CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 11 (2) (2026), Pages 1580–1599
p-ISSN: 2086-0382; e-ISSN: 2477-3344

A Multiscale Extension of Geographically Weighted Spline Nonparametric Regression: Model Formulation and Theoretical Properties

Indi Rizqy Fahrani, Henny Pramoedyo, and Atiek Iriany*

Department of Statistics, Faculty of Mathematics and Natural Sciences, Brawijaya University, Malang, Indonesia

Abstract

This study proposes a Multiscale Geographically Weighted Spline Nonparametric Regression (MS-GWSNR) model to simultaneously accommodate spatial heterogeneity, multiscale spatial variation, and nonlinear relationships within a unified regression framework. The proposed model extends Geographically Weighted Spline Nonparametric Regression (GWSNR) by incorporating component-specific bandwidths adapted from the Multiscale Geographically Weighted Regression (MGWR) approach. Parameter estimation is developed using a Weighted Least Squares (WLS) framework combined with a backfitting algorithm to address the absence of a closed-form estimator under multiple spatial weighting matrices. The theoretical properties of the estimator are derived through a smoothing matrix representation, including the smoothing-bias representation, variance, and Mean Squared Error (MSE). A numerical illustration using simulated spatial data with controlled multiscale heterogeneity shows that the proposed estimator converges stably within 12 iterations, with a final Score of Change (SOC) value of 2.48×10^{-5}. The estimator also recovers the underlying response surface closely, with an MAE of 0.0680 and an RMSE of 0.0880. These results indicate that the proposed MS-GWSNR model can be estimated stably and is able to reconstruct nonlinear spatial relationships across different spatial scales.

Keywords: Backfitting Algorithm; GWSNR; MGWR; spatial Nonparametric Regression; Spline Regression.

1. Introduction

Spatial regression analysis has received increasing attention due to its ability to model spatially varying relationships and accommodate spatial heterogeneity in geographic data [1]. Conventional global regression models assume constant relationships between response and predictor variables across all locations, an assumption that is often unrealistic for spatial data exhibiting heterogeneous local processes [2]. To address this issue, Geographically Weighted Regression (GWR) was developed by allowing regression parameters to vary spatially according to observation locations [3]. Consequently, GWR is capable of capturing local spatial variation more effectively than global regression models [1].

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

Received: May 26, 2026 Revised: June 17, 2026 Accepted: June 19, 2026
DOI: https://doi.org/10.18860/cauchy.v11i2.42954

Despite its flexibility, GWR employs a single bandwidth to represent the overall spatial scale of the process, whereas different predictors may operate at different spatial scales [4]. To overcome this limitation, Multiscale Geographically Weighted Regression (MGWR) was introduced by allowing each parameter to have its own bandwidth [5]. Recent studies have shown that MGWR outperforms GWR in identifying spatial heterogeneity and variable-specific spatial scales [6–10]. The development of multiscale spatial models has also continued to expand in order to improve parameter estimation accuracy and interpretation of local spatial processes [11–13].

Although multiscale approaches improve the representation of spatial variation, most existing geographically weighted models still assume linear relationships among variables. In practice, some predictors may exhibit nonlinear patterns that cannot be adequately represented using parametric assumptions [14]. Therefore, nonparametric regression approaches have been widely adopted because they provide greater flexibility in modeling unknown functional relationships [15]. One commonly used approach is truncated spline regression, which is characterized by high flexibility and good adaptability to local data characteristics [16]. The integration of spline-based nonparametric regression into geographically weighted regression led to the development of Geographically Weighted Spline Nonparametric Regression (GWSNR), which accommodates both spatial heterogeneity and nonlinear relationships [17]. Previous studies show that knot selection and bandwidth determination substantially affect the resulting model performance [18].

Nevertheless, several limitations remain. MGWR is generally restricted to linear relationships, whereas complex spatial processes are often influenced by nonlinear mechanisms [19]. Conversely, GWSNR still employs a single bandwidth for all model parameters, making it unable to accommodate variable-specific spatial scales among predictors [20, 21]. Therefore, the development of a model capable of simultaneously accommodating spatial heterogeneity, multiscale spatial variation, and nonlinear relationships remains an important challenge in spatial nonparametric regression.

Based on these issues, this study proposes a Multiscale Geographically Weighted Spline Nonparametric Regression (MS-GWSNR) model by integrating a multiscale bandwidth framework into the GWSNR model. A parameter estimation procedure based on the backfitting algorithm is developed to address the estimation complexity arising from the use of multiple bandwidths. Furthermore, the theoretical properties of the resulting estimator are investigated through smoothing matrix representation, bias analysis, variance analysis, and Mean Squared Error (MSE) derivation. Numerical illustrations using simulated multiscale spatial data are also conducted to evaluate the behavior of the proposed estimator and its capability in reconstructing underlying spatial patterns.

The remainder of this paper is organized as follows. Section 2 describes the research methodology employed in this study. Section 3 presents the proposed model formulation, parameter estimation procedure, theoretical properties of the estimator, and the results of the numerical illustration. Finally, Section 4 provides the conclusions of the study.

2. Methods

This study was conducted through several stages to develop the Multiscale Geographically Weighted Spline Nonparametric Regression (MS-GWSNR) model. The research stages are described as follows.

1. First, the MS-GWSNR model was constructed by integrating the multiscale bandwidth approach from MGWR, first introduced by [5], into the framework of GWSNR, first introduced by [17]. In this stage, each model component was assumed to have a different spatial bandwidth to accommodate multiscale spatial heterogeneity among predictor variables.
2. Next, a parameter estimation procedure was developed using the Weighted Least Squares (WLS) approach combined with a backfitting algorithm. This approach was employed to address the estimation complexity arising from the use of different bandwidths for

each model component. Therefore, the estimation process was performed iteratively until convergence was achieved.

3. After that, a theoretical study of the resulting estimator was conducted through smoothing matrix representation. The study included the analysis of estimator variance, bias representation, and the derivation of the Mean Squared Error (MSE) estimator to evaluate the mathematical properties of the proposed model.
4. Furthermore, numerical illustrations were conducted using simulated spatial data with a controlled multiscale spatial heterogeneity structure. The simulations were used to illustrate the behavior of the estimator as well as the capability of the MS-GWSNR model in reconstructing the underlying multiscale spatial patterns.
5. Finally, the results of the numerical illustrations were analyzed based on the convergence pattern of the backfitting algorithm, the degree of closeness between the actual and estimated values, and the capability of the model in reconstructing multiscale spatial coefficient surfaces.

3. Results and Discussion

This section presents the formulation of the MS-GWSNR model, the parameter estimation procedure using the WLS approach and the backfitting algorithm, as well as the theoretical study of the estimator through smoothing matrix representation, variance analysis, bias representation, and the derivation of the MSE estimator. In addition, numerical illustrations using multiscale spatial simulated data are presented to evaluate the behavior of the estimator and the capability of the model to reconstruct the underlying spatial patterns.

3.1. Model Formulation

This section formulates a multiscale extension of the GWSNR, referred to as the Multiscale Geographically Weighted Spline Nonparametric Regression (MS-GWSNR). The proposed model is developed by integrating the multiscale bandwidth mechanism of MGWR into the GWSNR framework.

MGWR extends the GWR framework by allowing each predictor variable to operate at a different spatial scale through predictor-specific bandwidths [5]. The MGWR model can be expressed as

$$y_i = \beta_{(bw_0)}(u_i, v_i) + \sum_{k=1}^{p} \beta_{(bw_k)}(u_i, v_i) x_{ik} + \varepsilon_i.$$

where y_i denotes the response variable at the i-th location, (u_i, v_i) represents the spatial coordinates of the i-th observation, $\beta_{(bw_k)}(u_i, v_i)$ denotes the spatially varying coefficient estimated using bandwidth bw_k, x_{ik} is the value of the k-th predictor variable at location i, and ε_i is the random error term. The MGWR framework enables each predictor variable to influence the response variable at different spatial scales. However, the relationship between the response and predictor variables is still assumed to be linear.

To accommodate possible nonlinear relationships, spline nonparametric regression has been incorporated into the GWR framework through the GWSNR model [17], which is formulated as

$$y_i = \sum_{k=1}^{p} \left[\sum_{j=0}^{m} \beta_{kj}(u_i, v_i) x_{ik}^{j} + \sum_{h=1}^{r_k} \beta_{k(m+h)}(u_i, v_i)(x_{ik} - K_{kh})_+ \right] + \varepsilon_i.$$

where $\beta_{kj}(u_i, v_i)$ denotes the spatially varying coefficient associated with the polynomial spline component of order j, $\beta_{k(m+h)}(u_i, v_i)$ represents the coefficient associated with the truncated spline component corresponding to the h-th knot, K_{kh} denotes the knot location for the k-th predictor, and $(x_{ik} - K_{kh})_+ = \max(0, x_{ik} - K_{kh})$. The GWSNR model allows the relationship between the response and predictor variables to vary nonlinearly and spatially. However, all spline basis components within each predictor are still assumed to operate under a common spatial scale through a single bandwidth parameter. Based on the limitations, the proposed MS-GWSNR model incorporates spatial heterogeneity and nonlinear relationships while allowing each predictor to operate at its own spatial scale through predictor-specific bandwidths. Let (u_i, v_i) denote the spatial coordinates of the i-th observation, for $i = 1, 2, \ldots, n$. The MS-GWSNR model is formulated as an additive nonparametric regression model

$$y_i = \sum_{k=1}^{p} f_k(x_{ik}) + \varepsilon_i.$$

where $f_k(\cdot)$ denotes the nonparametric regression function associated with the k-th predictor variable. Each component function $f_k(x_{ik})$ is approximated using a truncated spline basis consisting of polynomial and truncated components. Specifically, the function $f_k(x_{ik})$ is expressed as

$$f_k(x_{ik}) = \sum_{j=0}^{m} \beta_{kj}(u_i, v_i; bw_k) x_{ik}^j + \sum_{h=1}^{r_k} \beta_{k(m+h)}(u_i, v_i; bw_k)(x_{ik} - K_{kh})_+.$$

where $\beta_{kj}(u_i, v_i; bw_k)$ are spatially varying coefficients associated with the polynomial spline components of order $j = 0, 1, \ldots, m$, $\beta_{k(m+h)}(u_i, v_i; bw_k)$ are coefficients associated with the truncated spline components corresponding to the h-th knot, bw_k denotes the bandwidth parameter specific to the k-th predictor, and $(x_{ik} - K_{kh})_+ = \max(0, x_{ik} - K_{kh})$.

Substituting the expression of $f_k(x_{ik})$ into the additive model yields

$$y_i = \sum_{k=1}^{p} \left[\sum_{j=0}^{m} \beta_{kj}(u_i, v_i; bw_k) x_{ik}^j + \sum_{h=1}^{r_k} \beta_{k(m+h)}(u_i, v_i; bw_k)(x_{ik} - K_{kh})_+ \right] + \varepsilon_i.$$

The key feature of the proposed MS-GWSNR model lies in the use of predictor-specific bandwidths bw_k, which are assigned to all spline basis components within each predictor. This structure enables the model to capture heterogeneous spatial scales across predictors rather than across spline components. Consequently, the proposed MS-GWSNR model provides greater flexibility in modeling spatially varying nonlinear relationships by simultaneously accommodating spatial heterogeneity, nonlinear effects, and multiscale spatial variation across predictors.

To ensure that the additive decomposition is mathematically well-defined, an identifiability constraint is imposed on the component functions. Since each spline component f_k includes a constant term through the polynomial basis with $j = 0$, the additive components are subject to the centering constraint

$$\sum_{i=1}^{n} \hat{f}_k(x_{ik}) = 0, \quad k = 1, 2, \ldots, p,$$

while the overall intercept is represented separately. This constraint ensures identifiability of the additive components and eliminates competition between the intercept and individual components during the backfitting procedure [22].

3.2. Parameter Estimation

In the conventional GWSNR, parameter estimation can be performed using a weighted least squares approach with a single spatial weighting matrix. Let $\mathbf{W}(u_i, v_i)$ denote a diagonal matrix containing spatial weights for location i. Assuming that $\mathbf{X}^\top \mathbf{W}(u_i, v_i)\mathbf{X}$ is nonsingular, the weighted least squares estimator can be expressed in closed form as

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

where $\mathbf{X}$ is the design matrix containing the spline basis functions and $\mathbf{Y}$ is the response vector. The use of a single weighting matrix implies that all predictors and their associated spline basis components operate under the same spatial scale. Consequently, parameter estimation can be conducted simultaneously for all model components, resulting in a closed-form solution.

In the proposed MS-GWSNR model, each predictor is associated with its own spatial bandwidth bw_k, which induces a distinct weighting matrix $\mathbf{W}_k(u_i, v_i)$ for each predictor variable. Consequently, the weighted least squares objective function for the k-th component is expressed as

$$Q(\boldsymbol{\beta}_k) = (\mathbf{r}_k - \mathbf{X}_k\boldsymbol{\beta}_k)^\top \mathbf{W}_k(u_i, v_i)\,(\mathbf{r}_k - \mathbf{X}_k\boldsymbol{\beta}_k)\,,$$

where $\mathbf{r}_k = \mathbf{Y} - \sum_{l \neq k} \hat{\mathbf{f}}_l$ is the partial residual representing the portion of the response not explained by the remaining components, and $\mathbf{X}_k\boldsymbol{\beta}_k$ represents the spline basis component associated with the k-th predictor variable. The detailed iterative procedure for computing $\mathbf{r}_k$ is described in the following section.

Unlike the conventional GWSNR case, the presence of multiple weighting matrices prevents the normal equations from being represented under a single weighted least squares system. In particular, the estimator cannot be expressed in the form

$$\left(\mathbf{X}^\top \mathbf{W} \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{Y},$$

since $\mathbf{W}_1 \neq \mathbf{W}_2 \neq \cdots \neq \mathbf{W}_p$. Each weighting matrix is only compatible with the subset of columns in the design matrix corresponding to the associated predictor, making simultaneous estimation infeasible within a single optimization step. To address this issue, parameter estimation is conducted iteratively through a backfitting algorithm, following the multiscale estimation strategy introduced in the MGWR framework by [5].

3.2.1. Backfitting Algorithm

To address the absence of a closed-form solution, parameter estimation in the MS-GWSNR model is carried out using a backfitting algorithm. This approach exploits the additive structure of the model and estimates each component function iteratively while treating the remaining components as fixed.

For the k-th predictor, define the partial residual as

$$\mathbf{r}_k = \mathbf{Y} - \sum_{l \neq k} \hat{\mathbf{f}}_l,$$

which represents the portion of the response not explained by the remaining predictor components. The component $f_k(x_k)$ is then estimated by fitting a locally weighted spline regression of $\mathbf{r}_k$ on x_k. Let $\mathbf{X}_k$ denote the design matrix associated with the spline basis of the k-th predictor. Assuming that $\mathbf{X}_k^\top \mathbf{W}_k(u_i, v_i)\mathbf{X}_k$ is nonsingular, the corresponding parameter estimator is given by

$$\hat{\boldsymbol{\beta}}_k(u_i, v_i) = \left(\mathbf{X}_k^\top \mathbf{W}_k(u_i, v_i)\mathbf{X}_k\right)^{-1} \mathbf{X}_k^\top \mathbf{W}_k(u_i, v_i)\mathbf{r}_k.$$

The fitted component is subsequently updated as

$$\hat{\mathbf{f}}_k = \mathbf{X}_k\hat{\boldsymbol{\beta}}_k.$$

The estimation proceeds iteratively over all predictors $k = 1, 2, \ldots, p$, where each component function is updated sequentially using the most recent estimates of the remaining components. At iteration t, the updating scheme is expressed as

$$\mathbf{r}_k^{(t)} = \mathbf{Y} - \sum_{l \neq k} \hat{\mathbf{f}}_l^{(t)},$$

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_k^{(t+1)}(u_i, v_i) &= \left(\mathbf{X}_k^\top \mathbf{W}_k \mathbf{X}_k\right)^{-1} \mathbf{X}_k^\top \mathbf{W}_k \mathbf{r}_k^{(t)}, \\
\hat{\mathbf{f}}_k^{(t+1)} &= \mathbf{X}_k \hat{\boldsymbol{\beta}}_k^{(t+1)}.
\end{aligned}$$

The backfitting algorithm is assumed to converge to a stable solution through iterative updating of each component function. Following [5], convergence is evaluated using the Score of Change (SOC), defined as

$$\mathrm{SOC} = \sqrt{\frac{\sum_{k=1}^{p}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{f}_{ik}^{\text{new}} - \hat{f}_{ik}^{\text{old}}\right)^2\right)}{\sum_{i=1}^{n}\left(\sum_{k=1}^{p}\hat{f}_{ik}^{\text{new}}\right)^2}} \tag{1}$$

where $\hat{f}_{ik}^{\text{new}}$ and $\hat{f}_{ik}^{\text{old}}$ denote the updated and previous estimates of the k-th component function at the i-th observation, respectively. The iterative process is considered convergent when the SOC value falls below a predefined tolerance level. The complete iterative estimation procedure of the proposed MS-GWSNR model is summarized in Algorithm 1.

Algorithm 1 Backfitting Algorithm for MS-GWSNR Estimation

Require: Response vector $\mathbf{Y}$; predictor variables $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p$; spatial weighting matrices $\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_p$; initial component estimates $\hat{f}_1^{(0)}, \hat{f}_2^{(0)}, \ldots, \hat{f}_p^{(0)}$; convergence tolerance ε

Ensure: Estimated component functions $\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_p$; parameter vectors $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$

```
Initialize iteration counter t ← 0
Compute initial fitted values
repeat
    for k = 1 to p do
        Compute partial residual:
```

$$\mathbf{r}_k^{(t)} = \mathbf{Y} - \sum_{l \neq k} \hat{f}_l^{(t)}$$

```
        Estimate spline coefficients:
```

$$\hat{\beta}_k^{(t+1)} = \left(\mathbf{X}_k^T \mathbf{W}_k \mathbf{X}_k\right)^{-1} \mathbf{X}_k^T \mathbf{W}_k \mathbf{r}_k^{(t)}$$

```
        Update fitted component:
```

$$\hat{f}_k^{(t+1)} = \mathbf{X}_k \hat{\beta}_k^{(t+1)}$$

```
    end for
    Compute Score of Change (SOC):
```

$$\mathrm{SOC} = \sqrt{\frac{\displaystyle\sum_{k=1}^{p}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{f}_{ik}^{\text{new}} - \hat{f}_{ik}^{\text{old}}\right)^2\right)}{\displaystyle\sum_{i=1}^{n}\left(\sum_{k=1}^{p}\hat{f}_{ik}^{\text{new}}\right)^2}}$$

```
    Update iteration counter:
```

$$t \leftarrow t + 1$$

```
until SOC < ε
return f̂_1, f̂_2, ..., f̂_p and β̂_1, β̂_2, ..., β̂_p
```

The estimator obtained from the backfitting procedure is subsequently used to construct the smoothing matrix representation and to derive the theoretical properties of the proposed MS-GWSNR model.

3.3. Theoretical Properties

This section presents the theoretical properties of the proposed MS-GWSNR estimator. The derivation begins with the hat matrix representation of the estimator, followed by the assumptions of the additive model and the error structure used in the estimation procedure. Based on these assumptions, several theoretical properties of the estimator are subsequently derived, including the smoothing-bias representation, variance representation, quadratic trace representation, and the Mean Squared Error (MSE) formulation.

Given the local parameter estimator for the k-th component:

$$\hat{\boldsymbol{\beta}}_k = (\mathbf{X}_k^\top \mathbf{W}_k \mathbf{X}_k)^{-1} \mathbf{X}_k^\top \mathbf{W}_k \mathbf{r}_k$$

the corresponding function estimator can be expressed as:

$$\hat{\mathbf{f}}_k = \mathbf{X}_k \hat{\boldsymbol{\beta}}_k = \mathbf{X}_k (\mathbf{X}_k^\top \mathbf{W}_k \mathbf{X}_k)^{-1} \mathbf{X}_k^\top \mathbf{W}_k \mathbf{r}_k$$

Define the smoothing matrix as:

$$\mathbf{S}_k = \mathbf{X}_k (\mathbf{X}_k^\top \mathbf{W}_k \mathbf{X}_k)^{-1} \mathbf{X}_k^\top \mathbf{W}_k$$

such that:

$$\hat{\mathbf{f}}_k = \mathbf{S}_k \mathbf{r}_k$$

This representation indicates that the MS-GWSNR estimator can be formulated as a linear smoothing operator depending on the local weighting matrix $\mathbf{W}_k$. Since the spatial weighting matrix $\mathbf{W}_k(u_i, v_i)$ changes across target locations in geographically weighted models, the smoothing matrix $\mathbf{S}_k$ is constructed locally and row by row. Specifically, for the i-th location, the fitted value is obtained as

$$\hat{f}_{k,i} = \mathbf{s}_{k,i}^\top \mathbf{r}_k,$$

where $\mathbf{s}_{k,i}^\top$ denotes the i-th row of $\mathbf{S}_k$, computed using the local weighting matrix $\mathbf{W}_k(u_i, v_i)$ centered at location (u_i, v_i). The full smoothing matrix $\mathbf{S}_k$ is then understood as the row-wise assembly of all local smoothing vectors. Consequently, different bandwidths bw_k are assigned to different predictors under the multiscale framework, so that the local smoothing effect varies across predictor components. In general, $\mathbf{S}_k$ is not symmetric because of the spatial weighting structure involved in the estimation process.

Suppose the additive MS-GWSNR model is given by:

$$\mathbf{Y} = \sum_{l=1}^{p} \mathbf{f}_l + \boldsymbol{\varepsilon}$$

where $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^\top$ denotes the random error vector assumed to be independently distributed with:

$$E(\boldsymbol{\varepsilon}) = \mathbf{0}, \qquad \mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n$$

where $\mathbf{I}_n$ is the identity matrix of size $n \times n$.

The validity of the proposed MS-GWSNR estimator relies on the following assumptions: (1) the random errors ε_i are independently distributed with zero mean and constant variance σ^2, as stated above; (2) the design matrix $\mathbf{X}_k^\top \mathbf{W}_k(u_i, v_i)\mathbf{X}_k$ is nonsingular for each predictor k and location (u_i, v_i), ensuring the existence of the local parameter estimator; and (3) the backfitting algorithm converges to a stable solution, such that the approximation $\mathbf{r}_k \approx \mathbf{f}_k + \boldsymbol{\varepsilon}$ holds at convergence. In the numerical illustration, the simulated data were generated to satisfy assumptions (1) and (3) by design, with errors drawn from $\varepsilon_i \sim N(0, \sigma^2)$ and convergence verified through the SOC criterion. In empirical applications, where these conditions are not guaranteed by construction, the underlying assumptions can be assessed through standard diagnostics: the error assumptions via residual analysis for homoscedasticity and normality, the independence assumption via tests for residual spatial autocorrelation such as Moran's I, and the local nonsingularity condition via the condition number of $\mathbf{X}_k^\top \mathbf{W}_k \mathbf{X}_k$ to detect local multicollinearity.

Under the convergence condition of the backfitting algorithm, the partial residual satisfies the approximation:

$$\mathbf{r}_k \approx \mathbf{f}_k + \boldsymbol{\varepsilon} \tag{2}$$

Based on these assumptions, the theoretical properties of the MS-GWSNR estimator are derived as follows. Note that since the backfitting procedure relies on the approximation in Equation (2), the estimator exhibits a smoothing bias rather than exact unbiasedness, as formalized in Theorem 1.

Theorem 1. *Under the convergence approximation $\mathbf{r}_k \approx \mathbf{f}_k + \boldsymbol{\varepsilon}$, the conditional expectation of the estimator $\hat{\mathbf{f}}_k = \mathbf{S}_k \mathbf{r}_k$ is given by*

$$E(\hat{\mathbf{f}}_k \mid \mathbf{X}_k) = \mathbf{S}_k \mathbf{f}_k,$$

so that the smoothing bias is represented as

$$\mathrm{Bias}(\hat{\mathbf{f}}_k) = (\mathbf{S}_k - \mathbf{I}_n)\mathbf{f}_k.$$

Proof. From the estimator representation and the convergence approximation:

$$\hat{\mathbf{f}}_k = \mathbf{S}_k \mathbf{r}_k \approx \mathbf{S}_k(\mathbf{f}_k + \boldsymbol{\varepsilon}) = \mathbf{S}_k \mathbf{f}_k + \mathbf{S}_k \boldsymbol{\varepsilon}.$$

Taking expectation on both sides and using $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ yields:

$$E(\hat{\mathbf{f}}_k \mid \mathbf{X}_k) = \mathbf{S}_k \mathbf{f}_k.$$

Therefore, the bias of the estimator is:

$$\mathrm{Bias}(\hat{\mathbf{f}}_k) = E(\hat{\mathbf{f}}_k \mid \mathbf{X}_k) - \mathbf{f}_k = \mathbf{S}_k \mathbf{f}_k - \mathbf{f}_k = (\mathbf{S}_k - \mathbf{I}_n)\mathbf{f}_k. \qquad \square$$

Theorem 2. *The variance of the estimator is given by:*

$$\mathrm{Var}(\hat{\mathbf{f}}_k) = \sigma^2 \mathbf{S}_k \mathbf{S}_k^\top$$

Proof. From:

$$\hat{\mathbf{f}}_k = \mathbf{S}_k \mathbf{f}_k + \mathbf{S}_k \boldsymbol{\varepsilon}$$

since $\mathbf{S}_k \mathbf{f}_k$ is deterministic, it follows that:

$$\mathrm{Var}(\hat{\mathbf{f}}_k) = \mathrm{Var}(\mathbf{S}_k \boldsymbol{\varepsilon})$$

Using the property:

$$\mathrm{Var}(A\boldsymbol{\varepsilon}) = A\,\mathrm{Var}(\boldsymbol{\varepsilon})A^\top$$

We obtain:

$$\mathrm{Var}(\hat{\mathbf{f}}_k) = \mathbf{S}_k(\sigma^2 \mathbf{I}_n)\mathbf{S}_k^\top = \sigma^2 \mathbf{S}_k \mathbf{S}_k^\top. \qquad \square$$

Lemma 1. *For any vector $x \in \mathbb{R}^n$ and matrix $A \in \mathbb{R}^{n\times n}$, the following identity holds:*

$$x^\top Ax = \mathrm{tr}(Axx^\top)$$

Proof. Let:

$$x = (x_1, \dots, x_n)^\top$$

and:

$$A = (a_{ij})$$

Then:

$$x^\top Ax = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j$$

On the other hand:

$$xx^\top = (x_ix_j)$$

so that:

$$(Axx^\top)_{ii} = \sum_{j=1}^{n} a_{ij}x_jx_i$$

Taking the trace yields:

$$\mathrm{tr}(Axx^\top) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j$$

Therefore:

$$x^\top Ax = \mathrm{tr}(Axx^\top). \qquad \square$$

Theorem 3. *The Mean Squared Error (MSE) of the estimator is given by:*

$$\mathrm{MSE}(\hat{\mathbf{f}}_k) = \|(\mathbf{S}_k - \mathbf{I}_n)\mathbf{f}_k\|^2 + \sigma^2\mathrm{tr}(\mathbf{S}_k\mathbf{S}_k^\top)$$

Proof. By definition:

$$\mathrm{MSE} = E[(\hat{\mathbf{f}}_k - \mathbf{f}_k)^\top(\hat{\mathbf{f}}_k - \mathbf{f}_k)]$$

Substituting:

$$\hat{\mathbf{f}}_k - \mathbf{f}_k = (\mathbf{S}_k - \mathbf{I}_n)\mathbf{f}_k + \mathbf{S}_k\boldsymbol{\varepsilon}$$

Let:

$$a = (\mathbf{S}_k - \mathbf{I}_n)\mathbf{f}_k, \qquad b = \mathbf{S}_k\boldsymbol{\varepsilon}$$

Then:

$$(a+b)^\top(a+b) = a^\top a + 2a^\top b + b^\top b$$

Thus:

$$\mathrm{MSE} = E(a^\top a) + 2E(a^\top b) + E(b^\top b)$$

For the first term:

$$E(a^\top a) = a^\top a = \|(\mathbf{S}_k - \mathbf{I}_n)\mathbf{f}_k\|^2$$

For the second term:

$$E(a^\top b) = a^\top E(b)$$

Since:

$$E(b) = \mathbf{S}_kE(\boldsymbol{\varepsilon}) = \mathbf{0}$$

then:

$$E(a^\top b) = 0$$

For the third term:

$$E(b^\top b) = E[(\mathbf{S}_k\boldsymbol{\varepsilon})^\top(\mathbf{S}_k\boldsymbol{\varepsilon})]$$

Using the transpose property:

$$(\mathbf{S}_k\boldsymbol{\varepsilon})^\top = \boldsymbol{\varepsilon}^\top\mathbf{S}_k^\top$$

gives:

$$(\mathbf{S}_k\boldsymbol{\varepsilon})^\top(\mathbf{S}_k\boldsymbol{\varepsilon}) = \boldsymbol{\varepsilon}^\top\mathbf{S}_k^\top\mathbf{S}_k\boldsymbol{\varepsilon}$$

Applying Lemma 1:

$$\boldsymbol{\varepsilon}^\top\mathbf{S}_k^\top\mathbf{S}_k\boldsymbol{\varepsilon} = \mathrm{tr}(\mathbf{S}_k^\top\mathbf{S}_k\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top)$$

Taking expectation:

$$E(b^\top b) = E[\mathrm{tr}(\mathbf{S}_k^\top\mathbf{S}_k\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top)]$$

Using the property:

$$E[\mathrm{tr}(X)] = \mathrm{tr}(E(X))$$

yields:

$$E(b^\top b) = \mathrm{tr}(\mathbf{S}_k^\top\mathbf{S}_k E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top))$$

Since:

$$E(\boldsymbol{\varepsilon}) = \mathbf{0} \Rightarrow E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top) = \mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}_n$$

then:

$$E(b^\top b) = \mathrm{tr}(\mathbf{S}_k^\top\mathbf{S}_k\sigma^2\mathbf{I}_n) = \sigma^2\mathrm{tr}(\mathbf{S}_k^\top\mathbf{S}_k)$$

Using the trace property:

$$\mathrm{tr}(\mathbf{S}_k^\top\mathbf{S}_k) = \mathrm{tr}(\mathbf{S}_k\mathbf{S}_k^\top)$$

Combining all terms gives:

$$\mathrm{MSE}(\hat{\mathbf{f}}_k) = \|(\mathbf{S}_k - \mathbf{I}_n)\mathbf{f}_k\|^2 + \sigma^2\mathrm{tr}(\mathbf{S}_k\mathbf{S}_k^\top). \qquad \square$$

The MSE expression in Theorem 3 reveals that the estimation error of the MS-GWSNR estimator decomposes into two components: a squared bias term $\|(\mathbf{S}_k - \mathbf{I}_n)\mathbf{f}_k\|^2$ and a variance term $\sigma^2\,\mathrm{tr}(\mathbf{S}_k\mathbf{S}_k^\top)$. Both components depend on the smoothing matrix $\mathbf{S}_k$, which is itself determined by the bandwidth bw_k through the spatial weighting matrix $\mathbf{W}_k$.

The effect of the bandwidth on this trade-off can be understood as follows. A smaller bandwidth bw_k produces highly localized kernel weights, causing $\mathbf{W}_k$ to concentrate mass near each target location. In this case, $\mathbf{S}_k$ approaches the identity matrix $\mathbf{I}_n$, which drives the bias term $\|(\mathbf{S}_k - \mathbf{I}_n)\mathbf{f}_k\|^2$ toward zero. However, the variance term $\sigma^2\,\mathrm{tr}(\mathbf{S}_k\mathbf{S}_k^\top)$ increases correspondingly, as the estimator relies on fewer effective observations and becomes more sensitive to noise. Conversely, a larger bandwidth bw_k produces smoother estimates by incorporating information from more distant observations, reducing the variance at the cost of increased bias.

A key advantage of the proposed MS-GWSNR model is that each predictor variable x_k is associated with its own bandwidth parameter bw_k, allowing the bias–variance trade-off to be managed independently for each additive component. This flexibility enables the model to simultaneously accommodate predictors that operate at different spatial scales, assigning larger bandwidths to components with smooth global variation and smaller bandwidths to components with rapid local variation, which is not achievable under conventional GWSNR with a single global bandwidth.

3.4. Numerical Illustration

To illustrate the performance of the proposed estimator, a numerical illustration was conducted using simulated data generated from a multiscale spatial nonparametric process. The simulation was designed to evaluate the capability of the proposed MS-GWSNR model in capturing spatial heterogeneity with different levels of smoothness across model components.

A total of $n = 200$ observations were randomly generated over a two-dimensional spatial domain $(u_i, v_i) \in [0, 1] \times [0, 1]$. This sample size is considered sufficient for the present illustration, given that a first-order truncated spline basis with a single knot yields only three basis functions per predictor, resulting in six estimable parameters per location across two predictors. The primary purpose of this numerical illustration is to serve as a proof-of-concept demonstration rather than a large-scale simulation study, and $n = 200$ provides adequate observations to support stable estimation and convergence of the backfitting algorithm. Two predictor variables, namely x_1 and x_2, were independently generated from the Uniform$(0, 1)$ distribution. The response variable was generated from the additive nonparametric model:

$$y_i = f_1(x_{1i}, u_i, v_i) + f_2(x_{2i}, u_i, v_i) + \varepsilon_i,$$

where

$$\varepsilon_i \sim N(0, \sigma^2),$$

with $\sigma = 0.1$.

The first component was designed to exhibit relatively smooth spatial variation, whereas the second component was constructed to contain more rapidly changing local spatial variation. The data generating process (DGP) was specified as:

$$f_1(x_{1i}, u_i, v_i) = \alpha_0(u_i, v_i) + \alpha_1(u_i, v_i)x_{1i} + \delta_1(u_i, v_i)(x_{1i} - K_1)_+,$$

and

$$f_2(x_{2i}, u_i, v_i) = \gamma_0(u_i, v_i) + \gamma_1(u_i, v_i)x_{2i} + \eta_1(u_i, v_i)(x_{2i} - K_2)_+,$$

where the spatially varying coefficients are defined as:

$$\begin{aligned}
\alpha_0(u, v) &= 1 + 0.1 \sin(\pi u), \\
\alpha_1(u, v) &= 0.5 + 0.1u, \\
\delta_1(u, v) &= 0.2 + 0.05v, \\
\gamma_0(u, v) &= 0.5 + 0.1 \sin(2\pi u), \\
\gamma_1(u, v) &= 0.4 + 0.1v,
\end{aligned}$$

and

$$\eta_1(u, v) = 0.1 + 0.05u.$$

Sinusoidal functions were incorporated into the intercept components α_0 and γ_0 to introduce spatially structured variation, while the slope coefficients α_1 and γ_1 were specified as smooth linear functions of the spatial coordinates. The two DGP components were thus designed to exhibit similarly smooth true slope surfaces, while the multiscale structure of the simulation arises from the component-specific bandwidth configuration rather than from the intrinsic complexity of the slope coefficients themselves.

Consequently, the DGP embeds two distinct spatial scales through its intercept components, with α_0 exhibiting smoother variation via $\sin(\pi u)$ and γ_0 exhibiting more rapid variation via $\sin(2\pi u)$. In this context, the multiscale aspect examined in this numerical illustration arises primarily from the differential smoothing induced by the component-specific bandwidths during estimation. Therefore, this numerical illustration aims to evaluate whether the proposed MS-GWSNR estimator can apply appropriately different levels of spatial smoothing across components assigned different bandwidths.

A first-order truncated spline basis was employed for each predictor variable using a single knot determined by the median of the predictor values. Parameter estimation was performed using the Weighted Least Squares (WLS) approach combined with a Gaussian kernel weighting scheme based on Euclidean spatial distance, given by [23]:

$$w_{ij}(u_i, v_i) = \exp\left\{-\frac{1}{2}\left(\frac{d_{ij}}{bw}\right)^2\right\},$$

where d_{ij} denotes the Euclidean distance between the i-th and j-th spatial locations, and bw represents the bandwidth parameter controlling the degree of spatial smoothing. Larger bandwidth values produce smoother global estimation, whereas smaller bandwidths allow the estimator to capture more localized spatial variation.

To accommodate the multiscale structure, different bandwidths were assigned to each model component, namely $bw_1 = 0.5$ for the first component and $bw_2 = 0.2$ for the second component. The larger bandwidth was intended to capture smoother spatial variation, whereas the smaller bandwidth was used to capture more localized spatial variation. These bandwidths were intentionally selected to reflect the smoothness characteristics embedded within each DGP component. In this controlled simulation setting, where the true spatial scales of each component are known by design, the bandwidths were specified directly in order to isolate and illustrate the multiscale smoothing mechanism rather than to confound it with the variability of a bandwidth-selection procedure. In practical applications where the true spatial scales are unknown, data-driven bandwidth selection procedures such as cross-validation or the Akaike Information Criterion can be employed, which represents a natural direction for future work.

The estimation procedure was conducted iteratively using the backfitting algorithm until convergence was achieved. Algorithm convergence was monitored using the SOC criterion based on Eq. (1), which was computed at each iteration from successive changes in the estimated functions.

The estimator performance was evaluated using several criteria, including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), convergence behavior of the backfitting algorithm, and the agreement between actual and fitted values. The RMSE and MAE are respectively defined as follows [24]:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{f}(x_i) - f_{\text{true}}(x_i)\right|,$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{f}(x_i) - f_{\text{true}}(x_i)\right)^2}.$$

where $\hat{f}(x_i)$ denotes the estimated function value obtained from the proposed MS-GWSNR model, while $f_{\text{true}}(x_i)$ represents the true function value generated from the DGP.

The MAE measures the average absolute deviation between the estimated and true functions, whereas the RMSE provides a quadratic measure of estimation error that is more sensitive to large deviations. Smaller values of MAE and RMSE indicate better estimation accuracy and improved reconstruction capability of the proposed model.

Furthermore, the reconstruction capability of the proposed model was assessed by comparing the estimated spatial coefficient surfaces with the true coefficient surfaces used in the data generating process.

The results of the numerical illustration are presented through several visualizations to demonstrate the behavior of the proposed estimator. The estimation results are summarized in Table 1, which presents the overall performance metrics of the proposed MS-GWSNR estimator obtained from the numerical illustration.

Table 1: Performance evaluation of the MS-GWSNR estimator

Evaluation Criterion	Value
MAE	0.0680
RMSE	0.0880
Iterations to Convergence	12
Final SOC	2.4812×10^{-5}

The relatively small MAE and RMSE values indicate that the proposed MS-GWSNR estimator recovers the underlying nonlinear multiscale spatial structure embedded in the simulated data.

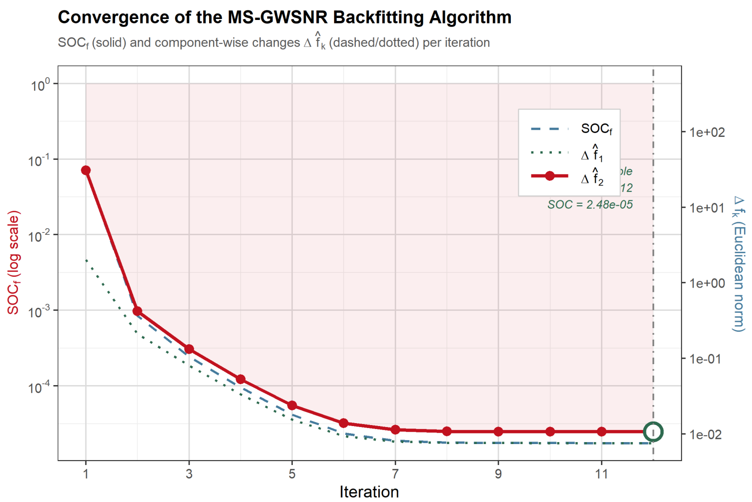


Fig. 1: Convergence of the backfitting algorithm

Fig. 1 presents the convergence behavior of the MS-GWSNR backfitting algorithm across iterations. The SOC criterion (dashed line, left axis) decreased monotonically from an initial value of approximately 7.06×10^{-2} at the first iteration to a final value of 2.48×10^{-5} at iteration 12, as also reported in Table 1, at which point the algorithm was declared stable. The component-wise changes $\Delta\hat{f}_1$ (dotted line) and $\Delta\hat{f}_2$ (solid red line with markers), displayed on the right axis, similarly exhibited monotonic decrease throughout the iterative process, indicating that both additive components converged simultaneously and consistently. The rapid reduction in SOC, spanning approximately three orders of magnitude within the first five iterations, followed by gradual stabilization confirms that the backfitting algorithm converges efficiently under the proposed multiscale estimation framework.

Fig. 2 illustrates the agreement between the actual values and the fitted values. The scatter plot demonstrates that the fitted values $\hat{y}_i$ are closely aligned with the actual values y_i, with the majority of observations lying in proximity to the perfect-fit diagonal line. The marginal density plots further confirm that the distributions of y_i and $\hat{y}_i$ are nearly identical in shape, indicating that the proposed estimator successfully reproduces the overall distributional characteristics of the response variable. As reported in Table 1, the estimation yielded an MAE of 0.0680 and an RMSE of 0.0880. Furthermore, a coefficient of determination of $R^2 = 0.8829$ was obtained, indicating that the MS-GWSNR model accounts for approximately 88.3% of the total variation in the response variable. These results collectively demonstrate that the proposed estimator reproduces the underlying response signal under the specified simulation setting.

Fig. 3(a) presents the estimated spatially varying coefficient surfaces $\hat{\alpha}_1(u,v)$ and $\hat{\gamma}_1(u,v)$, whereas Fig. 3(b) displays the corresponding true coefficient surfaces generated from the DGP. The estimated coefficient surfaces exhibit clear spatial variation across the domain, indicating that the proposed MS-GWSNR model successfully captures the presence of spatial heterogeneity within the simulated data. However, the estimated surfaces do not perfectly reproduce the exact

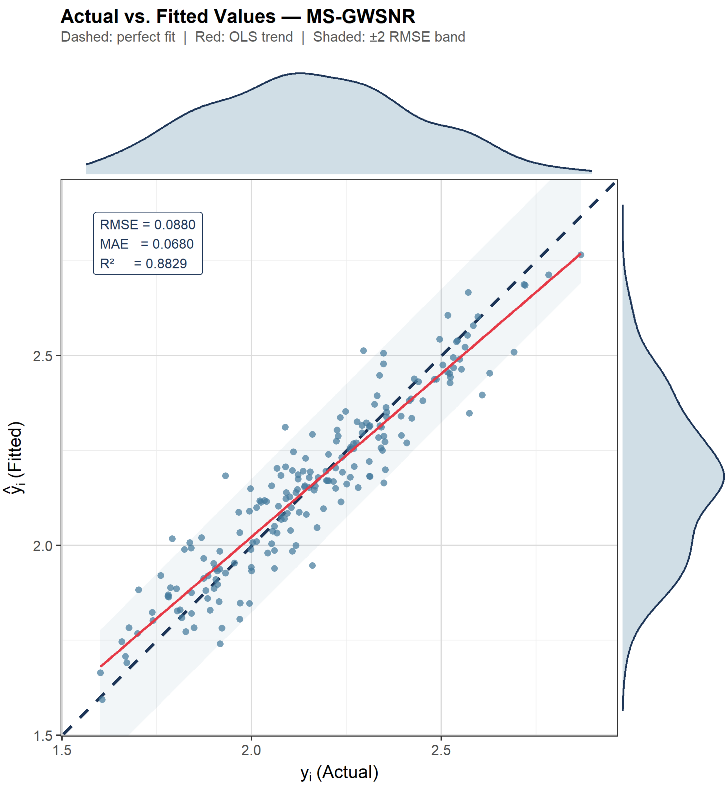
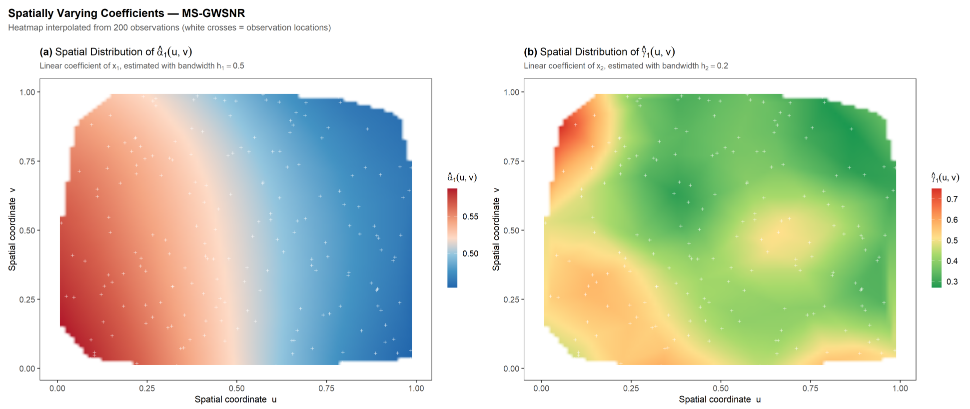


Fig. 2: The relationship between the actual values and the fitted values

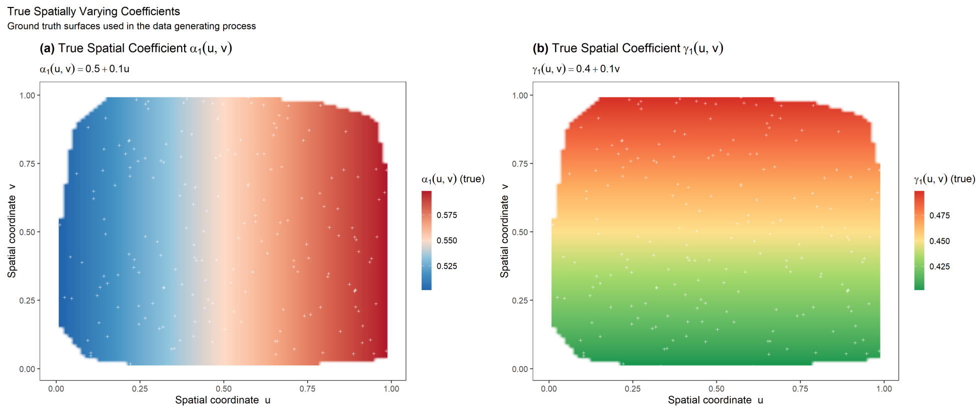
spatial patterns of the true coefficient surfaces. In particular, the estimated $\hat{\alpha}_1(u, v)$ surface exhibits a different directional gradient from the true $\alpha_1(u, v)$. The estimated surface $\hat{\gamma}_1(u, v)$, however, exhibits more localized spatial fluctuations relative to its true surface, which is defined as a smooth linear function $\gamma_1(u, v) = 0.4 + 0.1v$. These additional fluctuations do not reflect higher-frequency structure in the true DGP but rather the effect of the smaller bandwidth assigned to the second component ($bw_2 = 0.2$). Notably, since $\alpha_1(u, v)$ and $\gamma_1(u, v)$ are specified as equally smooth linear functions in the DGP, the markedly different smoothness observed between $\hat{\alpha}_1$ and $\hat{\gamma}_1$ can be attributed solely to the difference in their assigned bandwidths. This isolates and confirms the intended effect of the component-specific bandwidth mechanism. This behavior reflects the flexibility of the multiscale estimation framework, in which each component is estimated using different bandwidths and iteratively updated through the backfitting procedure.

Nevertheless, despite these differences at the component level, the estimator continues to recover the response surface closely. As shown in Fig. 2, the fitted values exhibit high agreement with the observed response values ($R^2 = 0.8829$), accompanied by relatively small RMSE and MAE values reported in Table 1. These results indicate that the proposed MS-GWSNR estimator is capable of effectively reconstructing the underlying multiscale spatial structure at the response level while simultaneously capturing spatial heterogeneity across different spatial scales. Furthermore, these results collectively demonstrate that the component-specific bandwidth mechanism functions as intended: a larger bandwidth produces globally smooth estimates, while a smaller bandwidth induces greater local adaptability, even when the underlying true surface is itself smooth.

The spatial distribution of estimation error is presented in Fig. 4, which shows the absolute deviation surfaces $|\hat{\alpha}_1 - \alpha_1|$ and $|\hat{\gamma}_1 - \gamma_1|$ across the spatial domain. For the first component (Fig. 4(a)), the estimation error exhibits a spatially structured pattern, with relatively smaller deviations observed in the central region and progressively larger deviations toward the boundary of the spatial domain, particularly in the high-u region. For the second component (Fig. 4(b)), the error is more broadly and heterogeneously distributed across the entire domain, with values



(a) Estimated spatial coefficient surfaces



(b) True spatial coefficient surfaces

Fig. 3: Spatial coefficient surfaces: (a) the estimated spatial coefficient surfaces and (b) the true spatial coefficient surfaces

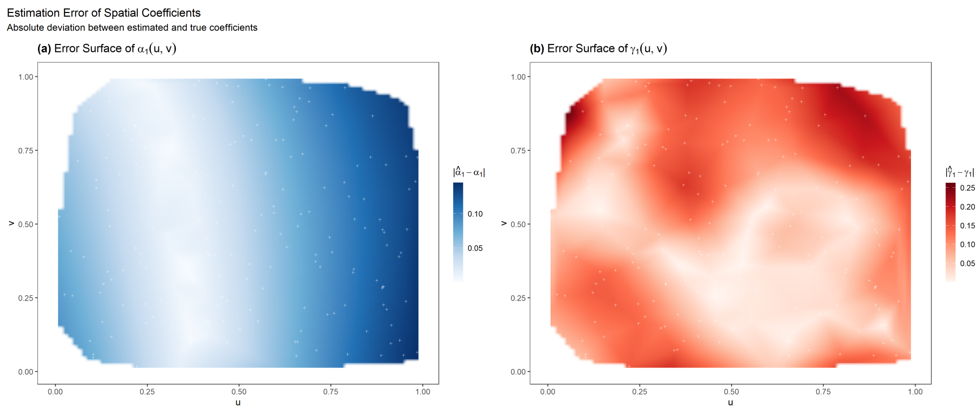


Fig. 4: The corresponding coefficient estimation error surfaces

predominantly ranging between 0.05 and 0.25. The comparatively larger and more spatially heterogeneous error observed for $\hat{\gamma}_1$ is consistent with the smaller bandwidth assigned to the second component ($bw_2 = 0.2$), which increases the estimation variance and renders the estimated surface more sensitive to local noise, even though the true $\gamma_1(u, v)$ surface is itself smooth.

3.5. Discussion

The numerical illustration presented in this study provides empirical support for the theoretical formulation and estimation framework of the proposed MS-GWSNR model. Overall, the results demonstrate that the proposed estimator, combined with the backfitting estimation algorithm, is capable of reconstructing nonlinear spatial relationships under heterogeneous multiscale spatial structures while maintaining stable computational performance throughout the estimation process. The findings further confirm that the use of component-specific bandwidths enables the estimator to adapt to different levels of spatial smoothness across model components.

As reported in Table 1 and illustrated in Fig. 1, the backfitting algorithm achieved stable convergence within 12 iterations with a final SOC value of 2.48×10^{-5}. The SOC criterion decreases monotonically throughout the iterative process, accompanied by simultaneous reductions in both $\Delta \hat{f}_1$ and $\Delta \hat{f}_2$. This behavior indicates that the proposed iterative estimation procedure produces stable updates for each additive component under the multiscale bandwidth configuration. Furthermore, the rapid decline of the SOC values during the early iterations suggests that the additive backfitting structure efficiently decomposes the response surface into its component functions without requiring computationally intensive joint optimization procedures. These findings provide empirical evidence that the proposed estimation algorithm is computationally feasible and numerically stable for estimating additive multiscale spatial models.

The numerical illustration also shows that the estimator recovers the simulated response surface closely. As presented in Table 1, the proposed estimator achieved relatively small error measures with MAE $= 0.0680$ and RMSE $= 0.0880$. In addition, Fig. 2 demonstrates strong agreement between the observed and fitted response values, with most observations closely distributed around the perfect-fit line. The obtained coefficient of determination ($R^2 = 0.8829$) further indicates that the proposed model successfully explains a substantial proportion of the variation in the simulated response variable. These findings collectively demonstrate that the proposed MS-GWSNR framework recovers the dominant nonlinear spatial signal embedded in the simulated data.

It should be noted that the remaining estimation error may arise from several sources, including the stochastic noise component $\varepsilon_i \sim N(0, 0.01)$ incorporated in the data generating process, as well as the use of fixed rather than data-driven optimal bandwidths. Under a fixed bandwidth setting, the estimator operates at a predetermined level of spatial smoothing that may not fully adapt to local variations in coefficient complexity. Nevertheless, the results demonstrate that even under this simplified setting, the proposed MS-GWSNR estimator reproduces the fitted response values closely relative to the true response surface.

The multiscale behavior of the proposed estimator is clearly illustrated in Fig. 3. The estimated coefficient surface $\hat{\alpha}_1(u, v)$, obtained using the larger bandwidth ($bw_1 = 0.5$), exhibits a smoother and more globally structured spatial pattern. In contrast, the estimated surface $\hat{\gamma}_1(u, v)$, estimated using the smaller bandwidth ($bw_2 = 0.2$), displays substantially more localized spatial variation and finer spatial fluctuations. These results empirically confirm that different bandwidth specifications produce different levels of spatial smoothing across model components, thereby enabling the proposed MS-GWSNR framework to adapt flexibly to heterogeneous spatial scales within the same model structure. It is worth noting that the true $\gamma_1(u, v)$ surface is itself a smooth linear function; the additional localized fluctuations observed in $\hat{\gamma}_1(u, v)$ therefore represent estimation variability induced by the narrow bandwidth rather than recovery of high-frequency structure in the DGP.

Although the estimated coefficient surfaces shown in Fig. 3 are not identical to the corresponding true coefficient surfaces from the DGP, the overall fitted response surface remains close to the true response, as evidenced by the goodness-of-fit results in Fig. 2 and Table 1. This indicates that the proposed estimator successfully captures the dominant multiscale spatial structure at the response level, even when localized discrepancies remain at the individual coefficient level. In the context of additive backfitting estimation, individual component functions may partially compensate for one another during the iterative fitting process, such that the total fitted response $\hat{y} = \hat{f}_1 + \hat{f}_2$ remains close to the true response surface despite discrepancies in the individual component estimates.

Additional insight into the estimator behavior is provided by the error surfaces shown in

Fig. 4. For the first component, the estimation error exhibits a relatively smooth spatial pattern with lower deviations across several central regions of the spatial domain. In contrast, the second component shows more heterogeneous and localized error patterns. This behavior is consistent with the smaller bandwidth assigned to the second component, which allows the estimator to respond more sensitively to local spatial variation. However, increased local adaptability is accompanied by greater spatial variability in the estimated coefficient surface. These findings illustrate an important trade-off within the multiscale estimation framework, namely that smaller bandwidths improve sensitivity to localized spatial structures but may simultaneously increase estimation variability in certain spatial regions.

Overall, the numerical illustration demonstrates that the proposed MS-GWSNR framework successfully integrates multiscale spatial weighting with spline-based nonparametric regression to accommodate heterogeneous spatial relationships with different levels of spatial smoothness. The results indicate that the proposed estimator is capable of producing stable estimation, close reconstruction of the response surface, and flexible adaptation to varying spatial scales within a unified modeling framework.

Nevertheless, the present study was conducted under a controlled simulation setting with predetermined bandwidth values, spline order, and knot locations. In practical applications, estimation performance may be further improved through the implementation of data-driven bandwidth selection procedures, adaptive knot selection strategies, and larger-scale simulation studies involving different sample sizes and spatial complexity levels. These extensions provide promising directions for future methodological development of the proposed MS-GWSNR framework.

4. Conclusion

This study proposed a Multiscale Geographically Weighted Spline Nonparametric Regression (MS-GWSNR) model by integrating the multiscale bandwidth framework of MGWR into the spline-based nonparametric structure of GWSNR. The proposed framework was developed to simultaneously accommodate spatial heterogeneity, varying spatial scales across predictors, and nonlinear relationships within a unified additive nonparametric regression model. To address the absence of a closed-form estimator arising from multiple spatial weighting matrices, parameter estimation was performed using a Weighted Least Squares (WLS) approach combined with a backfitting algorithm. The theoretical properties of the proposed estimator were derived through the smoothing matrix representation, including the smoothing-bias representation, variance representation, and Mean Squared Error (MSE) formulation. The numerical illustration demonstrated that the proposed estimation framework achieved stable convergence and recovered the underlying multiscale spatial structure. The results further showed that component-specific bandwidths enabled the estimator to adapt flexibly to different levels of spatial smoothness across model components. Overall, the proposed MS-GWSNR framework provides a flexible approach for modeling nonlinear spatial relationships under heterogeneous multiscale spatial structures. The integration of spline-based nonparametric regression with multiscale spatial weighting offers an important methodological contribution to the development of spatial nonparametric regression models. Nevertheless, the present study was conducted under controlled simulation settings with predetermined bandwidths and knot specifications. Future research may extend the proposed framework through adaptive bandwidth selection, optimal knot determination, larger-scale simulation studies, and applications to real spatial datasets.

CRediT Authorship Contribution Statement

Indi Rizqy Fahrani: Conceptualization, Methodology, Formal Analysis, Software, Writing–Original Draft Preparation. **Henny Pramoedyo:** Supervision, Validation, Writing–Review & Editing. **Atiek Iriany:** Supervision, Validation, Writing–Review & Editing.

Declaration of Generative AI and AI-assisted Technologies

Generative AI technology (ChatGPT, OpenAI) was used during the preparation of this manuscript for language refinement, grammar checking, code exploration, and LaTeX formatting assistance. The authors reviewed and edited all generated content and take full responsibility for the final content of the manuscript.

Declaration of Competing Interest

The authors declare no competing financial or personal interests.

Funding and Acknowledgments

This research received no external funding.

Data and Code Availability

The simulated data and code supporting the findings of this study are available from the corresponding author upon reasonable request.

References

[1] M. Naser Lessani and Zhenlong Li. "SGWR: similarity and geographically weighted regression". In: *International Journal of Geographical Information Science* 38.7 (2024), pp. 1232–1255. DOI: 10.1080/13658816.2024.2342319.

[2] Wei Kang and Taylor M. Oshan. "Scale and correlation in multiscale geographically weighted regression (MGWR)". In: *Journal of Geographical Systems* 27 (2025), pp. 399–424. DOI: 10.1007/s10109-025-00468-1.

[3] Ying Wu, Zhipeng Tang, and Shifeng Xiong. "A unified geographically weighted regression model". In: *Spatial Statistics* 55 (2023), p. 100753. DOI: 10.1016/j.spasta.2023.100753.

[4] Taylor M. Oshan, Zhenlong Li, Wei Kang, Levi J. Wolf, and A. Stewart Fotheringham. "MGWR: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale". In: *ISPRS International Journal of Geo-Information* 8.6 (2019), p. 269. DOI: 10.3390/ijgi8060269.

[5] A. Stewart Fotheringham, Wenbai Yang, and Wei Kang. "Multiscale geographically weighted regression (MGWR)". In: *Annals of the American Association of Geographers* 107.6 (2017), pp. 1247–1265. DOI: 10.1080/24694452.2017.1352480.

[6] Annisa Mutiara Priliyan and Ilham Faishal Mahdy. "Pemodelan Multiscale Geographically Weighted Regression (MGWR) pada Tingkat Pengangguran Terbuka di Provinsi Jawa Barat Tahun 2022". In: *Bandung Conference Series: Statistics* 4.2 (2024), pp. 329–339. DOI: 10.29313/bcss.v4i2.12708.

[7] Feng Chen, Yee Leung, Qiang Wang, and Yu Zhou. "Spatial non-stationarity test of regression relationships in the multiscale geographically weighted regression model". In: *Spatial Statistics* 62 (2024), p. 100846. DOI: 10.1016/j.spasta.2024.100846.

[8] Shuguang Deng, Shuyan Zhu, Xueying Chen, Jinlong Liang, and Rui Zheng. "A multi-scale geographically weighted regression approach to understanding community-built environment determinants of cardiovascular disease: Evidence from Nanning, China". In: *ISPRS International Journal of Geo-Information* 14.9 (2025), p. 362. DOI: 10.3390/ijgi14090362.

[9] Feili Wei, Dahai Liu, Ze Liang, Yueyao Wang, Jiashu Shen, Huan Wang, Yajuan Zhang, Yongxun Wang, and Shuangcheng Li. "Spatial heterogeneity and impact scales of driving factors of precipitation changes in the Beijing–Tianjin–Hebei region, China". In: *Frontiers in Environmental Science* 11 (2023). DOI: 10.3389/fenvs.2023.1161106.

[10] Yanmei Zhong, Song Hong, Zushuai Wei, Jeffrey P. Walker, Yanwen Wang, and Chaoqing Huang. "Spatial downscaling of SMAP soil moisture estimation using multiscale geographically weighted regression during SMAPVEX16". In: *Journal of Hydrology* 637 (2024), p. 131348. DOI: 10.1016/j.jhydrol.2024.131348.

[11] Zhihua Ma and Zhelin Huang. "A Bayesian implementation of the multiscale geographically weighted regression model with INLA". In: *Annals of the American Association of Geographers* 113.6 (2023), pp. 1501–1515. DOI: 10.1080/24694452.2023.2187756.

[12] Jinbiao Yan, Bo Wu, and He Zheng. "Multiscale cooperative optimization in multiscale geographically weighted regression models". In: *International Journal of Geographical Information Science* 39.2 (2025). Online first 2024, pp. 326–345. DOI: 10.1080/13658816.2024.2410346.

[13] Hanchen Yu. "Exploring Multiscale Spatial Interactions: Multiscale Geographically Weighted Negative Binomial Regression". In: *Annals of the American Association of Geographers* 114.3 (2024), pp. 574–590. DOI: 10.1080/24694452.2023.2289986.

[14] Sifriyani, Syaripuddin, M. Fathurahman, N. W. W. Sari, M. Fauziyah, A. T. R. Dani, R. Jannah, S. D. Juriani, and R. Kusuma. "Nonparametric spatio-temporal modeling: Construction of a geographically and temporally weighted spline regression". In: *MethodsX* 14 (2024), p. 103098. DOI: 10.1016/j.mex.2024.103098.

[15] Lilis Laome, I Nyoman Budiantara, and Vita Ratnasari. "Estimation Curve of Mixed Spline Truncated and Fourier Series Estimator for Geographically Weighted Nonparametric Regression". In: *Mathematics* 11.1 (2023), p. 152. DOI: 10.3390/math11010152.

[16] Robiansyah Putra, Muhammad Ghani Fadhlurrahman, and Gunardi. "Determination of the best knot and bandwidth in geographically weighted truncated spline nonparametric regression using generalized cross validation". In: *MethodsX* 10 (2023), p. 101994. DOI: 10.1016/j.mex.2022.101994.

[17] Sifriyani, Haryatmi, I. N. Budiantara, and Gunardi. "Geographically Weighted Regression with Spline Approach". In: *Far East Journal of Mathematical Sciences* 101.6 (2017), pp. 1183–1196. DOI: 10.17654/MS101061183.

[18] Sifriyani, I Nyoman Budiantara, Krishna Purnawan Candra, and Marisa Putri. "Selection of optimal knot point and best geographic weighted on geographically weighted spline nonparametric regression model". In: *MethodsX* 13 (2024), p. 102802. DOI: 10.1016/j.mex.2024.102802.

[19] Ziwei Huang, Shaoying Li, Yihuan Peng, and Feng Gao. "Spatial Non-Stationarity of Influencing Factors of China's County Economic Development Based on a Multiscale Geographically Weighted Regression Model". In: *ISPRS International Journal of Geo-Information* 12.3 (2023), p. 109. DOI: 10.3390/ijgi12030109.

[20] Laode Muhammad Saris, Henny Pramoedyo, and Adji Achmad Rinaldo Fernandes. "Modelling Geographically Weighted Truncated Spline Regression Using Maximum Likelihood Estimation for Human Development Disparities". In: *CAUCHY: Jurnal Matematika Murni dan Aplikasi* 10.1 (2025), pp. 109–120. DOI: 10.18860/cauchy.v10i1.31381.

[21] Nurul Mutiara Annisa, Dhita Hartanti Octavia, and Muhammad Ridzky Davala. "Deterministic Economic Resilience Through Gross Regional Domestic Product Using Nonparametric Geographically Weighted Regression Spline Truncated". In: *Jurnal Varian* 8.2 (2025), pp. 139–150. DOI: 10.30812/varian.v8i2.4303.

[22] Asad Haris, Noah Simon, and Ali Shojaie. “Generalized Sparse Additive Models”. In: *Journal of Machine Learning Research* 23.70 (2022), pp. 1–56. https://jmlr.org/papers/volume23/20-108/20-108.pdf.

[23] Rizky Ardhani, Nar Herrhyanto, and Fitriani Agustina. “Multiscale Geographically Weighted Regression (MGWR) untuk Memodelkan Nilai Angka Buta Huruf di Provinsi Sumatera Selatan Tahun 2021”. In: *Jurnal EurekaMatika* 11.2 (2023), pp. 89–98. DOI: 10.17509/jem.v11i2.63140.

[24] Ika Maulita, Chyntia Raras Ajeng Widiawati, and Arif Mu’amar Wahid. “Analisis Komparatif Linear Regression, Random Forest, dan Gradient Boosting untuk Prediksi Banjir”. In: *Jurnal Pendidikan dan Teknologi Indonesia* 4.8 (2024), pp. 369–379. DOI: 10.52436/1.jpti.599.