR model, following the standard formulation of GWR [10].

**Lemma 3.2** (Integration of the GWR Objective Function into the Flower Pollination Algorithm)**.** *Let $\theta_i$ denote the $i$-th candidate solution (flower) in the Flower Pollination Algorithm. For each $\theta_i$, the value of the objective function is computed as:*

$$f(\theta_i) = AICc(\theta_i). \tag{22}$$

***Global Pollination.*** *The global update rule is:*

$$\theta_i^{t+1} = \theta_i^t + L\left(\theta^* - \theta_i^t\right), \tag{23}$$

*where $\theta^*$ is the global best solution and $L$ is a Lévy-distributed step size:*

$$L \sim \frac{\lambda\Gamma(\lambda)\sin(\pi\lambda/2)}{\pi s^{1+\lambda}}. \tag{24}$$

***Local Pollination.*** *The local update rule is:*

$$\theta_i^{t+1} = \theta_i^t + \epsilon(\theta_j^t - \theta_k^t), \qquad \epsilon \sim U(0,1). \tag{25}$$

***Selection Rule.***

$$\theta_i^{t+1} = \begin{cases} \theta_i^{t+1}, & \text{if } f(\theta_i^{t+1}) < f(\theta_i^t), \\ \theta_i^t, & \text{otherwise.} \end{cases} \tag{26}$$

***Best Solution Update.***

$$\theta^* = \arg\min_{\theta_i} f(\theta_i). \tag{27}$$

Lemma 2 describes the mechanism for updating candidate solutions in the Flower Pollination Algorithm (FPA) based on their performance according to the GWR objective function. Specifically, each candidate solution (flower) $\theta_i$ is evaluated using the objective function $f(\theta_i) = AICc(\theta_i)$, which measures the goodness-of-fit of the GWR model with the given parameters.

The lemma establishes a selection rule whereby a candidate solution is updated only if the new position leads to an improvement in the objective function. This ensures that the search process consistently moves toward better solutions. Additionally, the global best solution $\theta^*$ is tracked and updated whenever a better candidate is found.

Thus, Lemma 2 is necessary to guarantee that the optimization process converges toward an optimal or near-optimal set of parameters for the GWR model, ensuring accurate local regression estimates while efficiently exploring the solution space using FPA.

## 3.1. Multicollinearity Test

### *3.1.1. Multicollinearity Test*

Multicollinearity is assessed using the Variance Inflation Factor (VIF). A VIF value greater than 10 generally indicates severe multicollinearity, while values between 5 and 10 suggest moderate multicollinearity.

**Table 3:** Variance Inflation Factor (VIF) Results

| Variable | VIF |
|---|---|
| $X_1$ | 2.837 |
| $X_2$ | 2.386 |
| $X_3$ | 4.946 |

**Interpretation.** The VIF values for all three independent variables ($X_1$, $X_2$, and $X_3$) are below the threshold of 10, and also below the moderate multicollinearity level of 5. This indicates that multicollinearity is not present in the model. Therefore, the independent variables do not exhibit strong linear relationships with one another, and the regression coefficients can be interpreted reliably without inflation due to collinearity.

## 3.2. Statistical Assumption Testing

### *3.2.1. Spatial Autocorrelation Test Using Moran's I*

The Moran's I test was conducted on the residuals of the OLS model to examine the presence of spatial dependence. The test produced a $p$-value of 0.012, which is less than the 0.05 significance level. Since $0.012 < 0.05$, the null hypothesis of spatial randomness is rejected, indicating the presence of significant spatial autocorrelation in the OLS residuals. This result suggests that the independence assumption of the global OLS model is violated, reflecting spatial dependence among neighboring observations.

It should be emphasized that Moran's I specifically detects spatial dependence in the residuals and does not directly indicate spatial non-stationarity of regression relationships. Therefore, this diagnostic is interpreted as evidence of spatial dependence rather than as a direct justification for the use of Geographically Weighted Regression (GWR).

### *3.2.2. Heteroskedasticity Test Using the Breusch–Pagan Test*

The Breusch–Pagan test was conducted to examine the homoskedasticity assumption of the regression residuals. The test yielded a $p$-value of 0.032, which is below the 0.05 significance level. Since $0.032 < 0.05$, the null hypothesis of homoscedasticity is rejected, indicating the presence of heteroskedasticity in the regression residuals. This result suggests that the variance of the errors is not constant across observations, violating a key assumption of the global Ordinary Least Squares (OLS) model.

The detected heteroscedasticity provides empirical justification for employing spatially adaptive models such as Geographically Weighted Regression (GWR), which are better suited to capture spatial heterogeneity and location-specific variance structures.

### *3.2.3. Parameter determination*

The implementation of the GWR–FPA model in this study involved determining several key model parameters, summarized as follows:

1. **GWR parameters**.
   (a) **Standard GWR:** Optimal bandwidth $b_{\text{GWR}} = 320{,}000$ was determined using Cross-Validation.
   (b) **GWR–FPA:** Optimal bandwidth $b_{\text{FPA}} = 393{,}560.6$ was obtained using the Flower Pollination Algorithm (FPA) with AICc as the objective function.
2. **FPA parameters**. The Flower Pollination Algorithm was executed using the following settings: scaling factor $\gamma = \mathbf{0.1}$, Lévy parameter $\lambda = \mathbf{1.5}$, global pollination probability $p = \mathbf{0.8}$, number of agents = **20**, and maximum iterations = **100**. These values ensured stable optimization in the bandwidth search process.
3. **Model evaluation criteria**. The corrected Akaike Information Criterion (AICc) was used as the optimization objective for the Flower Pollination Algorithm (FPA). The GWR–FPA model produced improved AIC and AICc values compared to the global OLS and standard GWR models (see Table 7), indicating enhanced model fit and effective capture of spatial heterogeneity.

## 3.3. GWR-FPA

### *3.3.1. GWR Local Regression Equations for All Districts/Cities*

All variables in the GWR model, including both the response (HDI, $Y$) and predictors ($X$), are standardized. Consequently, the local GWR coefficients represent the relative influence of each predictor on HDI in standardized units, allowing for direct comparison across variables and locations. The local intercept corresponds to the standardized HDI when all predictors are at their mean levels.

**Table 4:** GWR Local Regression Equations for All Districts/Cities in East Java

| **District/City** | **GWR-FPA Equation** |
|---|---|
| Pacitan | $\hat{Y} = 1.07 + 0.58X_1 + 0.96X_2 + 2.15X_3$ |
| Ponorogo | $\hat{Y} = 4.44 + 0.53X_1 + 0.97X_2 + 2.16X_3$ |
| Trenggalek | $\hat{Y} = 7.02 + 0.49X_1 + 0.99X_2 + 2.16X_3$ |
| Tulungagung | $\hat{Y} = 9.35 + 0.46X_1 + 1.00X_2 + 2.16X_3$ |
| Blitar | $\hat{Y} = 12.08 + 0.42X_1 + 1.01X_2 + 2.17X_3$ |
| Kediri | $\hat{Y} = 9.58 + 0.46X_1 + 1.00X_2 + 2.17X_3$ |
| Malang | $\hat{Y} = 14.16 + 0.39X_1 + 1.01X_2 + 2.18X_3$ |
| Lumajang | $\hat{Y} = 15.86 + 0.37X_1 + 1.00X_2 + 2.20X_3$ |
| Jember | $\hat{Y} = 18.18 + 0.33X_1 + 1.01X_2 + 2.22X_3$ |
| Banyuwangi | $\hat{Y} = 22.12 + 0.27X_1 + 1.05X_2 + 2.24X_3$ |
| Bondowoso | $\hat{Y} = 18.04 + 0.34X_1 + 1.01X_2 + 2.24X_3$ |
| Situbondo | $\hat{Y} = 17.37 + 0.34X_1 + 1.00X_2 + 2.25X_3$ |
| Probolinggo | $\hat{Y} = 15.38 + 0.38X_1 + 0.99X_2 + 2.21X_3$ |
| Pasuruan | $\hat{Y} = 13.27 + 0.41X_1 + 0.99X_2 + 2.19X_3$ |
| Sidoarjo | $\hat{Y} = 11.65 + 0.43X_1 + 0.99X_2 + 2.19X_3$ |
| Mojokerto | $\hat{Y} = 10.89 + 0.44X_1 + 0.99X_2 + 2.18X_3$ |
| Jombang | $\hat{Y} = 9.52 + 0.46X_1 + 0.99X_2 + 2.18X_3$ |
| Nganjuk | $\hat{Y} = 9.75 + 0.45X_1 + 1.00X_2 + 2.16X_3$ |
| Madiun | $\hat{Y} = 4.52 + 0.53X_1 + 0.97X_2 + 2.17X_3$ |
| Magetan | $\hat{Y} = 1.27 + 0.58X_1 + 0.95X_2 + 2.16X_3$ |
| Ngawi | $\hat{Y} = -0.21 + 0.60X_1 + 0.93X_2 + 2.17X_3$ |
| Bojonegoro | $\hat{Y} = 4.16 + 0.53X_1 + 0.96X_2 + 2.18X_3$ |
| Tuban | $\hat{Y} = 3.09 + 0.55X_1 + 0.93X_2 + 2.18X_3$ |
| Lamongan | $\hat{Y} = 7.79 + 0.48X_1 + 0.97X_2 + 2.18X_3$ |
| Gresik | $\hat{Y} = 8.30 + 0.48X_1 + 0.95X_2 + 2.20X_3$ |
| Bangkalan | $\hat{Y} = 11.37 + 0.43X_1 + 0.97X_2 + 2.20X_3$ |
| Sampang | $\hat{Y} = 12.97 + 0.41X_1 + 0.97X_2 + 2.21X_3$ |
| Pamekasan | $\hat{Y} = 14.04 + 0.40X_1 + 0.97X_2 + 2.23X_3$ |
| Sumenep | $\hat{Y} = 17.49 + 0.34X_1 + 1.01X_2 + 2.27X_3$ |
| Kota Kediri | $\hat{Y} = 8.99 + 0.46X_1 + 0.99X_2 + 2.17X_3$ |
| Kota Blitar | $\hat{Y} = 11.43 + 0.43X_1 + 1.01X_2 + 2.17X_3$ |
| Kota Malang | $\hat{Y} = 13.45 + 0.40X_1 + 1.00X_2 + 2.18X_3$ |
| Kota Probolinggo | $\hat{Y} = 14.71 + 0.39X_1 + 0.99X_2 + 2.20X_3$ |
| Kota Pasuruan | $\hat{Y} = 13.26 + 0.41X_1 + 0.99X_2 + 2.19X_3$ |
| Kota Mojokerto | $\hat{Y} = 10.28 + 0.45X_1 + 0.99X_2 + 2.18X_3$ |
| Kota Madiun | $\hat{Y} = 3.26 + 0.55X_1 + 0.96X_2 + 2.17X_3$ |
| Kota Surabaya | $\hat{Y} = 11.18 + 0.44X_1 + 0.98X_2 + 2.19X_3$ |

$$\hat{Y}_{\text{Kota Batu}} = 12.29 + 0.42X_1 + 1.00X_2 + 2.18X_3 \tag{28}$$

The GWR local regression results for 38 districts and cities in East Java reveal substantial spatial variation in the effects of the predictor variables on the Human Development Index (HDI).

As shown in Eq. (28), the local model for Kota Batu indicates that population size ($X_1$), literacy rate ($X_2$), and mean years of schooling ($X_3$) all have positive effects on HDI. The coefficient of $X_1$ ranges from 0.273 (Banyuwangi) to 0.598 (Ngawi), indicating substantial local heterogeneity across districts. In contrast, $X_2$ exhibits relatively limited spatial variation, with coefficients ranging from 0.934 to 1.050. Meanwhile, $X_3$ shows the strongest and most stable influence on HDI, with coefficient values between 2.147 and 2.273, highlighting its dominant role in explaining HDI variation. Negative coefficient values, although not observed in this study, would indicate an inverse relationship with HDI, whereas positive coefficients—as illustrated in Eq. (28) reflect a direct positive contribution.

To further ensure that the spatial heterogeneity captured by GWR does not leave residual spatial dependence, Moran's I test was applied to the residuals. The results show a $p$-value greater than 0.05, indicating no significant spatial autocorrelation and confirming that GWR improves the error structure relative to the global OLS model.

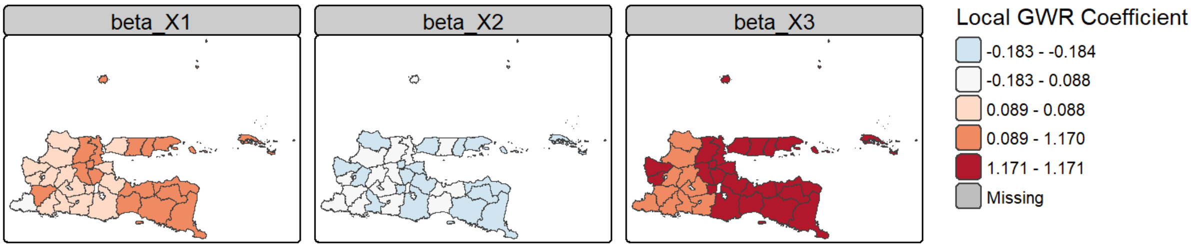


**Figure 2:** Spatial Distribution of Local GWR–FPA Coefficients

### *3.3.2. Modelling GWR–FPA*

The bandwidth value obtained using the fixed Tricube kernel corresponds to a single optimal bandwidth. The bandwidth was optimized using the Flower Pollination Algorithm (FPA) with the objective of minimizing the corrected Akaike Information Criterion (AICc). The resulting optimal bandwidth was $b = 393{,}560.6$, which was used uniformly across all observation locations. After determining the optimal bandwidth and constructing the spatial weighting matrix using the Tricube kernel, local regression parameters were estimated for each district/city. The distribution of the main GWR parameter estimates is summarized in Table 5.

**Table 5:** Distribution of GWR Parameter Estimates (Tricube Kernel)

| Variable | Min | Median | Max |
|---|---|---|---|
| $X_1$ | 0.273 | 0.435 | 0.598 |
| $X_2$ | 0.934 | 0.991 | 1.050 |
| $X_3$ | 2.147 | 2.183 | 2.273 |

Table 5 shows that the coefficient of $X_1$ is consistently positive across all districts/cities, ranging from 0.273 to 0.598. This indicates that increases in $X_1$ are consistently associated with improvements in the Human Development Index (HDI), although the strength of its influence varies spatially. The coefficient of $X_2$ is also consistently positive, ranging between 0.934 and 1.050, this indicates that higher literacy rates ($X_2$) are consistently associated with improvements in the Human Development Index (HDI). Meanwhile, $X_3$ exhibits the largest impact among all predictors, with coefficients ranging from 2.147 to 2.273. This implies that variations in $X_3$ produce the strongest improvement in HDI throughout the study area.

### *3.3.3. Significance Analysis of Local GWR Parameters*

Local inference for the GWR coefficients was conducted using location-specific $t$-statistics, computed as:

$$t_{ik} = \frac{\hat{\beta}_k(u_i, v_i)}{\mathrm{SE}(\hat{\beta}_k(u_i, v_i))}, \tag{29}$$

where $\hat{\beta}_k(u_i, v_i)$ is the local coefficient for predictor $k$ at location $i$, and $\mathrm{SE}(\hat{\beta}_k(u_i, v_i))$ is the corresponding local standard error.

A local coefficient is classified as statistically meaningful if $|t_{ik}| > 2$, following a descriptive guideline for exploratory spatial analysis. No formal multiple comparison adjustment was applied.

**Table 6:** Summary of Local Significance for GWR–FPA Parameters across 38 Districts/Cities in East Java

| **Parameter** | **Number of Districts/Cities with Significant Local Effect** |
|---|---|
| Intercept | 38 / 38 |
| $X_1$ (Population Size) | 3 / 38 |
| $X_2$ (Literacy Rate) | 38 / 38 |
| $X_3$ (Mean Years of Schooling) | 38 / 38 |

**Interpretation.** The intercept is significant in all districts/cities. $X_1$ (population size) shows significant effects in only 3 out of 38 districts/cities, indicating **localized influence** in a few areas while being negligible elsewhere. This pattern highlights spatial heterogeneity, suggesting that the impact of population size on HDI is **context-dependent**, possibly driven by regional characteristics. Meanwhile, $X_2$ (literacy rate) and $X_3$ (mean years of schooling) are consistently significant across all districts/cities, confirming their dominant and spatially stable contribution to HDI variation.

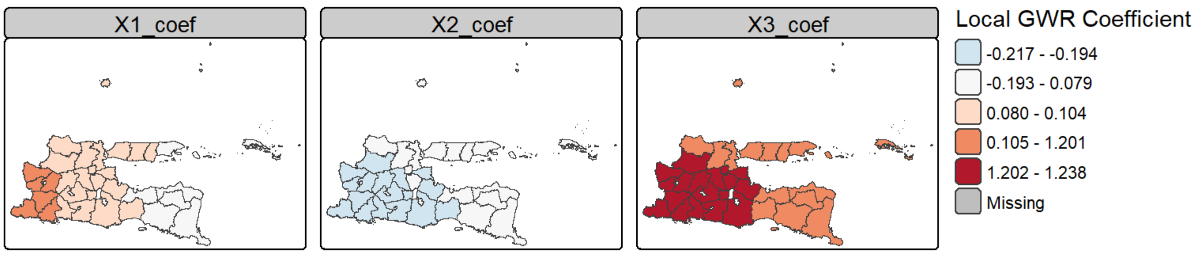


**Figure 3:** Spatial Distribution of Local GWR Coefficients

### *3.3.4. Best Model Selection*

The selection of the best model between the OLS regression and various specifications of the *Geographically Weighted Regression* (GWR) model was carried out to determine the most appropriate model for the characteristics of the data. The main criteria used were the *Akaike Information Criterion* (AIC) and AICc, where lower values indicate a better model in terms of balancing goodness-of-fit and model complexity.

Table 7 presents a comparison of the AIC and AICc values for the OLS model, the GWR–Tricube model, and the GWR–FPA–Tricube model.

**Table 7:** Comparison of AIC Values for OLS, GWR, and GWR–FPA Models

| **Model** | **AIC** | **AICc** |
|---|---|---|
| OLS Regression (Global) | 138.0064 | 138.0079 |
| GWR–Tricube | 136.9215 | 140.9693 |
| GWR–FPA–Tricube | **135.8821** | **137.0045** |

Based on Table 7, Although the AIC value of the GWR–Tricube model is numerically close to that of the global OLS model, the corrected criterion (AICc) increases due to the larger effective number of parameters in GWR, as reflected by the trace of the GWR smoothing matrix. This highlights the importance of using AICc rather than AIC when comparing global and local models. However, its AICc value increases due to the penalty for model complexity, resulting in a slight decrease in overall performance. In contrast, the GWR–FPA–Tricube model generates the lowest AIC and AICc values among all models evaluated. Therefore, the GWR–FPA–Tricube model is identified as the best-performing model and the most appropriate for capturing the spatial variation of the Human Development Index (HDI) in East Java Province. For the GWR models, the parameter count used in AIC/AICc computation corresponds to the effective number of parameters, approximated by the trace of the GWR smoothing matrix, following standard GWR model selection practice.

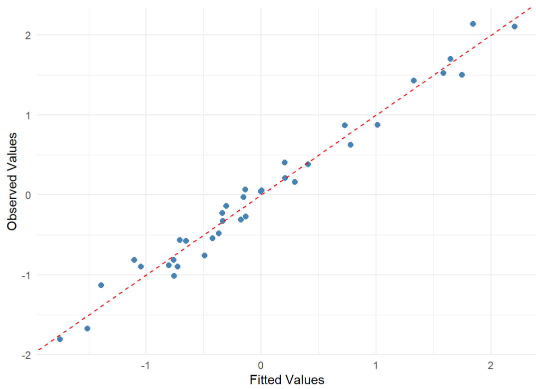


**Figure 4:** Observed vs Fitted Values

Table 7 presents the model comparison results. It should be noted that standard GWR utilized Cross-Validation for bandwidth selection, whereas the GWR–FPA model applied FPA optimization targeting AICc minimization. This separation ensures reproducibility and clarifies the optimization strategy for each model.

#### *3.3.5. Post-GWR Residual Diagnostics*

To evaluate whether the GWR–FPA model successfully removed spatial dependence, Moran's I test was conducted on the residuals of the fitted GWR model. A Tricube spatial weighting scheme was employed, consistent with the model kernel. The Moran's I value of 0.028 with a p-value of 0.248 indicates that the residuals of the GWR–FPA model do not exhibit significant spatial autocorrelation. This confirms that the model adequately accounts for spatial heterogeneity in HDI across East Java. The Tricube-based weighting scheme ensures that nearby observations are weighted more heavily, effectively capturing local spatial variation. Both statistical (Moran's I) and visual diagnostics confirm that the GWR–FPA model residuals are spatially random. This implies that the model has successfully removed spatial dependence and is appropriate for modeling the spatial variation of HDI in the study area.

## 4. Conclusion

This study analyzed the spatial variation of the Human Development Index (HDI) in East Java using a Tricube-based Geographically Weighted Regression–Flower Pollination Algorithm (GWR–FPA). The results show clear spatial heterogeneity in the determinants of HDI: population size ($X_1$) was not significant in any district/city, while education-related variables—literacy rate ($X_2$) and mean years of schooling ($X_3$)—were consistently significant across all 38 regions. Model comparison based on AIC and AICc demonstrated that the GWR–FPA–Tricube model outperformed both global OLS and standard GWR, indicating superior optimization and spatial representation. Overall, the study successfully met its objectives by identifying spatially varying HDI determinants, confirming the superiority of the optimized GWR–FPA model, and These findings suggest that spatial disparities in HDI largely mirror disparities in education-related components, which may inform region-specific development prioritization within the existing HDI framework.

"The dataset and code analyzed during the current study are publicly available in the Badan

Pusat Statistik East Java 2024[1]."

## CRediT Authorship Contribution Statement

**Friansyah Gani**: Conceptualization, methodology, writing—original draft. **Henny Pramoedyo**: Supervision, validation, editing. **Achmad Efendi**: Supervision, validation, editing.

## Declaration of Generative AI and AI-assisted Technologies

The authors declare that no generative AI tools were used to generate or modify the data, results, or analysis of this study. AI tools were used only for grammar improvement and formatting.

## Declaration of Competing Interest

The authors declare no competing financial or personal interest.

## Funding and Acknowledgments

This research received no external funding. The authors thank BPS Jawa Timur for providing the dataset.

## Data and Code Availability

The data used in this study are publicly available from the Badan Pusat Statistik (BPS) at https://jatim.bps.go.id/id/publication/2025/05/27/47fde052cb353c601c21c209/indeks-pembangunan-manusia-provinsi-jawa-timur-2024.html. The code used for analysis in this study can be provided by the corresponding author upon reasonable request.

## References

[1] A. Cohen-Solal et al., "Geographic variation and human development index in heart failure with reduced ejection fraction: Insights from victoria," *JACC: Heart Failure*, vol. 13, no. 9, p. 102 548, 2025. DOI: 10.1016/j.jchf.2025.102548. https://www.sciencedirect.com/science/article/pii/S2213177925004755.

[2] A. Al-Maadid, M. S. Ben Ali, and K. Si Mohammed, "The effect of climate risk on the human development index using the panel time-varying interactive fixed effects," *Environmental and Sustainability Indicators*, vol. 27, p. 100 757, 2025. DOI: 10.1016/j.indic.2025.100757. https://www.sciencedirect.com/science/article/pii/S2665972725001783.

[3] F. Jiang, "Comment on "favorable breast cancer mortality-to-incidence ratios of countries with good human development index rankings and high health expenditures"," *Taiwanese Journal of Obstetrics and Gynecology*, vol. 64, no. 6, p. 1130, 2025. DOI: 10.1016/j.tjog.2025.07.027. https://www.sciencedirect.com/science/article/pii/S1028455925002797.

[4] B. Zulfiqar, L. Madureira, S. Abbas, F. Shahzad, and Z. Fareed, "Asymmetric impact of human development index on terrorism in pakistan: New findings from qardl," *Socio-Economic Planning Sciences*, vol. 100, p. 102 226, 2025. DOI: 10.1016/j.seps.2025.102226. https://www.sciencedirect.com/science/article/pii/S0038012125000758.

[5] J. Casellas et al., "Classification of light yorkshire pigs at different production stages using ordinary least squares and machine learning methods," *Animal*, vol. 18, no. 1, p. 101 047, 2024. DOI: 10.1016/j.animal.2023.101047. https://www.sciencedirect.com/science/article/pii/S1751731123003646.

[1] https://jatim.bps.go.id

[6] K. Tong, "Urbanization moderates the transitional linkages between energy resource use, greenhouse gas emissions, socio-economic and human development: Insights from subnational analyses in china," *Journal of Cleaner Production*, vol. 476, p. 143 776, 2024. DOI: 10.1016/j.jclepro.2024.143776. https://www.sciencedirect.com/science/article/pii/S0959652624032256.

[7] T. Goswami and S. Ghosal, "Domestic water poverty in a semi-arid district of eastern india: Multiple dimensions, regional pattern, and association with human development," *Environmental Development*, vol. 44, p. 100 742, 2022. DOI: 10.1016/j.envdev.2022.100742. https://www.sciencedirect.com/science/article/pii/S2211464522000446.

[8] M. d. R. Ahumada et al., "Spatial analysis of fasciola hepatica prevalence in sheep flocks from córdoba, argentina using gwr and idw models," *Veterinary Parasitology: Regional Studies and Reports*, vol. 66, p. 101 367, 2025. DOI: 10.1016/j.vprsr.2025.101367. https://www.sciencedirect.com/science/article/pii/S2405939025001753.

[9] S. Mondal, K. K. Gavsker, and B. Mandal, "Decoding spatial non-stationarity of urban heat island in a million-plus indian city: An integrated analysis harnessing gwr, mgwr, and geodetector models for urban climate resilience," *Advances in Space Research*, 2025. DOI: 10.1016/j.asr.2025.10.012. https://www.sciencedirect.com/science/article/pii/S0273117725011433.

[10] S. W. Tyas, Gunardi, and L. A. Puspitasari, "Geographically weighted generalized poisson regression model with the best kernel function in the case of postpartum maternal mortality in east java," *MethodsX*, vol. 10, p. 102 002, 2023. DOI: 10.1016/j.mex.2023.102002. https://www.sciencedirect.com/science/article/pii/S2215016123000079.

[11] R. Putra, M. G. Fadhlurrahman, and Gunardi, "Determination of the best knot and bandwidth in geographically weighted truncated spline nonparametric regression using generalized cross validation," *MethodsX*, vol. 10, p. 101 994, 2023. DOI: 10.1016/j.mex.2022.101994. https://www.sciencedirect.com/science/article/pii/S2215016122003685.

[12] M. Chu and W. Chen, "A bi-objective discrete flower pollination algorithm for planning the collaborative disassembly of retired power batteries by humans and robots," *Applied Soft Computing*, vol. 177, p. 113 213, 2025. DOI: 10.1016/j.asoc.2025.113213. https://www.sciencedirect.com/science/article/pii/S1568494625005241.

[13] Y. Hu, L. Qin, S. Li, X. Li, Y. Li, and W. Sheng, "Optimal chiller loading based on flower pollination algorithm for energy saving," *Journal of Building Engineering*, vol. 93, p. 109 884, 2024. DOI: 10.1016/j.jobe.2024.109884. https://www.sciencedirect.com/science/article/pii/S2352710224014529.

[14] R. Neog and B. Ghasemzadeh, "Drivers and dynamics of urban sprawl in dimapur, india (1994–2024): A gini, ueii, and geographically weighted regression-based assessment," *Advances in Space Research*, 2025. DOI: 10.1016/j.asr.2025.11.091. https://www.sciencedirect.com/science/article/pii/S0273117725013754.

[15] D. Zhang and Y. Zhang, "Moran's i of vrpad: A human activity-sensitive spatial pattern index for vegetation restoration evaluation," *Journal of Environmental Management*, vol. 387, p. 125 948, 2025. DOI: 10.1016/j.jenvman.2025.125948. https://www.sciencedirect.com/science/article/pii/S0301479725019243.

[16] B. Güloğlu, S. Taşpınar, O. Doğan, and A. K. Bera, "Testing homoskedasticity in spatial panel data models," *Econometrics and Statistics*, 2024. DOI: 10.1016/j.ecosta.2024.04.003. https://www.sciencedirect.com/science/article/pii/S2452306224000339.

[17] F. Gani, H. Pramoedyo, and A. Efendi, "Modeling fuzzy geographically weighted clustering with flower pollination algorithm for spatial optimization and clustering," *CAUCHY: Jurnal Matematika Murni dan Aplikasi*, vol. 10, no. 2, pp. 1205–1218, 2025. DOI: 10.18860/cauchy.v10i2.36800. https://doi.org/10.18860/cauchy.v10i2.36800.

[18] N. Chauhan, N. Kaur, K. S. Saini, S. Verma, R. Abu Khurma, and P. A. Castillo, "Maximizing resource efficiency in cloud data centers through knowledge-based flower pollination algorithm (kb-fpa)," *Computers, Materials and Continua*, vol. 79, no. 3, pp. 3757–3782, 2024. DOI: 10.32604/cmc.2024.046516. https://www.sciencedirect.com/science/article/pii/S1546221824000304.

[19] M. Abdel-Basset and L. A. Shawky, "Flower pollination algorithm: A comprehensive review," *Artificial Intelligence Review*, vol. 52, no. 4, pp. 2533–2557, 2019. DOI: 10.1007/s10462-018-9624-4.

[20] Z. Huisheng et al., "Exploring groundwater potential: Combining gis techniques with ols, gwr, and exploratory regression," *Journal of Hydrology: Regional Studies*, vol. 61, p. 102 564, 2025. DOI: 10.1016/j.ejrh.2025.102564. https://www.sciencedirect.com/science/article/pii/S2214581825003891.

[21] B. Mandal and K. P. Goswami, "Evaluating the influence of biophysical factors in explaining spatial heterogeneity of lst: Insights from brahmani-dwarka interfluve leveraging geodetector, gwr, and mgwr models," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 138, p. 103 836, 2025. DOI: 10.1016/j.pce.2024.103836. https://www.sciencedirect.com/science/article/pii/S1474706524002948.

[22] G. Flood-Page, L. Boutonnier, and J.-M. Pereira, "Application of the akaike information criterion to the interpretation of bender element tests," *Soil Dynamics and Earthquake Engineering*, vol. 177, p. 108 373, 2024. DOI: 10.1016/j.soildyn.2023.108373. https://www.sciencedirect.com/science/article/pii/S0267726123006188.

[23] S. Kudo, M. Fujimoto, T. Sato, and A. Nagano, "Determination of the optimal number of linked rigid-bodies of the trunk during walking and running based on akaike's information criterion," *Gait & Posture*, vol. 77, pp. 264–268, 2020. DOI: 10.1016/j.gaitpost.2020.02.009. https://www.sciencedirect.com/science/article/pii/S0966636220300709.