ISSN 2090-3359 (Print)
ISSN 2090-3367 (Online)

ΑΔΣ

Advances in Decision Sciences

Volume 28
Issue 2
June 2024

亞洲大學 ASIA UNIVERSITY

SCIENTIFIC & BUSINESS WORLD

Published by Asia University, Taiwan and Scientific and Business World

Exploring Geographical Variability in Sugarcane Yields: A Geographically Weighted Panel Regression Approach with MM Estimation

Yani Quarta Mondiana
Brawijaya University, Malang, Indonesia
***Corresponding author** Email: yqmondiana@gmail.com

Henny Pramoedyo
Brawijaya University, Malang, Indonesia
Email: hennyp@ub.ac.id

Atiek Iriany
Brawijaya University, Malang, Indonesia
Email: atiekiriany@ub.ac.id

Marjono
Brawijaya University, Malang, Indonesia
Email: marjono@ub.ac.id

Received: January 24, 2024; First Revision: February 21, 2024;

Last Revision: May 26, 2024; Accepted: May 30, 2024;

Published: December 31, 2024

Abstract

Purpose: This study aims to apply Geographically Weighted Panel Regression (GWPR) to panel data analysis, specifically to examine the influence of geographical variables and local variability on sugarcane yields in East Java. GWPR integrates the principles of panel regression with geographically weighted regression (GWR) analysis to capture varying relationships across different locations, considering panel fixed effects in its model**.** In the context of Decision Sciences, this research develops an innovative method for more accurate decision-making in the agricultural sector, taking into account geographical variability often overlooked in traditional decision models.

Design/Methodology/Approach: The study adopts a weighted least squares approach, sensitive to outliers, for parameter estimation within the GWPR model. This paper addresses the limitations of conventional analysis models, which often neglect the importance of location variability in data-driven decision-making. This approach is then applied to a dataset of sugarcane yields from East Java to assess how it can manage variability and outliers in the data.

Findings: The analysis reveals that the size of plantation areas plays a crucial role in determining sugarcane yields, with significant variability detected across locations in East Java. The study identifies other factors contributing to sugarcane yield variations, such as soil conditions, climate, and farming practices. This paper's contributions include the application of the GWPR methodology in agriculture, providing new insights and enriching the literature on the impact of geographical and local factors on agricultural yields.

Practical Implications: These findings have significant implications for developing agricultural strategies in East Java, particularly in the context of land management and resource allocation.

Originality/Value: This study is original because it integrates GWR methods into panel data analysis, providing a new analytical framework to accommodate geographical variability in panel data.

Keywords: GWPR, fixed effects, outliers, M-estimation, sugarcane yields, geographical variability.

JEL Classifications : C21, C23, C36, C52

1. Introduction

Classic regression, often known as ordinary least squares (OLS), is a statistical approach for modelling the relationship between one or more independent variables and a dependent variable. This approach seeks the best-fitting linear equation describing these variables' relationship.

Traditional regression models imply that the relationship among variables is continuous across the dataset, meaning the same coefficients apply to all observations. In some instances, varying responses across regions result in spatial heterogeneity. Evaluating data with spatial heterogeneity issues using regular linear regression models can lead to misinterpreted model parameters (Thompson, 2018). Several approaches have been proposed to address these issues, including robust estimation techniques and improvements to portfolio optimization theories (Bian**, et al.,** 2013; Wong**, et al.,** 2011). Additionally, quantile regression analysis can provide more accurate insights into economic relationships and growth (Guo**, et al.,** 2023). Other robust estimation models have also been developed to address issues in capital asset pricing model estimation (Wong & Bian, 2000)

In cases of spatial heterogeneity, Geographically Weighted Regression (GWR) can be used to model the relationship among variables. GWR accounts for the geographic location of data points and allows regression coefficients to vary spatially. This means the relationship among variables can differ in strength and direction as you move across different locations or regions. The GWR model estimates parameters using the Weighted Least Squares (WLS) method, which assigns weights to each observation point. Consequently, parameter values in the GWR model differ at each observation point (they are local).

Yu (2010) introduced Geographically Weighted Panel Regression (GWPR), a technique for analyzing panel data and identifying varying correlations. This method combines panel regression and geographically weighted regression (GWR). By allowing regression coefficients to fluctuate, GWPR considers regional variation. Various disciplines, including economics, agriculture, and environmental research, have extensively used this methodology. Weighted Least Squares is one of the approaches that can be used to estimate parameters in GWPR (Zhang & Mei, 2011). The WLS approach can be used to estimate regression model parameters if the error variance is not constant. Estimated parameters in the GWR model are also determined based on the WLS approach. The researcher defines weights used in the WLS method, and the sum of weights for each location is equal. In the GWR model, weights used in parameter estimation vary with distance. The sum of weights for each area in the GWR model varies. One weighting function is the Gaussian kernel used in this study. There is a value indicating the bandwidth in the Gaussian kernel weighting function. GWR model parameter estimates are not affected by the weighting type but by the bandwidth selection. Consequently, the first step in the parameter estimation process is determining the ideal bandwidth value for the kernel weighting function. Cross-validation (CV) is one approach to determining the optimal bandwidth value (Zhang & Mei, 2011).

One disadvantage of parameter estimation methods using WLS is their sensitivity to outliers (Liu**, et al.,** 2016). Outliers are observations in a dataset that have different patterns or values from other observations in that dataset. They are rare or unusual observations occurring at one extreme of the data. Extreme points in observations are values significantly different from most other values in the group, for example, values too small or too large. According to Weisberg (2005), outliers are data that do not follow the general or overall data pattern.

Outliers can affect the outcome of regression parameter estimation and lead to violations of the data's normality assumption. If outliers are caused by errors in recording observations or preparing equipment, they can be ignored or discarded before data analysis is conducted. One way to address outliers is by using robust estimation methods. Well-known robust estimation methods include M-estimation, S-estimation, and MM-estimation (Susanti**, et al.,** 2014). M-estimation is an extension of the maximum likelihood method and is a robust estimate, while S-estimation and MM-estimation are developments of the M-estimation method.

GWR modelling has produced robust parameter estimates. GWR parameters are estimated using Least Absolute Deviation (Harris**, et al.,** 2014; Zhang & Mei, 2011). Fotheringham**, et al.** (2002) introduced two techniques for strengthening GWR. The first approach involves removing samples, taking abnormally large residuals, and running GWR. Harris**, et al.** (2011) and Harris**, et al.** (2014) further described this practical approach to outlier detection. The second approach uses iterative GWR modelling to reduce the weight of data with large residuals. This technique of reducing weight has seen significant expansion. A Bayesian GWR with non-constant variance using priors to regularize (or downweight) outliers was proposed by LeSage (2004).

Geographically and Temporally Weighted Regression (GTWR) modelling was done using robust techniques with the S-estimator. According to study findings, robust GTWR modelling produced a superior model compared to GTWR. Erda and Djuraidah (2019) obtained the same findings as LeSage (2004) after modelling data with outliers using the M-estimator in GTWR. In this study, the MM method was used to estimate the fixed effect model parameters of GWPR.

Research gaps related to the integration of the Geographically Weighted Panel Regression (GWPR) model with robust estimator approaches, as explored in foundational research by Ningrum**, et al.** (2020), Yu**, et al.** (2021), and Xu**, et al.** (2017), and further supported by methodological advances in robust standard errors and spatial heterogeneity by Vogelsang (2012), Hoechle (2007), and Sugasawa and Murakami (2022), highlight a critical gap in current research efforts. These studies underscore a growing understanding of spatial dynamics in econometrics and the need for models to accommodate the inherent spatial, temporal, and cross-sectional heterogeneity in panel data. Specifically, Yu**, et al.** (2021) emphasized the subtle impact of high-speed rail systems on county development in China through GWPR analysis, showcasing spatially variable effects that traditional panel regression models might obscure. Similarly, Ningrum**, et al.** (2020) demonstrated the utility of GWPR in predicting village classification indices, reinforcing the model's applicability in various geographical contexts.

Research by Xu**, et al.** (2017) on CO_2 emissions in China's manufacturing industry, using the geographically weighted regression model, further elucidates geographic disparities in environmental impacts and the necessity for spatially aware econometric models. Vogelsang (2012) and Hoechle (2007) contribute to this discourse by addressing methodological challenges of heteroskedasticity, autocorrelation, and spatial correlation in linear panel models with fixed effects, thus laying the groundwork for more robust inference mechanisms in spatial econometrics.

Additionally, studies by Nilsson (2014), Naudé (2004), Rasekhi**, et al.** (2013), Thissen**, et al.** (2016), Xu and Huang (2015), and Ma**, et al.** (2021) extend the application of GWPR and related methodologies across various domains, including urban planning, economic growth, R&D investment spillovers, trade network competitiveness, spatial heterogeneity of accidents, and spatial economic data analysis. These contributions collectively identify a critical research gap: the need to integrate robust estimator approaches into the GWPR model to effectively capture and analyze the complex, multi-dimensional spatial and temporal variability characteristic of economic, environmental, and social phenomena. The emerging consensus points toward the development of more adaptive, resilient, and nuanced econometric models that can tackle the complexity of spatial heterogeneity, affirmed by pioneering work on robust adaptive GWPR by Sugasawa and Murakami (2022), marking a significant step forward in addressing these challenges.

Exploration of gaps in spatial econometrics literature has revealed a significant need for more dynamic approaches responsive to spatial and temporal heterogeneity and challenges posed by outliers in panel data analysis. This realization spurred the development of the research 'Geographically Weighted Panel Regression Fixed Effect Model: Robust Estimator Approach', focusing on integrating robust estimators into the GWPR framework to address existing methodological weaknesses. This approach aims to strengthen the model against outlier sensitivity and enhance precision in estimating geographically and temporally varying dynamics, showcasing the novelty of the research. This study is highly relevant to the field of Decision Sciences, as it provides advanced spatial–temporal modelling tools that support more informed and accurate decision-making in complex economic, environmental, and policy contexts. Unlike prior works that applied GWPR or robust estimators in isolation, this research uniquely integrates robust estimation methods (M, S, and MM) within the GWPR fixed effect framework, enabling improved resilience to outliers while simultaneously capturing spatial and temporal heterogeneity in panel data. By introducing robust estimators like M, S, and MM into GWPR, this study targets improved reliability and accuracy in parameter estimation while providing a more in-depth and detailed analysis of geographical and temporal effects. Hence, this research initiative directly addresses the identified gaps in the literature, contributing to the evolution of spatial econometrics by offering more sophisticated tools for understanding and analysing economic, social, and environmental phenomena within complex panel data frameworks.

2. Literature Review and Theoretical Framework

2.1 *Robust Regression with MM Estimation*

Box (1953) introduced the term "robust" to the statistical literature. The high sensitivity of several traditional statistical approaches to tiny assumptions violations, however, was initially identified by Tukey (1960). According to Chen (2002), robust parameter estimates in the regression model are:

(1) Developed by Huber (1973), M-estimation or M-estimate examines data assuming that most outliers are in the dependent variable. It is a type of robust regression that aims to provide stable results despite outliers. M-estimation functions by minimizing a loss function that is less sensitive to outliers than the traditional squared loss function used in ordinary least squares. An advanced model of this is discussed in the work by Combettes and Muller (2020),

where they explore the broad applicability of M-estimation through a perspective model that generalizes a large class of existing statistical models, including various robust estimators like Huber's (1973) concomitant M-estimator.

(2) Least Trimmed Squares (LTS) estimation is a method with a high breakdown point, as introduced by Rousseeuw and Yohai (1984). The breakdown point measures the smallest proportion of data contaminated by outliers compared to all observational data.

(3) Scale (S) estimation was developed by Rousseeuw and Yohai (1984). It has a high breakdown point and outperforms LTS at the same breakdown value.

(4) MM estimation (Method of Moments) was introduced by Yohai (1987). MM estimation is designed to offer a balance between robustness against outliers and statistical efficiency, making it superior to many other estimators in terms of handling diverse data conditions. It achieves this by starting with a robust initial estimate with a high breakdown point and refining it to improve efficiency, particularly under Gaussian distributions.

The MM estimation technique, as developed by Yohai in 1987, represents a significant advance in robust statistical methodologies. This method initially utilizes S estimation to handle the substantial presence of outliers by minimizing the scale of residuals. This initial robust estimate sets the stage for the subsequent application of M-estimation, which refines these estimates by optimizing them for efficiency, especially under conditions approximating normal distributions. This dual-phase approach ensures that the robustness to outliers is not compromised while still achieving a high level of statistical efficiency.

The utility and effectiveness of MM estimation are well documented in the literature. For instance, Susanti**, et al.** (2014) explored the practical applications of M, S, and MM estimations in robust regression. They highlight the method's balance between handling outlier-contaminated data and maintaining statistical integrity, providing a detailed discussion on the iterative algorithm that typifies MM estimation approaches. This discussion underscores the adaptability of MM estimation in managing datasets with significant anomalies, thus supporting its use in complex analytical scenarios.

Further extending the relevance of MM estimation, Özdemir and Arslan (2021) integrate this approach into the empirical likelihood methods for linear regression models. Their study showcases how MM estimation significantly outperforms traditional estimators, particularly in datasets where outliers impact both response and explanatory variables. This robustification through MM estimation enhances the reliability of parameter estimates, ensuring their viability across a broader range of real-world applications (Özdemir & Arslan, 2021),

Moreover, Kudraszow and Maronna (2011) proposed a class of robust estimates for multivariate linear models based on the MM estimation framework. They demonstrate the method's high breakdown point and asymptotic efficiency under Gaussian errors. Their findings, validated through both simulated and real data, illustrate the advantages of MM estimation over other estimation techniques, especially in terms of its robustness and efficiency in handling complex multivariate datasets.

These contributions collectively highlight MM estimation's crucial role in modern statistical analysis, particularly in applications where the presence of outliers challenges data integrity. This robust estimation technique mitigates the adverse effects of these outliers and ensures that the resulting statistical models are reliable and applicable across various research contexts.

The MM estimation technique involves first estimating the regression parameter using S estimation to minimize the residual scale from M estimation, followed by M estimation. MM estimation aims for high breakdown values and efficiency. Breakdown value measures how many outliers can be addressed before they affect the model. The MM estimator method takes the following form:

$$w_i(u_i) = \frac{\psi(u_i)}{u_i} = \begin{cases} \left[1 - \left(\frac{u_i}{c}\right)^2\right]^2 & , \quad |u_i| \le c; \\ 0 & , \quad |u_i| > c, \end{cases} \tag{1}$$

where $w_i(u_i)$ is the weight function, $\psi(u_i)$ is the psi function, u_i is the residual, and c is a constant.

2.2 *Geographically Weighted Regression (GWR)*

GWR is a geographic-based variant of global regression. According to Fotheringham**, et al.** (2002), the GWR model is a regression method that generates estimates of local model parameters for each point or place where data is collected. The response variable is predicted using explanatory factors, each with a regression coefficient that varies based on where the data is observed. The GWR model is expressed as follows:

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

where y_i represents the response variable at the i^th^ location, x_{ik} denotes the k^th^ predictor variable at the i^th^ location, $\beta_0(u_i, v_i)$ indicates the intercept model for the i^th^ location, and $\beta_k(u_i, v_i)$ signifies the k^th^ regression parameter for the i^th^ location. The coordinates for the i^th^ location are given by (u_i, v_i), and ε_i represents the i^th^ residual, which follows a normal distribution with mean 0 and variance σ^2, i.e., $\varepsilon_i \sim N(0, \sigma^2)$.

GWR modelling is performed on data with geographical heterogeneity. Spatial heterogeneity occurs when one region's nature and attributes differ from those of another. Ignoring spatial heterogeneity will yield inefficient estimation results and less acceptable conclusions. Spatial heterogeneity testing determines whether there is variation between places because each location has unique structures and interactions. The Breusch-Pagan test statistic (BP test) assesses spatial heterogeneity.

Geographically Weighted Regression (GWR) handles datasets with geographical differences where distinct regions exhibit unique characteristics. Overlooking spatial heterogeneity can lead to results and conclusions that are not effective or applicable across different areas. Thus, it is essential to address this diversity to ensure that spatial analyses are accurate and relevant (Herwartz, 2007).

Evaluating spatial heterogeneity is critical in ensuring the suitability of statistical models such as GWR. This evaluation helps determine whether the differences between locations justify a region-specific approach. The Breusch-Pagan (BP) test is frequently employed for this assessment. It examines the presence of varying variance across spatial units, suggesting that each location may have distinct structural and interactional features that cannot be captured by a generic model (Koenker, 1981).

The BP test is structured around a hypothesis where the null hypothesis presumes that variance is uniform across the spatial units, while the alternative hypothesis assumes variance heterogeneity. This test is instrumental in detecting spatial heterogeneity, leading researchers to adopt localized models that better fit the specific characteristics of each area. This approach significantly enhances the robustness and applicability of the analysis results. Verifying spatial heterogeneity allows researchers to adequately justify the use of GWR and similar localized models, facilitating more precise and impactful policies or interventions tailored to the geographical nuances of the data (Halunga**, et al.,** 2017).

The Breusch-Pagan test is written as follows to examine the presence of spatial heterogeneity with the hypotheses:

$H_0: \sigma^2_{(u_1,v_1)} = \sigma^2_{(u_2,v_2)} = \cdots = \sigma^2_{(u_n,v_n)} = \sigma^2$ (there is no spatial heterogeneity);

$H_1: at\ least\ one\ i\ where\ \sigma_i^2 \neq \sigma^2; i = 1,2,\dots,n$ (there is spatial heterogeneity).

The formula for the Breusch-Pagan statistical test is as follows (Anselin, 2019):

$$BP = \left(\frac{1}{2}\right) \boldsymbol{h}^T (\boldsymbol{Z^T Z})^{-\mathbf{1}} \boldsymbol{Z^T} \sim \chi^2_{p+1}, \tag{3}$$

where $\boldsymbol{h}$ is the vector element $h_i = \left(\frac{e_i^2}{\sigma^2} - 1\right)$, and $\boldsymbol{Z}$ is the explanatory variables matrix.

H_0 is accepted if $BP \leq \chi^2_{(p+1)}$, where $\chi^2_{(p+1)}$ is the critical value of the chi-square distribution, and p is the number of explanatory variables.

A spatial weighting, a kernel function, is required for GWR modelling. The Gaussian kernel can be utilized with the following equation:

$$w_{ij} = exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h_i}\right)^2\right), \tag{4}$$

where d_{ij} is Euclidean distance between the i^{th} and j^{th} location. h_i is a different *bandwidth* value for the i^{th} location.

Euclid distance can be calculated with Equation 5.

$$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2}, \tag{5}$$

with (u_i, v_i) is the longitude and latitude i^{th} location, and (u_j, v_j) is the longitude and latitude j^{th} location.

The value of each parameter in the GWR model is calculated using points at each site, so the parameter values differ. According to Fotheringham**, et al.** (2002), the Weighted Least Squares (WLS) approach can estimate GWR model parameters. The principle of parameter estimation in WLS is nearly identical to that of OLS (Ordinary Least Squares), which is to minimize the sum of residual squares. WLS parameter estimations are derived by reducing the following Q function: the sum of the residual squares.

$$Q = \sum_{j=1}^{n} \{w_{ij}^{\frac{1}{2}} \varepsilon_i\}^2$$
$$= \sum_{j=1}^{n} (w_{ij}^{\frac{1}{2}}(u_i, v_i)\{y_i - \beta_0(u_i, v_i) - \beta_1(u_i, v_i)x_{1i} - \cdots - \beta_k(u_i, v_i)x_{pi}\})^2. \quad (6)$$

Equation 6 can be written in matrix as follows:

$$Q = \boldsymbol{\varepsilon^T W_i \varepsilon}$$
$$= \{\boldsymbol{Y - X\beta}(u_i, v_i)\}^{\boldsymbol{T}} \boldsymbol{W_i}\{\boldsymbol{Y - X\beta}(u_i, v_i)\}$$
$$= \boldsymbol{Y^T W_i Y - \beta}(u_i, v_i)^{\boldsymbol{T}} \boldsymbol{W_i Y - Y^T W_i X \beta}(u_i, v_i) + \boldsymbol{\beta}(u_i, v_i)^{\boldsymbol{T}} \boldsymbol{X^T W_i X \beta}(u_i, v_i)\boldsymbol{.}$$

Because $\{\boldsymbol{\beta}(u_i, v_i)^{\boldsymbol{T}} \boldsymbol{W_i Y}\}^T = \boldsymbol{Y^T W_i X \beta}(u_i, v_i) = \boldsymbol{\beta}(u_i, v_i)^{\boldsymbol{T}} \boldsymbol{X^T W_i X \beta}(u_i, v_i)$

$$Q = \boldsymbol{Y^T W_i Y - 2\beta}(u_i, v_i)^{\boldsymbol{T}} \boldsymbol{W_i Y} + \boldsymbol{\beta}(u_i, v_i)^{\boldsymbol{T}} \boldsymbol{X^T W_i X \beta}(u_i, v_i). \quad (7)$$

The parameter β is obtained by carrying out a partial derivative of Equation 7 on $\beta(u_i, v_i)$.

So, the estimation of the model parameters can be obtained by Equation 8:

$$\widehat{\boldsymbol{\beta}}(u_i, v_i) = (\boldsymbol{X^T W_i X})^{\boldsymbol{-1}} \boldsymbol{X^T W_i Y}, \qquad i = 1,2, \ldots, n. \quad (8)$$

2.3 *Geographically Weighted Panel Regression (GWR-Panel)*

GWR panels are based on the same concept as cross-sectional GWR. The GWR panel assumes that a time series of measurements at a certain geographic place represents the realization of a spatiotemporal process.

Geographically Weighted Panel Regression (GWR-Panel) is an advanced statistical technique that builds upon cross-sectional Geographically Weighted Regression (GWR) foundations. This method is designed to handle datasets that vary across space and evolve. The GWR-Panel model postulates that a series of measurements at a specific geographical location represents a realization of a spatiotemporal process, suggesting that both spatial and temporal dynamics influence the observed outcomes.

This modelling approach is particularly valuable for exploring the complexities of spatiotemporal data, which is increasingly common in fields like environmental science, urban planning, and public health. GWR-Panel can offer more precise insights into the dynamics that drive changes within a region by accommodating each location's unique structures and interactions over multiple time points. It allows the statistical relationships between variables to vary across space and over time.

Recent studies have utilized the GWR-Panel to address various complex issues. For instance, Wu**, et al.** (2020) have extended traditional GWR by incorporating neural networks to better capture the nuanced interactions of space and time, thus enhancing the model's ability to deal with spatiotemporal non-stationarity. Their research underscores the potential of GWR-Panel to provide a more nuanced understanding of environmental and geographical processes, highlighting its superiority in fitting and predicting spatiotemporal data compared to more traditional models.

Wrenn and Sam (2014) further explored this methodology by applying it to land use changes, demonstrating how temporal heterogeneity can significantly influence the performance of models dealing with nonlinear panel data. Their findings illustrate the robustness of GWR-Panel in capturing the complex, evolving patterns of land development, providing a compelling case for its application in regional planning and policy analysis.

These examples make it clear that Geographically Weighted Panel Regression is a methodological innovation and a practical tool for researchers and policymakers aiming to understand and respond to the intricate dynamics of spatial and temporal variation in their data. This approach represents a significant step forward in the spatial sciences, enabling more targeted, effective interventions based on rich, contextually detailed data analyses.

This process follows a distribution in which adjacent observations (whether geographic or temporal) are more correlated than distant observations. GWR-panel analysis aims to incorporate overall location (cross-sectional) and observations (Yu, 2010).

The general form of the panel regression model with Fixed Effect Model (FEM) is as written in Equation 9.

$$\boldsymbol{Y_{it}} = \boldsymbol{\alpha_i} + \boldsymbol{X_{it}^T \beta} + \boldsymbol{\varepsilon_{it}}, \tag{9}$$

where Y_{it} is the dependent variable, α_i is the constant, X_{it} is the vector of independent variables, β is the vector of coefficients, and ε_{it} is the error term.

Equation 8 can be expressed as a matrix as follows:

$$\begin{bmatrix} \mathrm{y}_1 \\ \vdots \\ \mathrm{y}_\mathrm{N} \end{bmatrix} = \begin{bmatrix} \mathrm{e} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \alpha_1 + \begin{bmatrix} 0 \\ \mathrm{e} \\ \vdots \\ 0 \end{bmatrix} \alpha_2 + \cdots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \mathrm{e} \end{bmatrix} \alpha_N + \begin{bmatrix} \mathrm{X}_1 \\ \mathrm{X}_2 \\ \vdots \\ \mathrm{X}_\mathrm{N} \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix};$$

$$\mathbf{y_i} = \mathbf{e}\alpha_i + \mathbf{X_i}\boldsymbol{\beta} + \boldsymbol{\varepsilon_i}, \tag{10}$$

where

$$\mathbf{y}_\mathrm{i} = \begin{bmatrix} \mathrm{y_{i1}} \\ \mathrm{y_{i2}} \\ \vdots \\ \mathrm{y_{iT}} \end{bmatrix} \quad , \quad \mathbf{X}_\mathrm{i} = \begin{bmatrix} x_{1i1} & x_{2i1} & \cdots & x_{pi1} \\ x_{1i2} & x_{2i2} & \cdots & x_{pi2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1iT} & x_{2iT} & \cdots & x_{piT} \end{bmatrix};$$

$\mathbf{e}^T = (1,1, \dots, 1)$, $\boldsymbol{\varepsilon}_i^T = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT})$;

$E(\boldsymbol{\varepsilon}_i) = 0$, $E(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T) = \sigma_u^2 \boldsymbol{I}_T$, $E(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T) = 0$ if $i \neq j$;

$\boldsymbol{I}_T$ is $T \times T$ identity matrix.

OLS estimator for α_i and β can be calculated by minimized

$$s = \sum_{i=1}^{N} \boldsymbol{\varepsilon}_i^T \boldsymbol{\varepsilon}_i = \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{e}\alpha_i - \mathbf{X}_i\boldsymbol{\beta})^T (\mathbf{y}_i - \mathbf{e}\alpha_i - \mathbf{X}_i\boldsymbol{\beta}). \tag{11}$$

The next step is to calculate the partial derivative of α_i. The result is

$$\hat{\alpha}_i = \bar{y}_i - \boldsymbol{\beta}^T \bar{\mathrm{x}}_i, \ i = 1, 2, \dots, N; \tag{12}$$

with

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \ \bar{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_{it}.$$

The next step is to substitute $\hat{\alpha}_i$ into Equation 11, then determine the partial derivative of β to obtain an estimator using the within group method as follows:

$$\widehat{\boldsymbol{\beta}} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)^T\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(y_{it} - \bar{y}_i)\right]. \tag{13}$$

According to Hsiao (2003), in the calculation procedure to estimate the slope coefficient in the model, it is only necessary to find the average of separate time series observations for each cross-section unit and then transform the research variables by subtracting the corresponding time series average and applying the OLS method to the transformed data.

In Equation 10, if we make an average for each time t=1, ..., T, it becomes a cross-section equation:

$$\bar{y}_{\mathrm{i}} = \bar{\mathbf{x}}_i\boldsymbol{\beta} + \alpha_i + \bar{\boldsymbol{\varepsilon}}_i, \tag{14}$$

where

$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \ \bar{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^{T} \mathrm{x}_{it}, \ \bar{\boldsymbol{\varepsilon}}_i = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\varepsilon}_{it}.$$

Then subtract Equation 9 from Equation 14, so it becomes:

$$(y_{it} - \bar{y}_i) = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\boldsymbol{\beta} + (\boldsymbol{\varepsilon}_{it} - \bar{\boldsymbol{\varepsilon}}_i); \tag{15}$$

or

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \ddot{\boldsymbol{\varepsilon}}_{it}, \qquad t = 1, 2, \dots, T; \tag{16}$$

with

$$\ddot{y}_{it} = (y_{it} - \bar{y}_i), \ddot{\mathbf{x}}_{it} = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i), \text{ and } \ddot{\boldsymbol{\varepsilon}}_{it} = (\boldsymbol{\varepsilon}_{it} - \bar{\boldsymbol{\varepsilon}}_i).$$

In order to estimate the value of β, the Ordinary Least Squares (OLS) method is used by creating a regression equation $\ddot{y}_{it}$ and $\ddot{\mathbf{x}}_{it}$. The fixed effects-within group model is formed by accumulating several observations. Still, for each unit of observation, it is formulated that each variable is a deviation from the average value and then estimated as an OLS regression on the corrected average value (Gujarati & Porter, 2009).

The following equation is a combination of the GWR equation and the FEM panel regression equation within group:

$$\ddot{y}_{it} = \beta_0(u_{it}, v_{it}) + \sum_{k=1}^{p} \beta_k\,(u_{it}, v_{it})\ddot{x}_{itk} + \ddot{\varepsilon}_{it}; \tag{17}$$

$$i = 1, 2, \dots, N\,, t = 1, 2, \dots, T,\ k = 1,\ 2, \dots, K.$$

3. Materials and Methods

3.1 *Data*

The main data used in this study are sugarcane production data. The data is obtained from BPS-Statistics Indonesia and the Directorate General of Plantations. Indonesia is the sixth-most sugar-consuming nation in the world, with 7.8 million metric tons consumed in 2020. In 2021, the amount of sugar used climbed by 200,000 metric tons, totaling 8 million tons. East Java, which comprises 38 districts and towns, is the Indonesian region with the most extensive sugar cane production. A rise in sugar output ought to coincide with increased sugar consumption. According to Indonesia's sugar import statistics for 2022, which climbed by 9.6% over 2021, the country's sugar production is still insufficient for consumption (Kementerian Pertanian, 2023). The considerable volume of sugar imports in Indonesia can be used to illustrate or reflect the country's still-weak local sugar sector. The sugar industry's weakness is caused by the low efficiency of Indonesian sugar facilities, which results in suboptimal sugar production and productivity and higher production costs. The production process of a sugar plant strongly depends on the availability of raw materials for the sugar industry, namely sugar cane. The sugar factory requires sugar cane as raw material for further processing into sugar.

However, the sugar cane plants required for these activities have seen limited production in recent years, resulting in insufficient sugar production to match the production capacity of the sugar mills available in each facility. Sugarcane production is constrained as planting and harvesting grounds for sugarcane commodities are dwindling. As a result, increasing sugarcane production by maximizing available inputs is critical with the shrinking of planting and harvesting areas. This research aimed to determine factors influencing sugarcane output in East Java Province. Six explanatory variables are used to model sugarcane yields. The information used is from three years (2019-2021). Table 1 displays the factors used in this research. Table 1 lists the variables used in this research.

Table 1. Variables of the study

Variables	Information of Variables	Unit
Y	Sugarcane Yields	tons
X_1	Sugarcane plantation area	Ha
X_2	Rainfall	mm
X_3	Number of farmers	Person
X_4	Duration of sunshine	Day
X_5	Temperature	%
X_6	Humidity	^{0}C

Note: Variable Descriptions for Sugarcane Yield Study in East Java.

Table 1, the primary outcome variable, Y (Sugarcane Yields), is quantified in tons and serves as a direct indicator of the productivity levels attained in the sugarcane sector of East Java. This measurement is essential for assessing the effectiveness and overall output of the regional agricultural practices. X_1, representing the Area under sugarcane cultivation, is measured in hectares. This variable is crucial as it reflects the scale of land devoted to sugarcane farming, which directly influences the potential total yield, assuming optimal conditions are maintained for growth. The variable X_2 denotes the Precipitation received, measured in millimeters. Rainfall is critical for sugarcane, which thrives under a sufficient water supply. The amount of rainfall is thus a significant predictor of agricultural yield, highlighting its role in sustainable crop management. X_3 concerns the Number of farmers engaged in sugarcane agriculture, expressed in terms of persons involved. This variable highlights the human resource investment in sugarcane cultivation, linking agricultural labor inputs to productivity outcomes.

Further, X_4, which captures the Sunshine duration during the cultivation period, is recorded in days. Adequate sunlight is vital for the photosynthetic activity essential for sugarcane growth, making this a critical factor in determining the health and yield of the crop. X_5, the Mean temperature during the crop growth period, is measured in degrees Celsius. Temperature regulates several biochemical processes in sugarcane cultivation, impacting growth rates and, ultimately, yield. Lastly, X_6 quantifies the relative humidity during the cultivation phase, which is reported as a percentage. Humidity affects transpiration and moisture availability to sugarcane, thereby influencing yield sustainability and quality.

These variables collectively contribute to a comprehensive model to understand and optimize sugarcane production in East Java. This model offers insights into the biophysical and human factors that drive agricultural output. This integrative approach underscores agricultural production systems' complexity and dependency on environmental conditions and human management practices.

3.2 *Econometric Model*

The procedures taken in this study are as follows:

(1) Exploring the data.
(2) Performing the Breusch- Pagan test to detect spatial heterogeneity with Equation 3.
(3) Using the Z value to discover outliers. A Z-value larger than +3 or less than -3 is considered an anomaly (Tyler, 2008).
(4) Performing the GWR model:
   a. Determine longitude and altitude coordinates of the observation area.
   b. Calculating Euclidean Distance (d_{ij}) between the location to – i and location to – j with Equation 5.
   c. Calculating weighting matrix (w_{ij}) with Equation 4.
   d. Determining bandwidth using CV optimal criteria:

      $CV(h) = \sum_{i=1}^{n}(y_i - \hat{y}_{-i}(h))^2;$
(5) Performing Geographically Weighted Panel Regression Fixed Effect Model with MM Estimation with the following algorithm:
   a. Preprocessing data

      For each cross sectional unit i and time t, transform all variabels into first differences:

      $\Delta y_{it} = y_{it} - y_{i,t-1}, for\ t = 2, \dots T$

$\Delta x_{kit} = x_{kit} - x_{kit-1}, for\ k = 1, \dots K$

b. Estimating $\hat{\beta}^0$ and get $\varepsilon_i^{(0)}$.

c. Calculating $\hat{\sigma}_i$=$\frac{median|\varepsilon_i - median(\varepsilon_i)|}{0,6745}$, where ε_i denotes the residuals obtained from step (b).

d. Calculating $u_i = \frac{\varepsilon_i}{\hat{\sigma}_i}$.

e. Determining the objective function and calculating the weighting value:

$$w_i = \begin{cases} [1 - (\frac{u_i}{4.685})^2]^2 & , |u_i| > 4.685; \\ 0, & , |u_i| > 4.685; \end{cases}$$

f. Calculating $\hat{\beta}_M$ using the Weighted Least Square (WLS) method with weighted w_i:

$$\widehat{\beta_{MM}} = (\boldsymbol{X}^T \boldsymbol{W}^{\boldsymbol{MM}} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W}^{\boldsymbol{MM}} \boldsymbol{y};$$

g. Setting residual in step (e) as residual step (a).

h. Iterating reweighted least squares (IRLS) on new weighting until $\hat{\beta}_M$ convergent.

Incorporating a robust estimation approach within the Geographically Weighted Panel Regression (GWPR) Fixed Effect Model necessitates an exhaustive and meticulous methodological framework that begins with the foundational aspect of data exploration. Consistent with Anselin's (1988) perspective on spatial econometrics, the initial stage comprises rigorous data scrutiny to ensure that spatial dependencies and heterogeneities are well understood before model application.

To address the prevalent issue of spatial heterogeneity, the Breusch-Pagan test, introduced by Breusch and Pagan (1979), is employed to evaluate the variance across the spatial units. This step is critical as it allows for the identification and correction of heteroscedasticity, thus ensuring that the model's assumptions align with the underlying spatial structure of the data.

Outlier detection and management follow, with procedures outlined by Barnett and Lewis (1994) guiding the identification and treatment of outliers. The outlier handling methodology is pivotal in mitigating undue influences that can skew model estimates and compromise inference (Barnett & Lewis, 1994). As elucidated by Fotheringham**, et al.** (2002), the GWR model serves as a precursor to the GWPR model, accommodating spatially localized relationships by integrating geographically sensitive parameters. This localization of parameters is further refined through the Fixed Effect Model, isolating time-invariant unobserved heterogeneity that is spatially variant.

The robust estimator approach within the GWPR framework utilizes the principles of locally weighted regression, as delineated by Cleveland and Devlin (1988), to iteratively refine parameter estimates. This ensures that the localized nature of the relationships is preserved while maintaining resilience against the distorting effects of outliers.

Harris**, et al.** (2011) further reinforce the necessity of considering spatial autocorrelation within the GWPR framework, prompting the use of geographically weighted principal components analysis to distill the essence of spatial relationships.

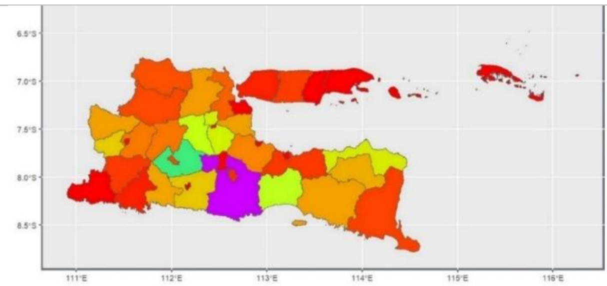
LeSage and Pace (2009) contribute to the robust estimator approach by emphasizing the importance of accounting for spatial econometric relationships. They offer a comprehensive introduction to the spatial econometrics framework that underpins the GWPR model. These references collectively scaffold the methodological rigor of the robust estimation approach within the GWPR Fixed Effect Model, ensuring the study's findings are both robust and reflective of the complex spatial and temporal dynamics at play.

This extended methodological description, complete with references, thoroughly covers the robust estimator approach within the GWPR framework. It provides a solid foundation for rigorous spatial econometric analysis, instilling confidence in the robustness of the study's methodology.

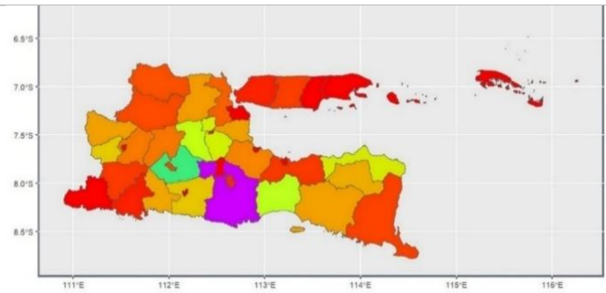
4. Results and Discussion

The heatmap of East Java sugar cane production in 2019-2021 is presented in Figure 1.

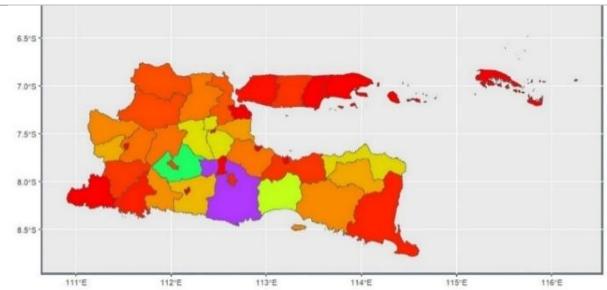
Figure 1. The heatmap of East Java sugar cane production in 2019-2021



(a) Sugarcane Yield Heatmap 2019



(b) sugarcane yield heatmap 2020



(c) Sugarcane Yield Heatmap 2021

Note: The heatmap displays the total sugarcane production (in tons) per district in East Java Province in 2019-2021. Darker colors indicate higher production levels

Figure 1 illustrates that the district/city with the most sugarcane production is highlighted in purple on the map. As a result, in 2019, 2020, and 2021, Malang Regency was the district/city with the most sugar cane output in East Java. However, Pacitan Regency, Pamekasan Regency, Blitar City, and Surabaya City were the only districts/cities that produced no sugarcane between 2019 and 2021. The orange colour map of the district/city demonstrates this.

The results of spatial heterogeneity testing are shown in Table 2.

Table 2. Breusch- Pagan test Result

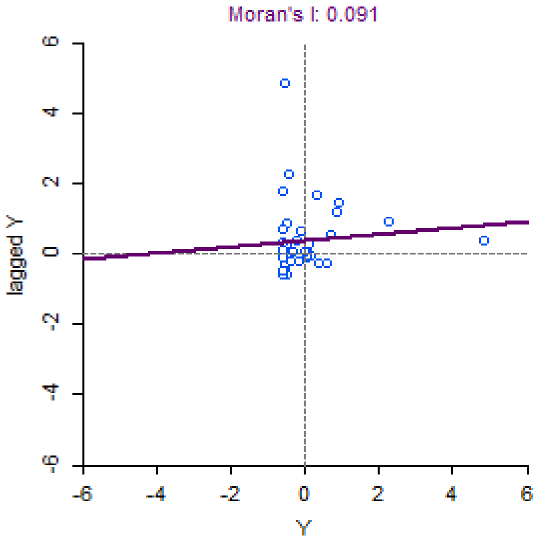
BP	*p-value*	Decision
12.7	0.0001	Reject H_0

Note: The Breusch–Pagan test was conducted to examine the presence of heteroskedasticity in the regression model. The decision is based on α = 5%. If p-value < 0.05, reject H_0.

Table 2 shows the p-value obtained is less than alpha (0,05), implying spatial heterogeneity. As a result, data can be evaluated spatially, such as with Geographically Weighted Panel Regression (GWPR). Subsequently, the Durbin–Watson (DW) test was applied to assess autocorrelation in the residuals. The DW statistic was 1.30, a value notably lower than the benchmark of 2, indicating positive autocorrelation. In other words, residuals in one location tend to be correlated with those in neighboring locations, reflecting a degree of spatial dependency in the data.

Moran's Scatterplot, a four-quadrant graph, analyzes groupings and regions. It divides data into four quadrants to represent potential clustering. A line with average and mean values delimits each quadrant. A region has high features if its values exceed the average. In contrast, locations with below-average qualities have low values. Moran's Scatter plot for sugarcane yield data is presented in Figure 2.

Figure 2. Moran’s Scatter Plot for Sugarcane Yield Data



Note: The Moran scatter plot shows the spatial autocorrelation of sugarcane productivity across districts in East Java in 2019-2021. The slope of the regression line indicates the value of Moran’s I statistic.

Figure 2 demonstrates a negative spatial autocorrelation with a Moran Index (I) of 0.091, falling from 0 to 1. This indicates no significant correlation between cities/districts in East Java regarding the sugarcane yield data at a 5% significance level.

Outlier identification is used to determine whether or not there are any outliers in the data. Outliers in the data that are disregarded cause the analysis results to be biased. The Z value is used for outlier detection. GWPR modelling was also performed in this study, employing panel data that have been changed (demeaned) according to the within-estimator approach. Based on the detection of outliers in GWPR data, it was discovered that East Java has eight districts/cities with outliers. Malang Regency will be in charge from 2019 to 2021, followed by Blitar Regency in 2020, Batu City in 2020 and 2021, Kediri Regency in 2021, and Ngawi Regency in 2021.

GWPR modelling in this research requires estimators that can accommodate the outlier data. Therefore, the MM estimator is used in this research to make the resulting parameter model robust. Modelling sugarcane production data from 2019 to 2021 obtained an optimum bandwidth value of 40400.27 with a fixed Gaussian weighting function. Table 3 displays the results of estimating global GWPR model parameters with the MM-Estimator approach.

Table 3. Global Parameter Estimation (First-Differencing Transformation)

Parameter	Coefficients	*p-value*
Intercept	-1063.3	0.58
X_1	0.985	0.00
X_2	-14.9242	0.51
X_3	0.5090	0.00
X_4	-26.525	0.95
X_5	0.0074	0.00
X_6	81.289	0.62

Note: The table reports the estimated global coefficients from the first-differenced model, where the dependent variable is the annual change in sugarcane productivity (ton/ha). Independent variables are expressed as annual changes in land area (ha), rainfall (mm), number of farmers, sunshine duration, temperature, and humidity. Significant predictors ($p < 0.05$) include land area, number of farmers, and temperature. For example, a one-unit increase in the annual change of land area is associated with a 0.985 change in the annual sugarcane productivity. Insignificant coefficients, such as for rainfall, sunshine, and humidity, indicate weak and statistically insignificant effects.

Table 3 reports the global estimates from the first-differenced model, capturing the average effect of year-to-year changes in explanatory variables on changes in sugarcane productivity. The results show that changes in land area, number of farmers, and temperature have significant positive effects. Specifically, an additional hectare of land compared to the previous year is, on average, associated with a 0.985 ton/ha increase in productivity change, while greater farmer participation and higher temperatures also contribute to productivity growth. By contrast, annual changes in rainfall, sunshine duration, and humidity do not significantly influence productivity dynamics. Although their coefficients suggest a weak negative tendency, the effects are statistically insignificant. Overall, the findings emphasize that productivity growth is primarily driven by the average effects of changes in land expansion, labor input, and temperature, while climatic variations such as rainfall and humidity play a less decisive role. Table 4 summarizes the local parameter estimation of the first-differenced model using the MM-Estimator for districts/cities in East Java.

Table 4. Summary of Local Estimates of the First-Differenced GWPR-MM Model in East Java

Parameter	min	max	Mean	stdev
β_0	-7765.1	12042.84	-520.26	3921.35
β_1	-0,375	2.811	1.428	0.8388
β_2	-32.26	16.256	-3.86	10.945
β_3	-0.62	2.41	1.048	0.645
β_4	-180.562	233.51	-2.005	81.091
β_5	-0.0002	0.0071	0.0026	0.0023
β_6	-93.59	296.764	-43.501	87.638

Note: This table summarizes the local parameter estimates from the first-differenced Geographically Weighted Panel Regression (GWPR) model using the MM-estimator. The summary includes the minimum, maximum, mean, and standard deviation of each local coefficient across districts/cities in East Java. The variation in coefficient values indicates spatial heterogeneity in the effects of explanatory variables on changes in sugarcane productivity.

Local First Differenced GWPR- MM regression models for each area are presented in Table 5.

Table 5. Local Regression Equations of the First-Differenced GWPR-MM Model in East Java

Number	District/City	Regression Model
1	Pacitan District	$\Delta Y = -7765.109 + 2.81X_1 - 22.43X_2 - 0.624X_3 + 233.511X_4 + 0.006X_5 - 93.59X_6$
2	Ponorogo District	$\Delta Y = -3840.54 + 2.11X_1 - 10.76X_2 + 0.67X_3 - 76.12X_4 + 0.002X_5 + 63.98X_6$
3	Trenggalek District	$\Delta Y = 1286.63 + 0.07X_1 - 9.70X_2 + 1.49X_3 - 89.87X_4 + 0.0016X_5 - 63.73X_6$
4	Tulungagung District	$\Delta Y = -2067.35 + 1.67X_1 - 17.96X_2 + 0.68X_3 + 124.42X_4 + 0.002X_5 - 1.148X_6$
5	Blitar District	$\Delta Y = 2256.2 + 1.35X_1 + 2.418X_2 + 1.59X_3 + 24.59X_4 + 0.0003X_5 + 1.395X_6$
6	Kediri District	$\Delta Y = 256.12 + 1.15X_1 + 3.18X_2 + 2.59X_3 + 25.44X_4 + 0.0025X_5 + 1.195X_6$
7	Malang District	$\Delta Y = -463.5 + 1.65X_1 - 3.96X_2 + 0.94X_3 - 13.58.19X_4 + 0.0031X_5 + 51.73X_6$
8	Lumajang District	$\Delta Y = 12042.8 + 0.049X_1 - 32.26X_2 + 2.37X_3 - 67.69X_4 + 0.003X_5 + 285.8X_6$
9	Jember District	$\Delta Y = 8048.3 + 0.028X_1 - 22.46X_2 + 2.41X_3 - 77.2X_4 + 0.001X_5 + 264.32X_6$
10	Banyuwangi District	$\Delta Y = 6316.76 + 0.37X_1 + 15.504X_2 + 1.29X_3 + 160.31X_4 + 0.007X_5 + 296.7X_6$
11	Bondowoso District	$\Delta Y = -3823.6 + 2.008X_1 - 8.801X_2 + 0.86X_3 + 1.95X_4 + 0.0013X_5 + 16.035X_6$
12	Situbondo District	$\Delta Y = -292.68 + 0.76X_1 - 0.18X_2 + 1.71X_3 + 215.05X_4 - 0.0001X_5 - 10.17X_6$
13	Probolinggo District	$\Delta Y = -1840.7 + 1.81X_1 + 2.88X_2 + 1.28X_3 + 6.10X_4 - 0.00005X_5 + 2.19X_6$
14	Pasuruan District	$\Delta Y = -191.7 + 2.078X_1 - 3.198X_2 + 1.24X_3 + 8.82X_4 + 0.0001X_5 + 18.72X_6$
15	Sidoarjo District	$\Delta Y = 2565.77 + 2.13X_1 - 16.32X_2 + 0.71X_3 + 9.002X_4 + 0.0049X_5 + 56.96X_6$
16	Mojokerto District	$\Delta Y = 590.33 + 1.87X_1 - 7.48X_2 + 0.89X_3 - 19.92X_4 + 0.0033X_5 + 37.91X_6$
17	Jombang District	$\Delta Y = 901.5 + 1.669X_1 - 18.49X_2 + 0.64X_3 - 58.05X_4 + 0.0058X_5 + 63.58X_6$
18	Nganjuk District	$\Delta Y = -1195.36 + 1.97X_1 + 3.19X_2 + 0.61X_3 - 180.5X_4 + 0.0045X_5 + 34.62X_6$
19	Madiun District	$\Delta Y = 435.47 + 1.85X_1 - 8.54X_2 + 0.90X_3 - 18.01X_4 + 0.0033X_5 + 44.11X_6$
20	Magetan District	$\Delta Y = -4844.62 + 1.77X_1 + 4.56X_2 + 0.11X_3 - 120.09X_4 + 0.0056 - 65.8X_6$
21	Ngawi District	$\Delta Y = -4005.63 + 1.66X_1 + 10.76 + 0.36X_3 - 14.48X_4 + 0.0049X_5 + 51.16X_6$
22	Bojonegoro District	$\Delta Y = -6421.36 + 1.399X_1 + 2.85X_2 + 0.88X_3 - 9.777X_4 + 0.0014X_5 - 83.17X_6$
23	Tuban District	$\Delta Y = -6554.08 + 2.075X_1 + 2.75X_2 + 0.80X_3 + 22.66X_4 + 0.002X_5 + 10.34X_6$
24	Lamongan District	$\Delta Y = -1164.75 + 2.33X_1 + 2.75X_2 + 0.80X_3 + 22.66X_4 + 0.002X_5 + 10.34X_6$
25	Gresik District	$\Delta Y = -2325.12 + 1.82X_1 + 5.70X_2 + 0.58X_3 + 11.92X_4 + 0.0042X_5 + 67.12X_6$
26	Bangkalan District	$\Delta Y = -2705.77 + 1.35X_1 + 0.92X_2 + 0.25X_3 + 44.2X_4 + 0.007X_5 + 142.13X_6$
27	Sampang District	$\Delta Y = 836.15 - 0.083X_1 + 0.39X_2 + 1.96X_3 - 11.9X_4 - 0.00023X_5 + 15.75X_6$
28	Pamekasan District	$\Delta Y = 750.00 - 0.087X_1 - 1.67X_2 + 1.93X_3 + 34.5X_4 - 0.00005X_5 + 0.23X_6$
29	Sumenep District	$\Delta Y = -504.3 - 0.11X_1 - 24.8X_2 + 1.85X_3 + 38.38X_4 + 0.0003X_5 - 88.7X_6$
30	Kediri City	$\Delta Y = -616.28 + 1.65X_1 - 3.10X_2 + 0.94X_3 - 26.87X_4 + 0.0031X_5 + 33.59X_6$
31	Blitar City	$\Delta Y = 2028.89 + 2.43X_1 + 8.33X_2 + 1.25X_3 - 91.78X_4 + 0.0009X_5 + 17.00X_6$
32	Malang City	$\Delta Y = 1792.93 + 2.067X_1 - 4.25X_2 + 1.48X_3 - 11.6X_4 + 0.0006X_5 + 3.18X_6$
33	Probolinggo City	$\Delta Y = 3885.65 + 0.39X_1 + 10.36X_2 + 2.08X_3 + 52.18X_4 - 0.00008X_5 + 137.15X_6$
34	Pasuruan City	$\Delta Y = -6340.96 + 1.28X_1 + 16.25X_2 + 0.97X_3 - 36.78X_4 + 0.00076X_5 + 63.90X_6$
35	Mojokerto City	$\Delta Y = 547.95 + 1.86X_1 - 8.67X_2 + 0.899X_3 - 22.6X_4 + 0.0033X_5 + 37.5X_6$
36	Madiun City	$\Delta Y = -3447.56 + 1.95X_1 + 4.15X_2 + 0.41X_3 - 72.5X_4 + 0.004X_5 + 70.21X_6$
37	Surabaya City	$\Delta Y = -2064.114 + 1.58X_1 - 5.51X_2 + 0.28X_3 - 9.47X_4 + 0.0069X_5 + 100.3X_6$

38	Batu City	$\Delta Y = -1852.50 + 2.15X_1 - 0.68X_2 + 0.61X_3 + 6.21X_4 + 0.0034X_5 + 28.84X_6$

Note: This table presents the significant variables identified from the First-Differenced Geographically Weighted Panel Regression (GWPR) model with MM-estimation for each district/city in East Java. The coefficients are based on first-differenced data to ensure stationarity. ΔY denotes the first difference of sugarcane yields (tons); X_1, first difference of plantation area (Ha); X_2, first difference of rainfall (mm); X_3, first difference of number of farmers (persons); X_4, first difference of sunshine duration (hours); X_5, first difference of temperature (°C); and X_6, first difference of humidity (%). Coefficients are estimated from the first-differenced Geographically Weighted Panel Regression (GWRP) model with MM-estimation to ensure stationarity and capture spatial heterogeneity in the relationship between these predictors and sugarcane productivity. Differences in parameter estimates across regions highlight spatial heterogeneity in the relationship between these predictors and sugarcane productivity.

Based on Table 5, the regression model of Lumajang district is

$$\Delta Y = 12042.8 + 0.049X_1 - 32.26X_2 + 2.37X_3 - 67.69X_4 + 0.003X_5 + 285.8X_6.$$

The coefficient of plantation area (X_1) is positive, indicating that an increase in the annual change of plantation area is associated with an increase in the annual change of sugarcane yield. Similarly, the number of farmers (X_3) and humidity (X_6) also show positive effects, meaning that districts experiencing growth in these variables tend to have higher growth in sugarcane productivity. In contrast, annual changes in rainfall (X_2) and sunshine duration (X_4) are negatively associated with sugarcane yield changes, suggesting that higher rainfall or temperature fluctuations contribute to yield reduction. The variable temperature (X_5) also shows a positive but relatively small effect.

The local regression equations in Table 5 indicate that the relationships between the annual changes in explanatory variables and sugarcane productivity differ across districts in East Java. In several districts, the annual change in plantation area shows a positive and significant effect, implying that expansion in cultivated land is generally followed by growth in sugarcane yield. Similarly, changes in the number of farmers, sunshine duration, and humidity tend to contribute positively to productivity dynamics, although the magnitude of the effect varies by location. Table 6 summarizes the variables that are significant according to the Wald test.

Table 6. Wald Test Results of Significant Variables in the First-Differenced GWPR-MM Model

No	Significant Variable	District/city	Number of Significant District/City
1	X_1	Pacitan District, Ponorogo District, Tulungagung district, Blitar District, Kediri District, Malang District, Situbondo District, Probolinggo District, Pasuruan District, Sidoarjo District, Mojokerto District, Jombang District, Nganjuk District, Madiun District, Magetan District, Ngawi District, Bojonegoro District, Tuban District, Lamongan District, Gresik District, Bangkalan District, Sumenep District, Kediri City, Malang City, Madiun City, Surabaya City, Batu City	27
2	X_2	Sumenep District	1
3	X_3	Blitar District, Kediri District, Malang District, Situbondo District, Probolinggo District, Pasuruan District, Bojonegoro District, Tuban District, Bangkalan District, Sumenep District, Pamekasan Districy, Kediri City,	12
4	X_4	Magetan District, Ngawi District,	2
5	X_3, X_5	Pacitan District, Ponorogo District, Trenggalek District Tulungagung district, Malang District, Lumajang District,	21

	Banyuwangi District, Bondowoso District, Sidoarjo District, Mojokerto District, Jombang District, Nganjuk District, Madiun District, Magetan District, Ngawi District, Lamongan District, Gresik District, Malang City, Madiun City, Surabaya City, Batu City

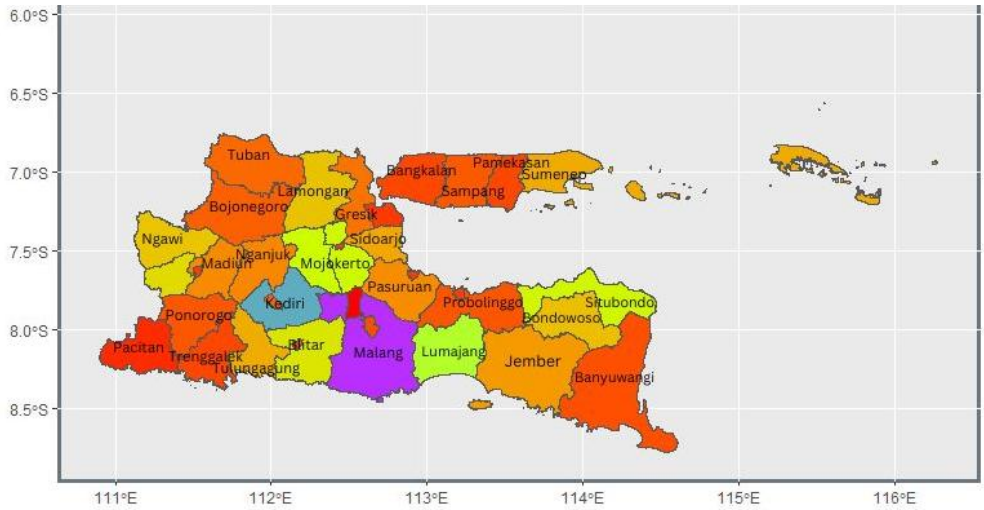
Note: This table presents the explanatory variables that are statistically significant ($p < 0.05$) in the local *Geographically Weighted Panel Regression* (GWPR) models estimated using the robust MM estimator, based on the Wald test. The dependent variable is sugarcane yield (tons/ha), while the explanatory variables (X_1–X_6) include sugarcane plantation area (hectares), rainfall (mm), number of farmers (persons), duration of sunshine (days), temperature (%), and humidity (°C). Entries such as "X_3, X_5" indicate that the corresponding variables are jointly significant in the respective district or city.

Table 6 presents the results of the Wald test to identify which explanatory variables are statistically significant in the first-differenced GWPR-MM model across districts and cities in East Java. The results show that the annual change in plantation area (X_1) emerges as the most dominant factor, being significant in 27 districts/cities. This finding reinforces the central role of land expansion or contraction in shaping the annual dynamics of sugarcane productivity. The annual change in the number of farmers (X_3) is also significant in 12 districts/cities, suggesting that variations in labor availability contribute meaningfully to changes in production. Interestingly, in 21 districts, both X_3 (farmers) and X_5 (sunshine duration) are jointly significant, indicating that the interplay between human resources and climatic conditions is crucial in explaining yield growth. Meanwhile, rainfall (X_2) and temperature (X_4) are significant only in a few districts (Sumenep, Magetan, and Ngawi), reflecting that their influence is more localized and context-dependent. The relatively limited significance of these variables highlights that climatic changes do not affect all regions uniformly but may exert strong impacts in specific ecological settings. Overall, these results confirm the presence of spatial heterogeneity: while land area remains the most consistent driver of productivity changes, other variables—especially labor and sunshine—show varying importance across districts. This implies that strategies to increase sugarcane productivity should be tailored to local conditions rather than relying solely on a uniform provincial approach.

Furthermore, the impact of climate change on sugarcane yield has been a subject of investigation (Jones**, et al.,** 2015), which simulated the impacts of climate change on water use and yield of irrigated sugarcane in South Africa, emphasizing the need to explore the implications of climate data simplifications in future work. Similarly, Black**, et al.** (2012) assessed the impacts of climate change on sugarcane production in Ghana, underscoring the importance of evaluating climate change effects on crop yields. Moreover, using satellite-based systems and remote sensing technologies has been instrumental in estimating and forecasting sugarcane yield (Morel**, et al.,** 2014). The potential of the open-access Sentinel-2 Earth observation system in overcoming limitations in sugarcane yield estimation was discussed. Additionally, Rahman and Robson (2016) compared different methods to estimate sugarcane yield using Landsat time series imagery, highlighting the accuracy of remote sensing-based integrated NDVI methods. In conclusion, the research on sugarcane yield encompasses a wide array of factors, including climate variables, advanced technologies such as remote sensing and GIS, and the impact of climate change. These studies collectively contribute to a comprehensive understanding of the variables influencing sugarcane yield and the methodologies employed for its prediction and estimation.

The sugarcane yield prediction map is presented in Figure 3. The prediction map shows the same pattern as the original data. Figure 3 illustrates that the district/city with the most sugar cane production is highlighted in purple on the map. As in the original data, Malang regency was the district/city with the most sugar cane output in East Java.

Figure 3. Sugarcane Yield Prediction Heatmap



Note: This heatmap illustrates the predicted sugarcane yield (in tons per hectare) across districts in East Java for the year 2021. The predictions are generated using the Geographically Weighted Panel Regression (GWPR) model with a robust MM estimator. Darker areas indicate higher predicted yields, while lighter areas represent lower yields.

Figure 3, based on the prediction results produced from the GWPR model with a robust estimator, the R-squared value was 92.4%, indicating that the regression model can explain 92.4% of the variability in the target variable. The Mean Absolute Percentage (MAPE) of 16.8% obtained from parameter estimation using the MM estimator indicates good model performance, falling within the range of 10% to 20% as suggested by Chang**, et al.** (2007) and Hodge and Austin (2004).

The GWPR model with MM estimator, a robust and reliable tool, outperforms the GWPR model with WLS. This is evident from the R-squared value of GWPR MM estimation, 92.4%, compared to 90.4% in GWPR WLS. The MAPE value for GWPR MM estimation, 16.8%, is also significantly better than the 48.74% for GWPR WLS. By prioritizing models with the largest R2 and smallest MAPE values, one aims to select a model that captures the data's variability well and provides accurate predictions. This approach ensures that the chosen model balances goodness of fit and predictive accuracy, both essential aspects of model performance evaluation in various fields such as forecasting, regression analysis, and machine learning. R-squared, or the coefficient of determination, indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. A larger R-squared value signifies a better fit of the model to the data, indicating how well the model explains the variability of the data points around the mean (Sumari**, et al.,** 2022). On the other hand, MAPE, or mean absolute percentage error, measures the accuracy of predictions by calculating the average percentage difference between predicted and actual values. A smaller MAPE value indicates higher accuracy in the model's predictions (Xu**, et al.,** 2019). This criterion suggests a balance between explaining the variance in the data and making accurate predictions. When selecting a model, it is crucial to consider both metrics to ensure a comprehensive assessment of model performance (Sumari**, et al.,** 2022).

This finding aligns with Erda and Djuraidah's (2019) research, which utilized the M estimator in Geographically Temporally Weighted Regression (GTWR) modelling on data containing outliers, concluding that robust estimators are suitable for modelling such data. This alignment with established research provides a strong foundation for the validity of our study's findings.

GWPR is a powerful modelling technique that allows for exploring spatial and temporal non-stationarity in data, particularly when dealing with outliers. One of the critical advantages of GWPR is its ability to account for spatial and temporal variations in the relationships between variables, making it particularly useful in scenarios where traditional regression models may not suffice (Brunsdon**, et al.,** 1996). GWPR has been widely applied in various fields such as disease mapping, crime analysis, and environmental health, where the spatial and temporal variations in relationships between variables are of particular interest (Nakaya**, et al.,** 2005; Poliart**, et al.,** 2021; Zeleke**, et al.,** 2022).

In modelling data containing outliers, GWPR offers a robust approach to account for spatial and temporal variations. The use of geographically weighted Poisson regression (GWPR) has been particularly effective in modelling count data with a Poisson distribution, such as the number of disease cases, crime incidents, or other count-based phenomena (Chen**, et al.,** 2020; Tavares & Costa, 2021; Zhou**, et al.,** 2019). This is especially relevant when dealing with outliers, as the spatially varying relationships captured by GWPR can provide insights into how the effects of independent variables vary depending on the location and time, thus offering a more nuanced understanding of the data (Lee**, et al.,** 2023; Liu**, et al.,** 2018; Manyangadze**, et al.,** 2021). The effectiveness of GWPR in handling outliers provides confidence in its applicability to various research scenarios.

Furthermore, the incorporation of geographically weighted Poisson regression (GWPR) in modelling count data with outliers is effective in various studies, including examining the spatial heterogeneity of disease incidence, crime rates, and environmental factors (Fundisi**, et al.,** 2023; Manyangadze**, et al.,** 2017; Zhang**, et al.,** 2021). The localized nature of GWPR allows for the identification of spatial clusters, hotspots, and areas of high or low incidence, which can be particularly valuable in identifying outliers and understanding their impact on the overall model (Atumo**, et al.,** 2022; Ehlkes**, et al.,** 2014).

Geographically Weighted Poisson Regression (GWPR) provides a robust and flexible approach to modelling data containing outliers. By accounting for spatial and temporal variations in the relationships between variables, GWPR offers a nuanced understanding of the data and can effectively capture the impact of outliers in count-based phenomena. Similarly, Isnaini**, et al.** (2019) also support using robust estimators for data containing outliers (Prasetya, 2023).

The study by Knight and Wang (2009) supports the effectiveness of MM-estimators and the L1-norm in excluding outliers, with rates up to ten percent higher than the outlier test as the number of outliers increased (Chen & Liu, 1993). Furthermore, the work of Kannan and Manoj (2015) emphasizes the importance of identifying outliers in regression analysis, particularly in multivariate data, and the relevance of robust estimators in dealing with outliers (Müller**, et al.,** 2011).

The references provide comprehensive support for using robust estimators, such as the MM and M estimators, in modelling data containing outliers. They also highlight the significance of outlier detection and the effectiveness of robust estimators in handling outliers in various types of data.

5. Conclusion

This study underscores the critical role of spatial analysis in agricultural economics research, particularly in the context of sugarcane production in East Java from 2019 to 2021. This study successfully addressed the dataset's challenges of spatial heterogeneity and outliers by utilizing the Geographically Weighted Panel Regression (GWPR) with a fixed effect model and robust MM estimator. The findings indicate that the plantation area is a significant variable affecting sugarcane yield in almost all the locations studied, including in districts with substantial sugarcane production, such as Malang Regency, Blitar City, and Ngawi Regency, where data showed the presence of outliers.

Further analysis confirmed the effectiveness of the Geographically Weighted Panel Regression with a robust estimator in modelling data containing outliers, as evidenced by a high R-squared value of 92.4% and a Mean Absolute Percentage Error (MAPE) of 16.8%. GWPR with a robust estimator is better than GWPR with WLS. It can be seen from the R-squared of GWPR with WLS that it is smaller than that of GWPR with the MM estimator, and the MAPE is smaller than that of GWPR with a robust estimator. This affirms that GWPR with a robust estimator is a suitable approach for data modelling in agricultural economics research, especially when the data contains outliers.

Additionally, this study expands the understanding of sugarcane production dynamics by highlighting geographical and climatic factors as key determinants of yield. It offers a significant contribution to the literature by identifying and validating the use of robust estimators in the analysis of complex and heterogeneous data.

Thus, this study not only provides new insights into the factors affecting sugarcane yield in East Java but also validates the effectiveness of the GWPR approach with a robust estimator in overcoming analytical challenges faced in agricultural economics research. Its implications range from developing evidence-based agricultural practices to more targeted agricultural policy-making, emphasizing the importance of robust approaches in analysing data containing outliers.

The findings of this study provide significant insights relevant to policymakers and practitioners in the agricultural sector, particularly concerning sugarcane cultivation in East Java. The government and related entities urgently need to enhance environmental management and land use policies to ensure that sugarcane farming activities are productive and environmentally sustainable. Additionally, providing financial and technological support to operators and farmers involved in sugarcane production is crucial, especially in adopting low-impact and efficient farming practices.

This study emphasizes the importance of considering geographical factors and climate conditions in decision-making related to sugarcane agriculture. Therefore, it is recommended for local governments and stakeholders in the agricultural sector to incorporate the use of Geographic Information Systems (GIS) and other spatial analysis methods in planning and managing sugarcane cultivation. This will assist in identifying the most suitable areas for sugarcane production and managing risks associated with climate change and other environmental factors.

This research contributes significantly to the literature on agricultural economics by highlighting the importance of spatial analysis in understanding the dynamics of agricultural production. By addressing the identified limitations, future research can provide broader and more detailed insights into improving efficiency and sustainability in Indonesia's agricultural production, particularly sugarcane.

This effect may significantly affect the development of more targeted and effective agricultural policies. One implication is that policymakers and farmers must carefully consider spatial differences and environmental factors in their decision-making processes. This has two important implications: first, it underscores the necessity for agriculture to adapt to geographical and climatic conditions, and second, it highlights the potential of spatial analysis to inform more sustainable and productive agricultural practices.

However, this study is limited by using data from 2019 to 2021 in East Java and applying the Geographically Weighted Panel Regression Model with a robust MM estimator. Future research has the opportunity to expand the geographical coverage and period of the study to enhance the applicability of the findings. Subsequent research could also explore other factors influencing sugarcane production, including government policies, market dynamics, and agricultural technological innovations.

References

Anselin, L. (1988). Spatial Econometrics: Methods and Models. Kluwer Academic Publishers.

Anselin, L. (2019). The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In *Spatial analytical perspectives on GIS* (pp. 111–126). Routledge

Atumo, E., Li, H., & Jiang, X. (2022). Segment-Level Spatial Heterogeneity Of Arterial Crash Frequency Using Locally Weighted Generalized Linear Models. Transportation Research Record *Journal of the Transportation Research Board*, 2677(3), 1637-1653. https://doi.org/10.1177/03611981221126510

Barnett, V., & Lewis, T. (1994). Outliers in Statistical Data. John Wiley & Sons.

Bian, G., McAleer, M., & Wong, W. K. (2013). Robust estimation and forecasting of the capital asset pricing model. *Annals of Financial Economics*, *8*(02), 1350007.

Black, E., Vidale, P., Verhoef, A., Cuadra, S., Osborne, T., & Hoof, C. (2012). Cultivating C4 Crops In A Changing Climate: Sugarcane In Ghana. *Environmental Research Letters*, 7(4), 044027. https://doi.org/10.1088/1748-9326/7/4/044027

Box, G. E. (1953). Non-Normality And Tests On Variances. *Biometrika*, *40*(3/4), 318-335.

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test For Heteroscedasticity And Random Coefficient Variation. *Econometrica*, 47(5), 1287-1294.

Brunsdon, C., Fotheringham, A., & Charlton, M. (1996). Geographically Weighted Regression: A Method For Exploring Spatial Nonstationarity. *Geographical Analysis*, 28(4), 281-298. https://doi.org/10.1111/j.1538-4632.1996.tb00936.x

Chang, P. C., Wang, Y. W., & Liu, C. H. (2007). The Development Of A Weighted Evolving Fuzzy Neural Network For PCB Sales Forecasting. *Expert Systems with Applications*, *32*(1), 86-96.

Chen, C. (2002). Paper 265-27 Robust Regression And Outlier Detection With The ROBUSTREG Procedure. In *Proceedings of the Proceedings of the Twenty-Seventh Annual SAS Users Group International Conference*.

Chen, C., & Liu, L. (1993). Forecasting Time Series With Outliers. *Journal of Forecasting*, 12(1), 13-35. https://doi.org/10.1002/for.3980120103

Chen, J., Liu, L., Liu, H., Long, D., Xu, C., & Zhou, H. (2020). The Spatial Heterogeneity Of Factors Of Drug Dealing: A Case Study From Zg, China. *International Journal of Geo-Information*, 9(4), 205. https://doi.org/10.3390/ijgi9040205

Cleveland, W. S., & Devlin, S. J. (1988). Locally Weighted Regression: An Approach To Regression Analysis By Local Fitting. *Journal of the American Statistical Association*, 83(403), 596-610.

Combettes, P. L., & Muller, C. L. (2020). Perspective Maximum Likelihood-Type Estimation Via Proximal Decomposition. *Electronic Journal of Statistics*, 14(1): 207-238. https://doi.org/10.1214/19-EJS1662

Ehlkes, L., Krefis, A. C., Kreuels, B., Krumkamp, R., Adjei, O., Ayim-Akonor, M., ... & May, J. (2014). Geographically Weighted Regression Of Land Cover Determinants Of Plasmodium Falciparum Transmission In The Ashanti Region Of Ghana. *International journal of health geographics*, *13*, 1-11. https://doi.org/10.1186/1476-072x-13-35

Erda, G., & Djuraidah, A. (2019). Outlier Handling Of Robust Geographically And Temporally Weighted Regression. In *Journal of Physics: Conference Series* (Vol. 1175, No. 1, p. 012041). IOP Publishing.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. E. (2002). *Geographically Weighted Regression : the Analysis of Spatially Varying Relationship*. John Wiley & Sons. England

Fundisi, E., Dlamini, S., Mokhele, T., Weir-Smith, G., & Motolwana, E. (2023). Exploring Determinants Of Hiv/Aids Self-Testing Uptake In South Africa Using Generalised Linear Poisson And Geographically Weighted Poisson Regression. *Healthcare*, 11(6), 881. https://doi.org/10.3390/healthcare11060881

Gujarati, D. N., & Porter, D. (2009). *Basic Econometrics*. Mc Graw Hill Inc. New York.

Guo, Y., Wong, W. K., Su, N., Ghardallou, W., Gavilán, J. C. O., Uyen, P. T. M., & Cong, P. T. (2023). Resource Curse Hypothesis And Economic Growth: A Global Analysis Using Bootstrapped Panel Quantile Regression Analysis. *Resources Policy*, *85*, 103790.

Halunga, A. G., Orme, C. D., & Yamagata, T. (2017). A Heteroskedasticity Robust Breusch–Pagan Test For Contemporaneous Correlation In Dynamic Panel Data Models. *Journal of econometrics*, *198*(2), 209-230.

Harris, P., Brunsdon, C., & Charlton, M. (2011). Geographically Weighted Principal Components Analysis. *International Journal of Geographical Information Science*, *25*(10), 1717-1736.

Harris, P., Brunsdon, C., Charlton, M., Juggins, S., & Clarke, A. (2014). Multivariate Spatial Outlier Detection Using Robust Geographically Weighted Methods. *Mathematical Geosciences*. 46(1): 1–31.

Herwartz, H. (2007). Testing For Random Effects In Panel Models With Spatially Correlated Disturbances. *Statistica Neerlandica*, *61*(4), 466-487.

Hodge, V., & Austin, J. (2004). A survey of outlier detection methodologies. Artificial Intelligence Review, 22(2), 85-126. https://doi.org/10.1007/s10462-004-4304-y

Hoechle, D. (2007). Robust standard errors for panel regressions with cross-sectional dependence. *The stata journal*, *7*(3), 281-312.

Hsiao, C. (2003). Analysis of Panel Data. Cambridge University Press. New York *International Journal of Finance & Banking Studies (2147-4486)*, *9*(1), 58-67.

Huber, P. J. (1973). Robust Regression: Asymptotics, Conjectures and Monte Carlo. *Annals of Statistics*, 1, 799-821. http://dx.doi.org/10.1214/aos/1176342503

Isnaini, B., Dyah, S. U., & Nur Aidi, M. (2019). Estimating the Parameters of a Robust Geographically Weighted Regression Model in Gross Regional Domestics Product in East Java. *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, *48*(3), 150–160.

Jones, M., Singels, A., & Ruane, A. (2015). Simulated Impacts Of Climate Change On Water Use And Yield Of Irrigated Sugarcane In South Africa. *Agricultural Systems*, 139, 260-270. https://doi.org/10.1016/j.agsy.2015.07.007

Kannan, K., & Manoj, K. (2015). Outlier Detection In Multivariate Data. Applied Mathematical Sciences, 9, 2317-2324. https://doi.org/10.12988/ams.2015.53213

Kementerian Pertanian. (2023). *Analisis Kinerja Perdagangan Gula*. Pusat Data dan Sistem Informasi Pertanian Sekretariat Jenderal, Kementerian Pertanian.

Knight, N., & Wang, J. (2009). A Comparison Of Outlier Detection Procedures And Robust Estimation Methods In Gps Positioning. *Journal of Navigation*, 62(4), 699-709. https://doi.org/10.1017/s0373463309990142

Koenker, R. (1981). A note on studentizing a test for heteroscedasticity. *Journal of econometrics*, *17*(1), 107-112.

Kudraszow, N. L., & Maronna, R. A. (2011). Estimates of MM type for the multivariate linear model. *Journal of Multivariate Analysis*, *102*(9), 1280-1292.

Lee, M., Lin, V., Mei, Z., Mei, J., Chan, E., Shipp, D., Chen, J. M., & Le, T. (2023). Examining The Spatial Varying Effects Of Sociodemographic Factors On Adult Cochlear Implantation Using Geographically Weighted Poisson Regression. *Otology & Neurotology*, 44(5), e287-e294. https://doi.org/10.1097/mao.0000000000003861

LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.

LeSage, J. P. (2004). A Family of Geographically Weighted Regression Models. *Advances in Spatial Econometrics* : 241-264

Liu, J., Wang, Y., Fu, C., Guo, J., & Yu, Q. (2016). A Robust Regression Based On Weighted LSSVM And Penalized Trimmed Squares. *Chaos, Solitons & Fractals*, *89*, 328-334.

Liu, Y., Ji, Y., Shi, Z., & Gao, L. (2018). The Influence Of The Built Environment On School Children's Metro Ridership: An Exploration Using Geographically Weighted Poisson Regression Models. Sustainability, 10(12), 4684. https://doi.org/10.3390/su10124684

Ma, Z., Xue, Y., & Hu, G. (2021). Geographically Weighted Regression Analysis For Spatial Economics Data: A Bayesian Recourse. *International Regional Science Review*, *44*(5), 582-604.

Manyangadze, T., Chimbari, M., Macherera, M., & Mukaratirwa, S. (2017). Micro-Spatial Distribution Of Malaria Cases And Control Strategies At Ward Level In Gwanda District, Matabeleland South, Zimbabwe. *Malaria Journal*, 16(1). https://doi.org/10.1186/s12936-017-2116-1

Manyangadze, T., Chimbari, M., & Mavhura, E. (2021). Spatial Heterogeneity Association Of Hiv Incidence With Socio-Economic Factors In Zimbabwe. *Journal of Geographical Research*, 4(3), 51-60. https://doi.org/10.30564/jgr.v4i3.3466

Morel, J., Todoroff, P., Bégué, A., Bury, A., Martiné, J., & Petit, M. (2014). Toward A Satellite-Based System Of Sugarcane Yield Estimation And Forecasting In Smallholder Farming Conditions: A Case Study On Reunion Island. *Remote Sensing*, 6(7), 6620-6635. https://doi.org/10.3390/rs6076620

Müller, E., Schiffer, M., & Seidl, T. (2011). Statistical Selection Of Relevant Subspace Projections For Outlier Ranking. *2011 IEEE 27th International Conference on Data Engineering*, Hannover, Germany, 2011, pp. 434-445, https://doi.org/10.1109/ICDE.2011.5767916.

Nakaya, T., Fotheringham, A., Brunsdon, C., & Charlton, M. (2005). Geographically Weighted Poisson Regression For Disease Association Mapping. *Statistics in Medicine*, 24(17), 2695-2717. https://doi.org/10.1002/sim.2129

Naudé, W. A. (2004). The Effects Of Policy, Institutions And Geography On Economic Growth In Africa: An Econometric Study Based On Cross-Section And Panel Data. *Journal of International Development*, *16*(6), 821-849.

Nilsson, P. (2014). Natural amenities in urban space–A Geographically Weighted Regression Approach. *Landscape and Urban Planning*, *121*, 45-54.

Ningrum, A. S., Rusgiyono, A., & Prahutama, A. (2020, April). Village Classification Index Prediction Using Geographically Weighted Panel Regression. In *Journal of Physics: Conference Series* (Vol. 1524, No. 1, p. 012040). IOP Publishing.

Özdemir, Ş., & Arslan, O. (2021). Empirical likelihood-MM (EL-MM) Estimation For The Parameters Of A Linear Regression Model. *Statistics*, *55*(1), 45-67.

Poliart, A., Kirakoya-Samadoulougou, F., Ouedraogo, M., Collart, P., Dubourg, D., & Samadoulougou, S. (2021). Using Geographically Weighted Poisson Regression To Examine The Association Between Socioeconomic Factors And Hysterectomy Incidence In Wallonia, Belgium. *BMC Women S Health*, 21(1). https://doi.org/10.1186/s12905-021-01514-y

Prasetya, R. (2023). Unpacking Outlier With Weight Least Square (Implemented On Pepper Plantations Data). *Parameter Journal of Statistics*, 2(3), 24-31. https://doi.org/10.22487/27765660.2022.v2.i3.16138

Rahman, M., & Robson, A. (2016). A Novel Approach For Sugarcane Yield Prediction Using Landsat Time Series Imagery: A Case Study On Bundaberg Region. *Advances in Remote Sensing*, 05(02), 93-102. https://doi.org/10.4236/ars.2016.52008

Rasekhi, S., Anousheh, S., Ranjbar, H., & Moghimi, M. (2013). Regional Spillover Research And Development Investment: A Geographically Weighted Regression Approach. *African Journal of Business Management*, *7*(33), 3212-3219.

Rousseeuw, P., & Yohai, V. (1984, January). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series Analysis: Proceedings of a Workshop Organized by the Sonderforschungsbereich 123 "Stochastische Mathematische Modelle", Heidelberg 1983* (pp. 256-272). New York, NY: Springer US.

Sugasawa, S., & Murakami, D. (2022). Adaptively Robust Geographically Weighted Regression. *Spatial Statistics*, *48*, 100623.

Sumari, A. D. W., Charlinawati, D. S., & Ariyanto, Y. (2022). A Simple Approach Using Statistical-Based Machine Learning To Predict The Weapon System Operational Readiness. Proceedings of the International Conference on Data Science and Official Statistics, 2021(1), 343-351.

Susanti, Y., Pratiwi, H., Sulistijowati, S., & Liana, T. (2014). M estimation, S estimation, and MM estimation in robust regression. *International Journal of Pure and Applied Mathematics*, *91*(3), 349-360. https://doi.org/10.12732/ijpam.v91i3.7.

Tavares, J., & Costa, A. (2021). Spatial Modelling And Analysis Of The Determinants Of Property Crime In Portugal. Isprs *International Journal of Geo-Information*, 10(11), 731. https://doi.org/10.3390/ijgi10110731

Thissen, M., de Graaff, T., & van Oort, F. (2016). Competitive Network Positions In Trade And Structural Economic Growth: A Geographically Weighted Regression Analysis For European Regions. *Papers in Regional Science*, *95*(1), 159-180.

Thompson, F. S. (2018). Characterisation Of Heterogeneity And Spatial Autocorrelation In Phase Separating Mixtures Using Moran's. *Journal of Colloid and Interface Science*. Vol. 513, pp. 180-187.

Tukey, J. W. (1960). Conclusions vs decisions. *Technometrics*, *2*(4), 423-433.

Tyler, D. E. (2008). Robust Statistics: Theory And Methods. Journal of the American Statistical Association, 103(482), 888-889.

Vogelsang, T. J. (2012). Heteroskedasticity, Autocorrelation, And Spatial Correlation Robust Inference In Linear Panel Models With Fixed-Effects. *Journal of Econometrics*, *166*(2), 303-319.

Weisberg, S. (2005). Applied Linear Regression. 3rd ed. Canada : JohnWiley and Sons Inc.

Wong, W., & Bian, G. (2000). Robust estimation in Capital Asset Pricing Model. *Journal of Applied Mathematics and Decision Science*. 4. 65-82. https://doi.org/10.1155/S1173912600000043.

Wong, W., Leung, P., & Ng, H. Y. (2011). An Improved Estimation to Make Markowitz's Portfolio Optimization Theory Users Friendly and Estimation Accurate with Application on the US Stock Market Investment. *European Journal of Operational Research*. https://doi.org/222. 10.2139/ssrn.1968889.

Wrenn, D. H., & Sam, A. G. (2014). Geographically And Temporally Weighted Likelihood Regression: Exploring The Spatiotemporal Determinants Of Land Use Change. *Regional Science and Urban Economics*, *44*, 60-74.

Wu, S., Wang, Z., Du, Z., Huang, B., Zhang, F., & Liu, R. (2020). Geographically And Temporally Neural Network Weighted Regression For Modelling Spatiotemporal Non-Stationary Relationships. *International Journal of Geographical Information Science*, *35*(3), 582-608.

Xu, B., Xu, L., Xu, R., & Luo, L. (2017). Geographical analysis of CO2 emissions in China's manufacturing industry: A geographically weighted regression model. *Journal of Cleaner Production*, *166*, 628-640.

Xu, P., & Huang, H. (2015). Modelling crash spatial heterogeneity: Random parameter versus geographically weighting. *Accident Analysis & Prevention*, *75*, 16-25.

Xu, Y., Zhang, J., Long, Z., Tang, H., & Zhang, X. (2019). Hourly Urban Water Demand Forecasting Using The Continuous Deep Belief Echo State Network. *Water*, 11(2), 351.

Yohai, V. J. (1987). High Breakdown-Point And High Efficiency Robust Estimates For Regression. *The Annals of statistics*, 642-656.

Yu, D. (2010). Exploring Spatiotemporally Varying Regressed Relationships: The Geographically Weighted Panel Regression Analysis. The International Archives of the Photogrammetry. *Remote Sensing and Spatial Information Sciences*. 38 (2) : 134-139

Yu, D., Zhang, Y., Wu, X., Li, D., & Li, G. (2021). The varying effects of accessing high-speed rail system on China's county development: A geographically weighted panel regression analysis. *Land Use Policy*, *100*, 104935.

Zeleke, A., Miglio, R., Palumbo, P., Malaguti, E., Chiari, L., & Due, U. (2022). Spatiotemporal heterogeneity of sars-cov-2 diffusion at the city level using geographically weighted poisson regression model: the case of bologna, italy. Geospatial Health, 17(2). https://doi.org/10.4081/gh.2022.1145

Zhang, H., Liu, Y., Chen, F., Mi, B., Zeng, L., & Pei, L. (2021). The Effect Of Sociodemographic Factors On Covid-19 Incidence Of 342 Cities In China: A Geographically Weighted Regression Model Analysis. *BMC Infectious Diseases*, 21(1). https://doi.org/10.1186/s12879-021-06128-1

Zhang, H., & Mei, C. (2011). Local Least Absolute Deviation Estimation Of Spatially Varying Coefficient Models: Robust Geographically Weighted Regression Approaches. *International Journal of Geographical Information Science*, *25*(9), 1467-1489. https://doi.org/10.1080/13658816.2010.528420

Zhou, S., Zhou, S., Liu, L., Zhang, M., Kang, M., Xiao, J., & Song, T. (2019). Examining The Effect Of The Environment And Commuting Flow From/To Epidemic Areas On The Spread Of Dengue Fever. *International Journal of Environmental Research and Public Health*, *16*(24), 5013. https://doi.org/10.3390/ijerph16245013.