Modeling Industrial and Maritime Sector Effects on Economic Growth in Riau Island Using GWNR-P Spline

Cinta Rizki Oktarina[1*], Andro Kurniawan[1], Putri Suci Aria[1], Sandy Salomo Saruan[1], and Winalia Agwil[2]

[1] Mathematics Study Program, Faculty of Information Technology, Institut Teknologi Batam, Batam, Indonesia

[2] Data Science Institute, Center for Statistics, Hasselt University, Agoralaan D, Diepenbeek, Belgium

*Corresponding author: cinta@iteba.ac.id

Received: 12 June 2026
Revised: 6 August 2026
Accepted: 7 August 2026

ABSTRACT – Economic growth is an important indicator of regional development and is influenced by various socioeconomic factors. In the Riau Islands Province, the effects of the Labor Force Participation Rate (LFPR) and Capture Fishery Production Value on the Gross Regional Domestic Product (GRDP) percentage distribution may vary across districts and cities due to spatial heterogeneity and nonlinear relationships. Therefore, this study applies the GWNR-PSpline model to analyze these relationships. The optimal Penalized Spline model was obtained using knot points of 67.83 and 15.67 with a smoothing parameter of 50, as indicated by the minimum Generalized Cross Validation (GCV) value of 1245.917. Subsequently, the Manhattan distance and Bisquare kernel weighting scheme produced the minimum Cross Validation (CV) value of 64.51, indicating the lowest prediction error among the evaluated weighting combinations. The estimated local parameters varied across districts and cities, indicating spatial heterogeneity in the relationships between the explanatory variables and economic growth. Spatial analysis revealed that Natuna and Anambas Island exhibited stronger local effects than other regions. Furthermore, the local coefficients of determination ranged from 0.3944 to 0.8603, substantially exceeding the global P-Spline model ($R^2 = 0.158$). These findings demonstrate that the GWNR-PSpline model effectively captures both nonlinear and spatially varying relationships.

Keywords– Capture Fishery Production Value, Economic Growth, GWNR-PSpline, Labor Force Participation Rate, Spatial Heterogeneity.

I. INTRODUCTION

The Riau Islands is one of Indonesia's provinces characterized by an archipelagic landscape, with an economy dominated by the industrial and marine sectors [1]. Its strategic geographical position and proximity to neighboring countries such as Singapore and Malaysia have established the region as an economic growth area grounded in industry and maritime activities [2]. Batam City, Bintan Regency, and Karimun are recognized as industrial centers, whereas areas such as Natuna and Lingga hold substantial potential in the fisheries and marine sectors [3]. These differences in regional characteristics give rise to variations in economic growth across the regencies and cities of the Riau Islands.

Within this regional context, economic growth, as a primary indicator of regional development, is influenced by the interaction of various economic sectors [4]. Previous studies have shown that both industrial and maritime activities contribute to economic growth, although their effects may differ across sectors and regions. [5] found that investment in the manufacturing sector had a positive and significant effect on economic growth in Indonesia. However, these effects were not uniform across economic sectors, indicating that sectoral contributions depend on their respective characteristics. Meanwhile, [6] reported that Indonesia's marine sector contributed to economic output, community income, and employment. Although both studies indicate positive economic contributions, their findings also suggest that sectoral effects cannot be assumed to be uniform across different regional contexts. Differences in infrastructure, market accessibility, resource dependence, geographical conditions, and local economic structures may cause the influence of the industrial and maritime sectors on economic growth to vary across regions. Therefore, these relationships need to be examined by considering the regional heterogeneity of the Riau Islands.

Conventional spatial regression models can account for spatial dependence, but their parameters are generally estimated globally and therefore may not adequately represent variations in relationships across locations [7]. Conventional GWR addresses this limitation by allowing regression coefficients to vary spatially; however, it commonly assumes a linear relationship between the response and predictor variables. Geographically weighted generalized regression extends GWR to accommodate nonnormal response distributions, but it does not specifically address complex nonlinear relationships among continuous variables. These limitations are important in the Riau Islands, where regional characteristics vary considerably and the relationships between economic growth and its predictors may be both spatially heterogeneous and nonlinear [8]. Therefore, the GWNR-P Spline approach is adopted because it combines geographically weighted local estimation with the flexibility of Penalized Spline functions, enabling spatial variation and nonlinear relationship patterns to be modeled simultaneously [9].

Geographically Weighted Nonparametric Regression (GWNR) extends conventional GWR by incorporating a nonparametric framework, allowing the relationship between the response and predictor variables to be modeled without assuming a specific functional form. In this study, the Penalized Spline approach is integrated into GWNR to provide flexible curve estimation while controlling model complexity through a smoothing penalty. Consequently, the resulting GWNR-P Spline model is able to estimate location-specific relationships while capturing nonlinear patterns in the effects of the industrial and maritime sectors on economic growth.

By estimating location-specific relationships, the GWNR-P Spline model allows each regency and city to be represented according to its own geographical and economic characteristics [10], [11]. This feature is particularly relevant for the Riau Islands, where industrial and maritime activities are unevenly distributed across regions. The application of this model is expected to provide a more adaptive and representative description of regional economic conditions. In addition, the integration of spatial and nonparametric approaches contributes methodologically to the analysis of complex spatial data and provides a more comprehensive understanding of economic growth dynamics in archipelagic and maritime regions.

II. LITERATURE REVIEW

A. Nonparametric Penalized Spline Regression

Nonparametric regression is a flexible regression method that does not impose a specific functional form on the relationship between the response variable and the predictors. This method is employed to estimate a regression curve that conforms to complex data patterns without assuming a particular function [12]. The general nonparametric regression model can be written as follows [13]:

$$\boldsymbol{y_i = f(t_i) + \varepsilon_i} \tag{1}$$

One widely used approach is spline regression, which is a regression method based on piecewise polynomial functions that are segmented and continuous. Spline regression offers high flexibility in capturing changes in data patterns through intersection points known as knots [12]. The number of knots is determined based on the pattern changes in the data; therefore, appropriate knot selection is essential to obtaining a well-fitted model. In Penalized Spline regression, in addition to the location and number of knots, the smoothing parameter λ also requires careful consideration. This parameter controls the degree of smoothness of the curve, where a large value of λ produces a smoother curve, while a small value results in a curve that more closely follows the data. This approach is applied so that the model can capture data patterns without experiencing overfitting. In general, the spline function with order m and the $j-th$ knot for each response can be expressed as follows [14]:

$$\boldsymbol{g(t_i) = \delta_0 + \sum_{g=1}^{G}\sum_{m=1}^{M}\left(\delta_{mg}t_{gi}^{m} + \sum_{k=1}^{K}\phi_{gk}\left(t_{gi}-\xi_{gk}\right)_+^{m}\right)} \tag{2}$$

where the truncated function is defined as follows [12]:

$$\boldsymbol{\left(t_{gi}-\xi_{gk}\right)_+^{m} = \begin{cases}\left(t_{gi}-\xi_{gk}\right)^{m}, t_{gi} \geq \xi_{gk} \\ 0, t_{gi} < \xi_{gk}\end{cases}} \tag{3}$$

Thus, the estimated function of y using Penalized Spline regression based on Equation (2) is obtained as follows:

$$\boldsymbol{y_i = \delta_0 + \sum_{g=1}^{G}\sum_{m=1}^{M}\left(\delta_{mg}t_{gi}^{m} + \sum_{k=1}^{K}\phi_{gk}\left(t_{gi}-\xi_{gk}\right)_+^{m}\right) + \varepsilon_i} \tag{4}$$

Previous studies have shown that Penalized Spline regression is effective in capturing complex nonlinear relationships through appropriate knot and smoothing parameter selection [7], [8]. However, these applications were generally formulated as global models, assuming that the relationship between the response and predictor variables was identical across observation locations. Therefore, although Penalized Spline can accommodate nonlinearity, it does not directly capture spatially varying effects, motivating its integration with a geographically weighted framework.

B. Testing for Spatial Heterogeneity

Nonparametric spline regression is used to model the relationship among variables in a flexible and complex manner [11]. However, in spatial data, differences in characteristics across observation regions generally exist, causing the relationship among variables to vary at each location. This condition is known as spatial heterogeneity. Spatial heterogeneity can be tested using the Breusch-Pagan (BP) test. The hypotheses employed are as follows:

$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$

$H_1: there\ exists\ at\ least\ one\ \sigma_1{}^2\ \neq\ \sigma^2$

there exists at least one $\sigma_i{}^2\ \neq\ \sigma^2$

$$BP = \frac{1}{2}\boldsymbol{f^T Z(Z^T Z)^{-1} Z^T f} \sim \chi^2_{(p)} \tag{5}$$

with: $f_i = \frac{e_i^2}{\sigma^2} - 1$

where e_i denotes the residual of the global Penalized Spline model at the i-th observation, $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n} e_i^2$ is the estimated error varience, $\boldsymbol{f} = (f_1, \dots, f_n)'$, and $\boldsymbol{Z}$ is the $n \times p$ matrix containing the variables used to explain the variation in the squared residuals. The testing decision is made by rejecting H_0 when the $p-value\ <\ \alpha$, indicating the presence of spatial heterogeneity [15].

C. Geographically Weighted Nonparametric Regression Penalized-Spline

Spatial data frequently exhibit relationships between the response variable and predictor variables that differ across observation regions. This variation occurs due to differences in geographic characteristics, socio-economic conditions, and local infrastructure development that cause each region to have a unique relationship structure between variables

[16]. The GWNR-PSpline method accommodates two key features simultaneously: the nonlinearity of relationships between variables through the spline function, and the spatial variation in those relationships through geographic weighting. The method allows model parameters to differ at each observation location, thereby enabling spatial variation to be captured more flexibly and locally [17]. Geographic weights are applied to assign greater influence to nearby regions compared to more distant ones, following the principle that spatially proximate observations tend to share more similar characteristics than those located farther apart [18]. The weighting function commonly employed is the kernel function, such as the Gaussian, bisquare, or exponential kernel, with the optimal bandwidth determined using *Cross Validation* (CV) or *Generalized Cross Validation* (GCV) criteria [18]. In general, the GWNR-PSpline model can be written as follows:

$$y_i = \delta_0(u_i, v_i) + \sum_{g=1}^{G} \sum_{m=1}^{M} \left(\delta_{mg}(u_i, v_i) t_{gi}^m + \sum_{k=1}^{K} \phi_{gk}(u_i, v_i) \left(t_{gi} - \xi_{gk}\right)_+^m \right) + \varepsilon_i \tag{6}$$

where $(\boldsymbol{u_i}, \boldsymbol{v_i})$ denotes the geographic coordinates of the $\boldsymbol{i}$-th observation location, $\boldsymbol{\delta_0}(\boldsymbol{u_i}, \boldsymbol{v_i})$ is the spatially varying intercept, $\boldsymbol{\delta_{mg}}(\boldsymbol{u_i}, \boldsymbol{v_i})$ and $\boldsymbol{\phi_{gk}}(\boldsymbol{u_i}, \boldsymbol{v_i})$ are the spatially varying polynomial and spline coefficients respectively, $\boldsymbol{\xi_{gk}}$ is the $\boldsymbol{k}$-th knot for the $\boldsymbol{g}$-th predictor, and $\boldsymbol{\varepsilon_i}$ is the random error term. All model parameters are estimated locally at each observation point, meaning that each region obtains its own model equation that reflects its specific geographic and economic conditions [19].

The local parameters are estimated using weighted penalized least squares at each observation location. The optimal knot points and smoothing parameter are selected based on the minimum GCV value, whereas the optimal distance metric, kernel function, and bandwidth are determined based on the minimum CV value. The estimation is implemented in R using a direct matrix solution; therefore, no iterative convergence criterion is required. The GWNR-P Spline model produces parameter estimates that vary across locations, enabling the relationships among variables to be represented more flexibly and accurately than in global modeling approaches. This characteristic makes the model particularly suitable for the Riau Islands, where the economic contributions of the industrial and maritime sectors vary considerably across districts and citie.

D. Weighting Function and Weight Matrix

Geographic weighting is a fundamental component in the GWNR-PSpline model, as the weight assigned to each observation determines the extent to which that observation influences the local parameter estimation at a given location. The underlying principle is that observations at nearby locations contribute more substantially to local estimation than those situated farther away, a concept formally referred to as spatial decay of influence [20]. The weight value assigned to observation j when estimating the parameter at location i is computed in two sequential steps: (a) calculating the geographic distance between locations using an appropriate distance metric, and (b) transforming the resulting distance into a weight value via a kernel function. Three distance metrics are commonly applied in geographically weighted models: Euclidean distance, Manhattan distance, and Haversine distance. The choice of metric should reflect the geographic context and the nature of the study area [21].

a) Euclidean Distance

Euclidean distance measures the straight-line distance between two coordinate points and is the most widely applied distance metric in spatial analysis. For observation locations i and j with geographic coordinates $(u_i,\ v_i)$ and (u_j, v_j), the Euclidean distance is defined as follows [22]:

$$d_{(ij)} = \sqrt{\left(\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2\right)} \tag{7}$$

Euclidean distance is appropriate for geographically compact study areas where the curvature of the Earth can be disregarded without significant loss of accuracy.

b) Manhattan Distance

Manhattan distance, also referred to as city-block distance, computes the sum of the absolute differences of the coordinate components between two locations. It is defined as follows:

$$d_{(ij)} = |u_i - u_j| + |v_i - v_j| \tag{8}$$

Manhattan distance is particularly suited to contexts where movement between locations follows a rectilinear path, such as transportation networks with grid-like road structures. In regional spatial analysis, this metric may better reflect accessibility patterns constrained by road infrastructure [23].

c) Haversine Distance

Haversine distance accounts for the spherical curvature of the Earth when computing the great-circle distance between two points defined by their latitude and longitude coordinates [24]. This metric is particularly appropriate for geographically dispersed study areas such as archipelagic regions. The Haversine distance is defined as follows:

$$d_{(ij)} = 2r\ \mathrm{asin}\left(\sqrt{\left(\sin^2\left(\frac{(\Delta u\ )}{2}\right) + \cos(u_i)\cos(u_j)\sin^2\left(\frac{(\Delta v\ )}{2}\right)\right)}\right) \tag{9}$$

where r denotes the mean radius of the Earth (6,371 km), $\Delta u\ =\ u_j - u_i$ is the difference in latitude (in radians), and $\Delta v\ =\ v_j - v_i$ is the difference in longitude (in radians). Given that the Riau Islands Province consists of geographically dispersed islands with considerable inter-district distances, the Haversine distance provides a geometrically accurate representation of great-circle distance. Nevertheless, the most appropriate distance metric for geographically weighted modeling should also consider the underlying spatial interactions represented by the data, particularly in regions where accessibility is influenced by inter-island transportation networks and port connectivity rather than solely by straight-line geographic proximity. Therefore, this study compares Euclidean, Manhattan, and Haversine distances, and the final metric is selected based on the Cross Validation (CV) criterion together with its suitability for representing regional accessibility patterns.

Geographic weighting in the GWNR-PSpline model is implemented through kernel functions that transform computed geographic distances into spatial weight values. Two fundamental kernel types are commonly distinguished: fixed kernel, which applies a constant bandwidth value h uniformly across all observation locations ($h_i = h, \forall i = 1,2,\dots,n$), and adaptive kernel, which adjusts the bandwidth individually at each location based on local observation density. Given the highly uneven spatial distribution of districts in Riau Islands Province ranging from the densely populated Batam metropolitan area to the remote islands of Natuna and Kepulauan Anambas the adaptive kernel is considered more appropriate for this study.

Each kernel type (fixed or adaptive) can be implemented using one of several kernel functions that define the rate at which spatial weights decay with increasing distance. Three commonly employed kernel functions in geographically weighted models are the Gaussian, Bisquare, and Tricube kernels [25].

a) Gaussian Kernel

The Gaussian kernel is a continuous, smooth kernel function that assigns positive weights to all observation locations regardless of distance [26]. The weight decreases exponentially as a function of the squared distance-to-bandwidth ratio. The Gaussian kernel function is defined:

$$w_{ij} = exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right) \tag{10}$$

A notable property of the Gaussian kernel is that it never assigns zero weight to any observation, meaning all locations regardless of distance contribute to the local parameter estimation, albeit with diminishing influence.

b) Bisquare Kernel

The Bisquare kernel is a compactly supported kernel function that assigns zero weight to observations beyond the bandwidth threshold, thereby confining the estimation influence to a defined spatial neighborhood [27]. It is defined as:

$$w_{ij} = \begin{cases} \left(1-\left(\frac{d_{ij}}{h}\right)^2\right)^2, & d_{ij} \le h \\ 0, & d_{ij} > h \end{cases} \tag{11}$$

The compact support property of the Bisquare kernel ensures computational efficiency and produces well-defined local neighborhoods, making it particularly useful when a clear spatial cutoff is desired.

c) Tricube Kernel

The Tricube kernel is also a compactly supported kernel function, similar to the Bisquare but employing a cubic rather than quadratic decay structure. It is defined as:

$$w_{ij} = \begin{cases} \left(1-\left(\frac{d_{ij}}{h}\right)^3\right)^3, & d_{ij} \le h \\ 0, & d_{ij} > h \end{cases} \tag{12}$$

The Tricube kernel produces a more gradual weight decay compared to the Bisquare, offering a smoother transition in the weighting scheme near the bandwidth boundary.

The spatial weights computed for all n observation locations relative to estimation point i are organized into a diagonal weight matrix $\boldsymbol{W}(u_i, v_i)$ of dimension $n \times n$, defined as follows:

$$\boldsymbol{W}(u_i, v_i) = diag(w_{i1}, w_{i2}, \dots, w_{in})$$

Where w_{ij} denotes the weight assigned to observation j when estimating the model parameters at location i. The diagonal structure of $\boldsymbol{W}(u_i, v_i)$ reflects the assumption that observations are spatially independent, with the influence of each observation encoded entirely in its corresponding diagonal element. Since the weight matrix is constructed separately for each estimation location $i = 1,2,\dots,n$, the GWNR-PSpline model yields a distinct set of locally estimated parameters for each district or city, thereby accommodating the spatially varying economic characteristics of Riau Islands Province.

III. METHODOLOGY

This study employs a quantitative approach with a nonparametric spatial analysis framework, utilizing cross-sectional secondary data for the year 2025 obtained from the Central Bureau of Statistics (BPS) of Riau Islands Province. The study covers seven districts and cities: Batam, Tanjungpinang, Bintan, Karimun, Natuna, Lingga, and Kepulauan Anambas. The seven observations represent the complete administrative population of the Riau Islands Province rather than a sample drawn from a larger regional population. Nevertheless, the limited number of spatial units may affect the stability of local parameter estimates, particularly because the GWNR-P Spline model estimates location-specific coefficients. Therefore, the resulting estimates should be interpreted cautiously as an exploratory description of spatial variation across the regencies and cities. This limitation is acknowledged, and future studies are expected to use broader spatial coverage or repeated observations over time to improve estimation stability.The dependent variable is percentage distribution of GRDP at current prices, while the independent variables consist of the Labor Force Participation Rate (LFPR) as a proxy for the industrial sector and the value of capture fishery production as a proxy for the maritime sector. Geographic coordinate data for each district and city are additionally used for the spatial component of the model. The analytical procedure consists of five sequential stages: (1) descriptive statistical analysis to characterize variable distributions across districts; (2) global Penalized Spline (P-Spline) modeling as a baseline specification with optimal knot selection via GCV; (3) Breusch-Pagan test to formally detect spatial heterogeneity; (4) GWNR-PSpline modeling with locally estimated parameters using CV-optimized bandwidth; and (5) spatial mapping of the estimated coefficients to identify district-level variation in the influence of the industrial and maritime sectors on regional economic growth.

IV. RESULTS AND DISCUSSIONS

A. *Spatial Distribution of GRDP and Relationship Pattern Analysis*

The percentage distribution of Gross Regional Domestic Product at Current Prices (GRDP-CP) across districts and cities in Riau Islands Province is presented in Figure 1. Natuna recorded the highest distribution, ranging from 8.27% to 66.10%, while Lingga and Kepulauan Anambas fell within the lowest range. Natuna's distinct position may be associated with its different economic structure, particularly the strong contribution of natural-resource-based activities, including the oil and gas sector, to regional output[28]. In addition, its remote geographical location and limited connectivity with the main industrial centers of the Riau Islands distinguish its economic characteristics from those of the other districts and cities. These structural and geographical differences may explain the substantially different GRDP distribution observed in Natuna. The substantial variation across districts indicates an uneven regional economic contribution, likely driven by differences in industrial and maritime sector characteristics, and serves as preliminary evidence of spatial heterogeneity among observation units.

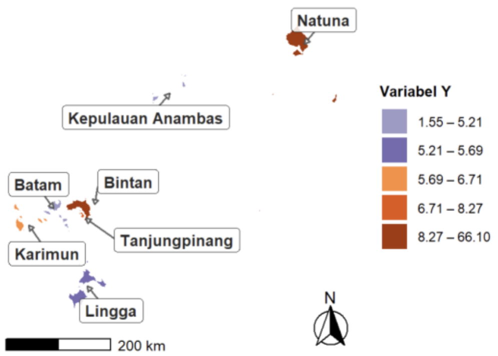


Figure 1. Thematic Map of GRDP in Riau Islands Province, 2025

The relationship pattern between the study variables was examined using the Terasvirta linearity test. The test results indicate that both the Labor Force Participation Rate (LFPR) and the value of capture fishery production yielded p-values less than 0.05, leading to the rejection of the null hypothesis of linearity. This confirms that the relationships between each independent variable and GRDP-CP are nonlinear in nature, necessitating the use of a nonparametric regression approach in the subsequent modeling process.

B. *Penalized Spline Regression*

Penalized Spline regression modeling was conducted to obtain an initial global estimation of the relationship between the percentage distribution of GRDP-CP and the two independent variables, namely LFPR and capture fishery production value. Optimal parameter selection was performed by minimizing the GCV value across all combinations of order, knot configuration, and lambda. The results of the optimal parameter selection are presented in Table 1.

Table 1. Optimal Parameter Selection Results of Penalized Spline Regression

Parameter	Optimal Value
Knot of LFPR (ξ_1)	67.83
Knot of Capture Fishery Production Value (ξ_2)	15.67
Lambda (λ)	50
Minimum GCV	1245.917
Minimum MSE	380.1557
R^2	0.158

Based on the optimal parameters obtained, the global Penalized Spline model is expressed as follows:

$$\hat{y}_i = -246.4082 + 3.9601\, x_{1i} - 0.0237(x_{1i} - 67.83)_+ - 0.3529\, x_{2i} - 0.3259(x_{2i} - 15.67)_+$$

The R^2 value of 0.158 indicates that the global P-Spline model explains only approximately 15.80% of the variation in GRDP-CP percentage distribution across districts and cities. This low explanatory power suggests that a single global relationship is insufficient to represent the substantial differences in economic structure, geographical accessibility, industrial development, and dependence on maritime resources across regions. Because the global model assumes that the effects of LFPR and capture fishery production are identical at all locations, contrasting local effects may be averaged in the estimation, thereby reducing the overall model fit. This result indicates the presence of spatially varying relationships and provides further justification for applying the GWNR-P Spline approach, which allows the model parameters to vary across locations.

C. *Spatial Heterogeneity Test*

The Breusch-Pagan test was conducted to examine whether spatial heterogeneity exists across observation units. The test results are presented in Table 2.

Table 2. Breusch-Pagan Test Results

Statistics	Value
F-statistic	9.2617
p-value	0.0315
Decision	H_0 rejected ($p < 0.05$)
Conclusion	Spatial heterogeneity exists

Since the p-value of 0.0315 is less than 0.05, the null hypothesis of homogeneous error variance is rejected, indicating the presence of spatial heterogeneity among districts and cities in Riau Islands Province. This result confirms that the relationship between the industrial sector, maritime sector, and GRDP-CP percentage distribution varies across locations, thereby providing statistical justification for proceeding to the GWNR-PSpline modeling approach.

The observed spatial heterogeneity may be associated with differences in regional economic structures across the Riau Islands. Batam, Bintan, and Karimun are more strongly supported by industrial, trade, and investment activities, whereas Natuna, Lingga, and Kepulauan Anambas rely more heavily on fisheries, marine resources, and other natural-resource-based activities. Differences in infrastructure, market accessibility, labor absorption, investment intensity, and inter-island connectivity may therefore cause LFPR and capture fishery production to affect GRDP-CP differently across locations. These economic mechanisms explain why the relationships among the study variables are not spatially uniform.

D. *GWNR-PSpline Modeling*

The GWNR-PSpline model was estimated using a fixed kernel approach across various combinations of distance metrics and kernel functions to determine the optimal weighting scheme, where a constant bandwidth value is applied uniformly across all observation locations. The selection of the optimal combination was determined based on the minimum Cross Validation (CV) value, where a smaller CV value indicates a better model fit with lower prediction error. The CV values for each combination are presented in Table 3.

Table 3. CV Values for Each Combination of Distance Metric and Kernel Function

Distance	Weighted		
	Gaussian	Bisquare	Tricube
Harvesine	299.42	232.56	272.99
Manhattan	219.10	**64.51**	67.19
Euclidean	282.68	195.88	233.75

The Manhattan distance combined with the Bisquare kernel yielded the minimum Cross Validation (CV) value of 64.51, indicating the best predictive performance among the evaluated weighting schemes. This result suggests that, although the Haversine distance is geometrically more accurate for measuring inter-island proximity, the Manhattan distance better represents the accessibility-driven spatial interactions among districts in the Riau Islands. Consequently, the Manhattan–Bisquare combination was selected not only because it achieved the lowest CV value, but also because it is theoretically consistent with the regional accessibility patterns of the study area.

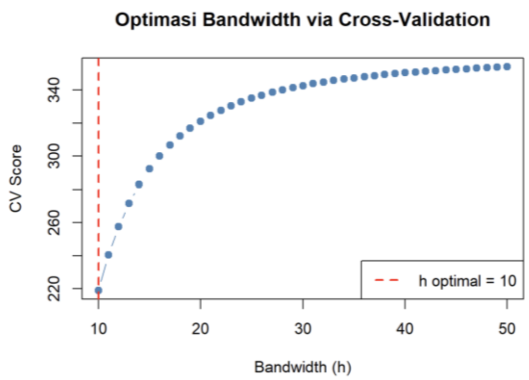


Figure 2. Cross-Validation Bandwidth Optimization Curve for Manhattan-Bisquare Kernel

The optimal bandwidth $h = 10$ is expressed in the same units as the Manhattan distance used in this study, which was calculated from the geographical coordinates of the districts/cities. Based on the optimal bandwidth and the Manhattan–Bisquare weighting scheme, the spatial weight matrix $\mathbf{W}(u_i, v_i)$ was constructed as follows:

$$\mathbf{W}(u_i, v_i) = \begin{bmatrix} 1.0000 & 0.9955 & 0.9823 & 0.7624 & 0.9307 & 0.2608 & 0.9924 \\ 0.9955 & 1.0000 & 0.9641 & 0.8104 & 0.9611 & 0.3197 & 0.9996 \\ 0.9823 & 0.9641 & 1.0000 & 0.6353 & 0.9089 & 0.1367 & 0.9698 \\ 0.7624 & 0.8104 & 0.6353 & 1.0000 & 0.6783 & 0.7781 & 0.7975 \\ 0.9307 & 0.9611 & 0.9089 & 0.6783 & 1.0000 & 0.1740 & 0.9685 \\ 0.2608 & 0.3197 & 0.1367 & 0.7781 & 0.1740 & 1.0000 & 0.3030 \\ 0.9924 & 0.9996 & 0.9698 & 0.7975 & 0.9685 & 0.3030 & 1.0000 \end{bmatrix}$$

The diagonal elements of the weight matrix equal 1.0000, indicating that each location receives full weight when estimating its own parameters. The off-diagonal elements reflect the degree of spatial proximity between districts, where values close to 1 indicate high geographic similarity and values approaching 0 indicate greater spatial distance. Notably, district 6 (Natuna) consistently exhibits the lowest weight values relative to other districts, reflecting its geographic remoteness as the most spatially isolated region in Riau Islands Province. The optimal bandwidth ($h = 10$) was selected using the Cross Validation (CV) criterion because it provides the best balance between local model flexibility and estimation stability by minimizing prediction error. The bandwidth value represents the distance threshold used in the Bisquare kernel weighting function. Although the Bisquare kernel theoretically assigns zero weights to observations located beyond the selected bandwidth, the maximum Manhattan distance among the seven districts/cities (7.939) is smaller than the optimal bandwidth ($h = 10$). Consequently, all observation pairs remain within the effective support of the kernel, resulting in positive spatial weights for every district. Therefore, the compact support property of the Bisquare kernel is preserved, although no observation exceeds the selected bandwidth to receive a zero weight.

E. GWNR-PSpline Local Parameter Estimation

Based on the optimal weighting scheme obtained in the previous stage, the GWNR-PSpline model was estimated locally for each district and city in Riau Islands Province. The locally estimated coefficients for each observation location are presented in Table 4.

Table 4. Local Parameter Estimates of GWNR-PSpline Model

District/City	$\hat{\delta}_0$	$\hat{\delta}_1$	$\hat{\phi}_1$	$\hat{\delta}_2$	$\hat{\phi}_2$	R^2
Batam	−0.0227	0.1219	1.3340	0.2175	−0.7203	0.7440
Bintan	−0.0278	0.1243	1.6042	0.2863	−0.8947	0.6910
Karimun	−0.0117	0.1148	0.7383	0.0626	−0.3143	0.8603
Anambas Island	−0.0791	0.1510	4.1548	0.9365	−2.5017	0.3944
Lingga	−0.0164	0.1155	0.9897	0.1177	−0.4503	0.8246
Natuna	−0.2320	0.1864	10.9035	1.9727	−4.8634	0.7127
Tanjung Pinang	−0.0264	0.1233	1.5257	0.2674	−0.8448	0.7058

Table 4 shows that the estimated local parameters vary across districts and cities in the Riau Islands Province, illustrating spatial variation in the estimated relationships between the predictor variables and the response variable. As an illustration, the local GWNR-P Spline model for Batam is expressed as follows:

$$\hat{y} = -0.0227 + 0.1219x_1 + 1.3340(x_1 - 67.83)_+ + 0.2175x_2 - 0.7203(x_2 - 15.67)_+.$$

Based on the spline model, the segmentation of $x_1$is given by

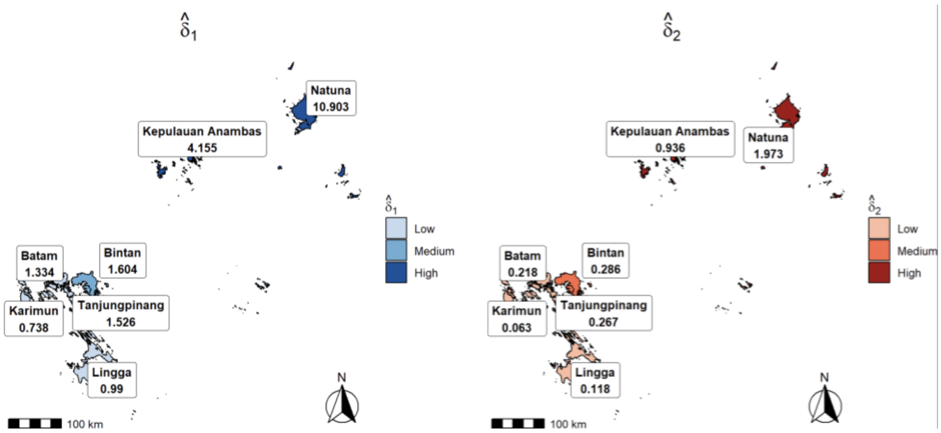
$$f(x_1) = \begin{cases} 0.1219x_1, & x_1 \le 67.83, \\ 1.4559x_1 - 90.4832, & x_1 > 67.83. \end{cases}$$

The estimated relationship between and the response variable differs across segments. In Segment I ($x_1 \le 67.83$), the estimated slope is 0.1219, whereas in Segment II ($x_1 > 67.83$), the estimated slope increases to 1.4559. This pattern suggests that the positive association between LFPR and GRDP-CP becomes stronger after the knot point of 67.83.

Meanwhile, the segmentation of $x_2$is given by

$$f(x_2) = \begin{cases} 0.2175x_2, & x_2 \le 15.67, \\ -0.5028x_2 + 11.2871, & x_2 > 15.67. \end{cases}$$

The estimated relationship between and the response variable also changes across segments. In Segment I ($x_2 \le 15.67$), the estimated slope is 0.2175, indicating a positive association. In Segment II ($x_2 > 15.67$), the estimated slope becomes -0.5028, suggesting that the relationship changes from positive to negative after the knot point. This change may reflect diminishing economic returns when higher fishery production is not accompanied by adequate processing capacity, market access, cold-storage facilities, distribution networks, or other value-added activities. This interpretation is particularly relevant for Batam, whose economy is predominantly driven by the manufacturing and industrial sectors. Compared with these sectors, capture fisheries contribute a relatively smaller share to regional economic output. Consequently, increases in capture fishery production beyond the estimated threshold may not proportionally stimulate GRDP unless they are supported by downstream processing industries, value-added activities, and efficient logistics and distribution systems. This finding suggests that improving the economic contribution of the fishery sector in Batam requires strengthening its downstream industries rather than relying solely on higher capture production.



(a) (b)

Figure 3. Spatial Distribution of Local Coefficients from the GWNR-PSpline Model (a) $\hat{\boldsymbol{\delta}}_1$ and (b) $\hat{\boldsymbol{\delta}}_2$

Figure 3 shows that the estimated local coefficients vary considerably across districts and cities in the Riau Islands Province. Natuna records the largest estimated coefficients, followed by Kepulauan Anambas. Natuna's distinct estimates may be related to its different economic structure, geographical isolation, strong dependence on natural-resource-based activities, and limited connectivity with the main industrial and market centers of the province. These characteristics may cause changes in LFPR and capture fishery production to be associated with larger changes in GRDP-CP than in other regions. In contrast, Karimun, Lingga, Batam, Bintan, and Tanjungpinang exhibit relatively smaller estimated coefficients, indicating that the relationships are not spatially uniform. The nonlinear patterns also have important policy implications. In regions where fishery production remains below the knot point, increasing production may still be associated with improved regional economic output. However, in regions above the knot point, additional production alone may not generate greater economic benefits. Policy interventions should therefore emphasize processing industries, cold-storage facilities, transportation networks, market integration, and value-added fishery products rather than focusing solely on production expansion. Similarly, the stronger estimated association of LFPR after its knot point suggests that improvements in labor participation should be supported by productive employment opportunities and sectoral absorption capacity. The local coefficients of determination range from 0.3944 to 0.8603, exceeding the global P-Spline R^2 of 0.158. This comparison indicates that the local model provides a better descriptive fit by accommodating spatially varying relationships.

V. CONCLUSIONS AND SUGGESTIONS

This study demonstrates that the GWNR-P Spline model can simultaneously accommodate nonlinear and spatially varying relationships between the Labor Force Participation Rate, capture fishery production, and GRDP at current prices across the regencies and cities of the Riau Islands Province. The scientific contribution of this study lies in integrating Penalized Spline regression with a geographically weighted framework, thereby providing a flexible approach for describing relationships that vary across an archipelagic region. The findings indicate that regional development policies should not be applied uniformly because each area has distinct economic structures and geographical characteristics. Labor policies should focus on creating productive employment opportunities, while fishery development should be supported by processing capacity, cold-chain infrastructure, distribution networks, market access, and value-added activities. Nevertheless, this study is limited by the relatively small number of spatial units and the absence of inferential testing for local parameters. Future research should consider broader spatial coverage or repeated observations over time and develop inferential procedures for estimating standard errors, confidence intervals, and the statistical significance of local parameters in the GWNR-P Spline model.

ACKNOWLEDGEMET

The authors would like to express their sincere gratitude to the Institut Teknologi Batam (ITEBA) for providing financial support through the Internal Research Grant (Penelitian Dosen Internal) scheme under the Hibah Penelitian Internal Institut Teknologi Batam, with Contract Number 002/LPPM/KP-ITEBA/V/2026 and Assignment Letter Number 002/LPPM/ST/PEN/V/2026. The authors also acknowledge the Institute for Research and Community Service (LPPM) of ITEBA for its support and facilitation throughout the completion of this research.

REFERENCES

[1] P. dan P. Badan Perencanaan, "Kajian Optimalisasi Potensi Kemaritiman Provinsi Kepulauan Riau," *Badan Perencanaan, Penelitian dan Pengembangan*, Kepulauan Riau, 2021.

[2] G. Prasiska *et al.*, "Analisis Peran Indonesia Dalam Kerjasama Dengan Singapura Dalam Penetapan Kawasan Spesial Economic Zone Di Wilayah Bintan Karimun," *J. Adm. Negara*, vol. 1, no. 5, pp. 282–291, 2023.

[3] R. J. Siagian, S. Syafri, D. R. Yuliandari, Nurhanifah, and O. Saputra, *Buku Profile Perikanan Kota Batan*, 1st ed. Dinas Perikanan Kota Batam, 2024.

[4] Deffrinica *et al.*, "Analisis Determinan Pertumbuhan Ekonomi Kawasan Perbatasan Di Kalimantan Barat Dengan Data Panel,"

Gorontalo Development Review.
[5] A. M. Fazaalloh, "FDI and economic growth in Indonesia : a provincial and sectoral analysis," *J. Econ. Struct.*, vol. 13, no. 3, p. 1022, 2024, doi: 10.1186/s40008-023-00323-w.
[6] Nurkholis, D. Nuryadin, N. Syaifudin, R. Handika, R. H. Setyobudi, and D. M. Udjianto, "The Economic of Marine Sector in Indonesia," *Aquat. Procedia*, vol. 7, pp. 181–186, 2016, doi: 10.1016/j.aqpro.2016.07.025.
[7] C. R. Oktarina, I. Sriliana, and S. Nugroho, "Penalized Spline Semiparametric Regression for Bivariate Response in Modeling Macro Poverty Indicators," *Indones. J. Appl. Stat.*, vol. 8, no. 1, pp. 13–23, 2025, doi: https://doi.org/10.13057/ijas.v8i1.94370.
[8] R. Pahlepi, I. Sriliana, W. Agwil, and C. R. Oktarina, "Biresponse Spline Truncated Nonparametric Regression Modeling for Longitudinal Data on Monthly Stock Prices of Three Private Banks in Indonesia," *Barekeng*, vol. 19, no. 4, pp. 2467–2480, 2025, doi: 10.30598/barekengvol19iss4pp2467-2480.
[9] S. Nugroho, D. S. Rini, T. Jomecho, C. R. Oktarina, S. Cahya, and E. E. Karuna, "Spatial Modeling Of Poverty In Bengkulu Province With Mixed Geographically Weighted Regression," vol. 18, no. 2, pp. 759–772, 2024.
[10] L. Laome, I. N. Budiantara, and V. Ratnasari, "Estimation Curve of Mixed Spline Truncated and Fourier Series Estimator for Geographically Weighted Nonparametric Regression," *Mathematics*, vol. 11, no. 1, p. 152, 2023, doi: https://doi.org/10.3390/math11010152.
[11] M. N. Saputri, Sifriyani, and Wasono, "Application Of Nonparametric Geographically Weighted Regression Method On Open Unemployment Rate Data In Indonesia," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, no. 4, pp. 2071–2080, 2023.
[12] I. N. Budiantara, *Regresi Nonparametrik Spline Truncated*. Surabaya: ITS Press, 2019.
[13] I. Sriliana, I. N. Budiantara, and V. Ratnasari, "Poverty Gap Index Modeling in Bengkulu Province using Truncated Spline Regression for Longitudinal Data," *AIP Conf. Proc.*, vol. 2975, no. 1, 2023, doi: 10.1063/5.0181178.
[14] D. Ruppert, M. P. Wand, and R. . Carroll, *Semiparametric Regression*. Cambridge University Press, 2003. doi: 10.1201/9781420091984-c17.
[15] N. F. Gamayanti, Junaidi, Fadjryani, and Nur'eni, "ANALYSIS OF SPATIAL EFFECTS ON FACTORS AFFECTING RICE PRODUCTION IN CENTRAL SULAWESI USING," *Barekeng J. Math. Its Appl.*, vol. 17, no. 1, pp. 361–370, 2023.
[16] A. R. Novianti, Aswi, and Z. Mar'ah, "GEOGRAPHICALLY WEIGHTED LASSO (GWL) MODELING TO IDENTIFY FACTORS INFLUENCE STUNTING INCIDENTS IN SOUTH SULAWESI," *J. Stat. dan Apl.*, vol. 9, no. 1, pp. 102–111, 2025.
[17] R. Putra, M. G. Fadhlurrahman, and Gunardi, "Determination of the best knot and bandwidth in geographically weighted truncated spline nonparametric regression using," *MethodsX*, vol. 10, no. January, p. 101994, 2023, doi: 10.1016/j.mex.2022.101994.
[18] Sifriyani, I. N. Budiantara, K. P. Candra, and M. Putri, "Selection of optimal knot point and best geographic weighted on geographically weighted spline nonparametric regression," *MethodsX*, vol. 13, no. July, 2024, doi: 10.1016/j.mex.2024.102802.
[19] L. Laome, I. N. Budiantara, and V. Ratnasari, "Construction of a geographically weighted nonparametric regression model fit test," *MethodsX*, vol. 12, no. October 2023, 2024, doi: 10.1016/j.mex.2023.102536.
[20] M. Liu, Y. Ge, S. Hu, and H. Hao, "The Spatial Effects of Regional Poverty : Spatial Dependence , Spatial Heterogeneity and Scale Effects," *Int. J. Geo-Information*, vol. 12, no. 501, pp. 1–23, 2023.
[21] C. R. Oktarina, I. Sriliana, E. N. F. Sidik, and M. A. Firmansyah, "Pemodelan Data Geospasial Balita Kurang Gizi Dengan Pendekatan Geographically Weighted Regression Principal Component Analysis," *J. Gaussian*, vol. 13, no. 2, pp. 339–350, 2024, doi: 10.14710/j.gauss.13.2.339-350.
[22] M. Ghaziasgar, H. Mahvash, and M. Peyman, "A New Geometry - Aware non - Euclidean Distance Metric," *Mach. Learn.*, vol. 114, no. 3, pp. 1–27, 2025, doi: 10.1007/s10994-024-06702-z.
[23] B. Uyanık and G. K. Orman, "A Manhattan distance based hybrid recommendation system," *Int. J. Appl. Math. Electron. Comput.*, vol. 11, no. 1, pp. 20–29, 2023, doi: 10.18100/ijamec.1232090.
[24] E. O. Helmi, O. Emam, and M. Abdel-salam, "Deep Learning Framework for Locating Physical Internet Hubs using Latitude and Longitude Classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 7, pp. 235–240, 2022.
[25] K. D. Widiyanti, A. Fadlurohman, D. Purwanto, U. M. Semarang, and A. Information, "Geographically Weighted Regression Modeling Using Fixed and Adaptive Kernel Weights for the Human Development Index Case in West Java Province," *J. Data Insights*, vol. 3, no. 2, pp. 92–103, 2025.
[26] S. M. Meliyana, R. Rusli, and A. Rahman, "Geographically Weighted Regression with Bi- Square Kernel Weights for Life Expectancy Data in East Java Province," *ARRUS J. Math. Appl. Sci.*, vol. 4, no. 2, pp. 56–62, 2024.
[27] M. Govorov, G. Beconytė, and G. Gienko, "Based Statistical Learning for Selecting Kernel Density Estimates of Spatial Point Patterns," *Trans. GIS*, vol. 29, no. 2, pp. 1–29, 2025, doi: 10.1111/tgis.70051.
[28] B. P. S. K. Natuna, *Analisis Sektor Unggulan Kabupaten Natuna 2021*. Ranai: Badan Pusat Statistik Kabupaten Natuna, 2022.

© 2026 by the authors. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (http://creativecommons.org/licenses/by-sa/4.0/).