Selection of the Bandwidth Matrix in Spatial Varying Coefficient Models to Detect Anisotropic Regression Relationships

Xijian Hu 1, Yaori Lu 1, Huiguo Zhang 1, Haijun Jiang 1 and Qingdong Shi 2 *

1 College of Mathematics and System Sciences, Xinjiang University, Urumqi 830046, China; xijianhu@xjtu.edu.cn (X.H.); yaorilu@126.com (Y.L.); jianghujunxju@163.com (H.J.)
2 College of Resources and Environmental Sciences, Xinjiang University, Urumqi 830046, China
* Correspondence: zhang_huiguo@163.com (H.Z.); shiqingdong@126.com (Q.S.)

Abstract: The commonly used Geographically Weighted Regression (GWR) fitting method for a spatial varying coefficient model is to select a bandwidth h for the geographic location (u, v), and assign the same weight to the two dimensions. However, spatial data usually present anisotropy. The introduction of a two-dimensional bandwidth matrix not only gives weight from two dimensions separately, but also increases the direction of kernel smoothness. The adaptive bandwidth matrix is more flexible. Therefore, in this paper, a two dimensional bandwidth matrix is introduced into the spatial varying coefficient model for parameter estimation. Through simulation experiments, the results obtained under the adaptive bandwidth matrix are compared with those obtained under the global bandwidth matrix, indicating the effectiveness of introducing the adaptive bandwidth matrix.

Keywords: spatial varying coefficient model; bandwidth matrix; plug-in insertion method; local linear GWR estimation

1. Introduction

The spatial varying coefficient model proposed by Brunsdon et al. (1996) [1] is a widely used and significant model in non-parametric statistics. It is a non-parametric smoothing method with the distance function between observation points as the weight.

The work of Brunsdon et al. (1998) [2] and Fotheringham et al. (1998) [3] shows the role of the Geographically Weighted Regression (GWR) method in studying the nonstationarity of spatial data regression relationships, and the effectiveness of the method was demonstrated through actual data. Since the two-dimensional region is larger than the one-dimensional region, it is easy to produce boundary effects. Wang et al. (2008) [4] extended the local linear fitting method from the varying coefficient model to the spatial varying coefficient model, and proposed the local linear GWR method. Recently, the backfitting technique has been successfully applied to the GWR models for dealing with the multiscale issue by manipulating the bandwidth in the kernel function (Fotheringham et al. (2017) [5], Leong and Yue (2017) [6], Yu et al. (2020) [7]). The kernel smoothing method is an effective method for solving non-parametric problems, in which the selection of the bandwidth h is more important than the selection of the kernel function itself. For the selection of bandwidth, the cross-validation method (CV) and its variants are widely used in practice, such as the generalized cross-validation method (GCV), or the AICc method to select the optimal bandwidth. If the estimated curve is smooth, the global bandwidth can be selected; while if the estimated curve has a complex structure, the adaptive bandwidth can be considered. There are two important studies on adaptive bandwidth: one is the K-nearest neighbor method proposed by Loftsgaarden and Quesenberry (1965) [8]; the second is the adaptive kernel density estimation proposed by Breiman et al. (1977) [9]. Subsequently, ref. [10] proposed using the square root method to select the adaptive bandwidth, that is, for any dimension, the bandwidth
in the form $h_i \propto f(x_i)^{-1/2}$ can be selected. Fan and Gijbels (1995) [11] used a data-driven method to select the bandwidth; based on the estimated curvature, they proposed a new method for estimating the approximate deviation and variance of local polynomial fitting. Tian (2019) [12] used the cutting-edge knowledge of statistics to research the four major contemporary non-parametric statistics in the selection of bandwidth.

Many studies are limited to one-dimensional cases, while non-parametric problems will have more important applications in higher dimensions. The bandwidth matrix obtained under high-dimensional data not only affects the smoothness, but also affects the smoothing direction. For high-dimensional data, there are mainly the Plug-in insertion method, SCV and other methods for selecting the bandwidth matrix (Gramacki (2018) [13], Chacón and Duong (2018) [14]). Among them, the Plug-in insertion method is more stable for selecting bandwidth matrix $H$.

Wand and Jones (1994) [15] used the Plug-in method for multi-dimensional data to select a constrained bandwidth matrix. Duong and Hazelton (2003) [16] used the Plug-in method to select a binary unconstrained bandwidth matrix. Subsequently, Chacón and Duong (2010) [17] improved it by selecting unconstrained bandwidth matrix for high-dimensional data. In addition, the Bayesian method is also applied to the field of bandwidth selection. There are also many ways to select high-dimensional bandwidths by using the prior information from the data.

Brewer (2000) [18] proposed a new method for adaptive bandwidth in univariate kernel density estimation based on likelihood cross-validation and Bayesian graphical model analysis. Zhang et al. (2006) [19] took the bandwidth matrix elements as the parameters, and the posterior density was obtained by the Likelihood Cross-Validation criterion. The optimal window width was estimated by MCMC algorithm. Zougab et al. (2014) [20] used Bayesian estimation method to select adaptive bandwidth matrix, and so forth. For multi-dimensional data, besides bandwidth matrix, the product kernel is usually used to solve high-dimensional problems. Belaid et al. (2018) [21] applied the adaptive bandwidth selected by the Bayesian method to the high-dimensional discrete kernel. Gramacki and Gramacki (2017) [22] used the FFT method to solve some problems of the bandwidth matrix.

Sain et al. (1994) [23] used the CV method to select multiple bandwidths by the form of the product kernel. Wu et al. (2007) [24] regard the local bandwidth as the product of the local bandwidth factor and the global bandwidth, calculate the local window width factor with the clustering analysis algorithm, and obtain the multiple adaptive bandwidth. However, the kernel form of the multivariate product is essentially similar to the diagonal matrix. Xu (2009) [25] extended the bandwidth $h$ of the coefficient function in the varying coefficient model to multidimensional, and theoretically tried to give a diagonal bandwidth matrix, where the values of diagonal elements were consistent. Wand and Jones (1993) [26] believe that the unconstrained bandwidth matrix is more general than the diagonal bandwidth matrix, and is applicable to a wider range. Unconstrained matrices can include many more smooth directions.

In many practical fields, such as geography, ecology, and meteorology, the data obtained usually contain geographic location information, and this kind of data is called spatial data. Spatial effects mainly come from spatial dependence (that is, spatial correlation) and spatial heterogeneity (that is, spatial non-stationarity). The spatial data are usually not in the same quantity distribution in all directions of geographical position, and the data distribution is not uniform, showing anisotropy. At present, there is no good way to deal with this phenomenon. The local linear geographically weight regression method is usually adopted to solve the fitting problem of the spatial variable coefficient model. In this method, the distance function between observation points is usually used as the weight value, and the same window width $h$ is selected for the geographical position $(u, v)$, that is, the data of two dimensions are given the same weight as isotropy. Considering the spatial anisotropy and dependence of the two variables $(u, v)$, this paper introduces the two-dimensional Gaussian kernel function into the coefficient function estimation of the
spatial variable coefficient model, and uses the binary unconstrained bandwidth matrix \( H \) to select different bandwidths from the two dimensions of geographical location, \( u \) and \( v \). This can increase the direction of the smooth kernel, which can also explain the anisotropy. Furthermore, the adaptive bandwidth matrix is studied, and the coefficient estimates of the spatial variable coefficient models of the binary bandwidth matrix and the adaptive window width matrix are obtained respectively. Through simulation experiments, it is shown that the introduction of adaptive bandwidth matrix is more effective and flexible than the global bandwidth matrix.

2. Selection and Estimation of Bandwidth of Spatial Varying Coefficient Model
2.1. Spatial Varying Coefficient Model

The spatial varying coefficient model is to embed the data with geographical location information into the varying coefficient model, which can not only avoid the boundary effects exists in this method, while the boundary of the two-dimensional geographic region is larger than that of the one-dimensional region, and the data of the boundary region is more prone to distortion. The local polynomial fitting method can automatically correct the spatially varying effects in this method, while the boundary of the two-dimensional geographic region is larger than that of the one-dimensional region, and the data of the boundary region is more prone to distortion. The local polynomial fitting method can automatically correct the boundary effect (Mei and Wang (2012) [27]). Wang et al. (2008) [4] extended the local linear fitting method from the varying coefficient model to the spatial varying coefficient model for the first time and proposed a method using the distance function between observation points as the weight, namely the geographically weighted regression method, which essentially belongs to the Nadaraya–Watson kernel estimation method. The boundary effects exists in this method, while the boundary of the two-dimensional geographic region is larger than that of the one-dimensional region, and the data of the boundary region is more prone to distortion. The local polynomial fitting method can automatically correct the boundary effect (Mei and Wang (2012) [27]). Wang et al. (2008) [4] extended the local linear fitting method from the varying coefficient model to the spatial varying coefficient model, and presented the locally linear GWR method. In order to avoid the boundary effects, the local linear GWR method is used for model fitting in this paper.

For a given location \((u_0, v_0)\) in the studied area, the estimates \(\hat{\beta}_j(u_0, v_0)\) of \(\beta_j(u, v)\) \((j = 1, 2, \cdots, p)\) at \((u_0, v_0)\), and the matrix notation as follows:

\[
Y_i = \beta_1(u_i, v_i)X_{1i} + \cdots + \beta_p(u_i, v_i)X_{pi} + \epsilon_i, \ i = 1, 2, \cdots, n, \tag{1}
\]

where \((Y_i; X_{1i}, X_{2i}, \cdots, X_{pi})\) represents the observed value of the dependent variable and independent variable at the geographic location \((u_i, v_i)\), \((i = 1, 2, \cdots, n)\), \(\beta_j(u, v)(j = 1, 2, \cdots, p)\) is a function of geographic location, and \(\epsilon_i(i = 1, 2, \cdots, n)\) is the i.i.d. error term, which satisfies \(E(\epsilon_i) = 0, \ Var(\epsilon_i) = \sigma^2\). The model allows a spatially varying interception by setting \(x_{1i} = 1, i = 1, 2, \cdots, n\).

Brunsdo et al. [1] explicitly gave the specific form of the spatial varying coefficient model for the first time and proposed a method using the distance function between observation points as the weight, namely the geographically weighted regression method, which essentially belongs to the Nadaraya–Watson kernel estimation method. The boundary effects exists in this method, while the boundary of the two-dimensional geographic region is larger than that of the one-dimensional region, and the data of the boundary region is more prone to distortion. The local polynomial fitting method can automatically correct the boundary effect (Mei and Wang (2012) [27]). Wang et al. (2008) [4] extended the local linear fitting method from the varying coefficient model to the spatial varying coefficient model, and presented the locally linear GWR method. In order to avoid the boundary effects, the local linear GWR method is used for model fitting in this paper.

For a given location \((u_0, v_0)\) in the studied area, the estimates \(\hat{\beta}_j(u_0, v_0)\) of \(\beta_j(u, v)\) \((j = 1, 2, \cdots, p)\) at \((u_0, v_0)\), and the matrix notation as follows:

\[
\hat{\beta}(u_0, v_0) = (\hat{\beta}_1(u_0, v_0), \hat{\beta}_2(u_0, v_0), \cdots, \hat{\beta}_p(u_0, v_0))^T, \tag{2}
\]

\[
\hat{\beta}(u_0, v_0) = (I_p, 0_p, 0_p)(X^T(u_0, v_0)W(u_0, v_0)X(u_0, v_0))^{-1}X^T(u_0, v_0)W(u_0, v_0)Y,
\]

where \(I_p\) is the p-order identity matrix and \(0_p\) is the p-order zero matrix, and \(Y = (Y_1, Y_2, \cdots, Y_n)^T\) is the observations vector of the response variable. The design matrix at point \((u_0, v_0)\) is:

\[
X(u_0, v_0) = \begin{pmatrix}
X_{11} & \cdots & X_{1p} & X_{11}(u_1 - u_0) & \cdots & X_{1p}(u_1 - u_0) & X_{11}(v_1 - v_0) & \cdots & X_{1p}(v_1 - v_0) \\
X_{21} & \cdots & X_{2p} & X_{21}(u_1 - u_0) & \cdots & X_{2p}(u_1 - u_0) & X_{21}(v_1 - v_0) & \cdots & X_{2p}(v_1 - v_0) \\
\vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\
X_{n1} & \cdots & X_{np} & X_{n1}(u_1 - u_0) & \cdots & X_{np}(u_1 - u_0) & X_{n1}(v_1 - v_0) & \cdots & X_{np}(v_1 - v_0)
\end{pmatrix},
\]

and \(W(u_0, v_0) = \text{diag}(w_1(u_0, v_0), w_2(u_0, v_0), \cdots, w_n(u_0, v_0))\) is a diagonal weight matrix with \(w_i(u_0, v_0) = K_h(d_i)(i = 1, 2, \cdots, n), K_h(\cdot)\) is a kernel function. The kernel functions
used in this paper are all Gaussian kernel functions. For the traditional GWR method, the specific form of the one-dimensional Gauss kernel function is as follows:

\[ K_h(d_{ij}) = \frac{1}{h} K\left(\frac{d_{ij}}{h}\right) = \frac{1}{h} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{d_{ij}}{h}\right)^2\right), \quad j = 1, 2, \ldots, n \]  

(3)

where, \(d_{ij}\) represents the Euclidean distance between \((u_i, v_i)\) and \((u_j, v_j)\) \((i = 1, 2, \ldots, n)\). Specifically, the set of weights at \((u_0, v_0)\) in Equation (2) are \(w_i(u_0, v_0) = K\left(\frac{d_{ii}}{h}\right), \quad i = 1, 2, \ldots, n\).

The specific form of the two-dimensional Gauss kernel function is as follows:

\[ K_H(t) = |H|^{-\frac{1}{2}} K(H^{-\frac{1}{2}}t) = |H|^{-\frac{1}{2}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} (t - t_i)^T H^{-1} (t - t_i)\right). \]  

(4)

The set of weights at \((u_1, v_1)\) is \(\omega_i(u_1, v_1) = K(H^{-\frac{1}{2}}(t_i - t_1)), \quad i = 1, 2, \ldots, n\). The spatial varying coefficient model involves two variables, \(u\) and \(v\), so \(d = 2, t_i\) is a binary vector about \(u\) and \(v\). \(H\) is a binary bandwidth matrix about \((u, v)\), the distance used here is formally known as the Mahalanobis distance.

The estimate of the dependent variable \(Y\) is:

\[ \hat{Y} = L Y, \]  

(5)

where \(L\) is the \(p \times p\) hat matrix. The hat matrix used for parameter estimation using the one-dimensional kernel function has the following form:

\[ L_1 = \begin{pmatrix} 
{X^T_1, 0}_{(1 \times 2p)} & {X^T(u_1, v_1)W(u_1, v_1)X(u_1, v_1)^{-1}X^T(u_1, v_1)W(u_1, v_1)} \\
{X^T_2, 0}_{(1 \times 2p)} & {X^T(u_2, v_2)W(u_2, v_2)X(u_2, v_2)^{-1}X^T(u_2, v_2)W(u_2, v_2)} \\
& \vdots \\
{X^T_n, 0}_{(1 \times 2p)} & {X^T(u_n, v_n)W(u_n, v_n)X(u_n, v_n)^{-1}X^T(u_n, v_n)W(u_n, v_n)} 
\end{pmatrix} \]  

(6)

The hat matrix used in parameter estimation using two-dimensional kernel function has the following form:

\[ L_2 = \begin{pmatrix} 
{X^T_1, 0}_{(1 \times 2p)} & {X^T(u_1, v_1)W_H(u_1, v_1)X(u_1, v_1)^{-1}X^T(u_1, v_1)W(u_1, v_1)} \\
{X^T_2, 0}_{(1 \times 2p)} & {X^T(u_2, v_2)W_H(u_2, v_2)X(u_2, v_2)^{-1}X^T(u_2, v_2)W(u_2, v_2)} \\
& \vdots \\
{X^T_n, 0}_{(1 \times 2p)} & {X^T(u_n, v_n)W_H(u_n, v_n)X(u_n, v_n)^{-1}X^T(u_n, v_n)W(u_n, v_n)} 
\end{pmatrix} \]  

(7)

2.2. Selection of Bandwidth Matrix

(1) The bandwidth matrix of the unary variable is \(H = h^2\). Here, the square of the window width is used to consider the problem. In this paper, the generalized cross validation method (GCV) is used to select the one-dimensional bandwidth in traditional GWR method:

\[ GCV(h) = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{Y_i - \hat{Y}_i(h)}{1 - \frac{1}{n} tr(L(h))} \right)^2 = \frac{n}{(n - tr(L(h)))^2} \sum_{i=1}^{n} \frac{n}{n} \left( Y_i - \hat{Y}_i(h) \right)^2, \]  

(8)

where \(L\) is the hat matrix, \(tr(\cdot)\) represents the trace of the matrix, \(Y_i\) represents the value of the \(i\)th dependent variable, \(\hat{Y}(h) = L(h)Y\). To select the optimal window width \(h_0\), in practical application, a range is first given in accordance with a certain step size, the corresponding GCV(h) value is calculated, and the \(h_0\) that makes GCV(H) reach the minimum is chosen, that is, \(GCV(h_0) = \min_{h>0} GCV(h)\), then \(h_0\) is the selected optimal bandwidth.
The form of the binary unconstrained window width matrix is as follows: 

\[ \mathbf{H} = \begin{bmatrix} h_1^2 & h_1 h_2 \\ h_1 h_2 & h_2^2 \end{bmatrix} \]

where \( \mathbf{H} \) is a symmetric positive definite matrix, \( h_1, h_2 > 0 \), \( |h_1| < h_1 h_2 \). For the selection of the binary unconstrained bandwidth matrix \( \mathbf{H} \), the Plug-in method proposed by Duong and Hazelton (2003) [16] is used for reference in this paper. The Plug-in method mainly relies on replacing the unknown function by the estimator in approximation, so that the asymptotic expression AMISE (Asymptotic Mean Integrated Squared Error) of the estimated MISE (Mean Integrated Squared Error) can reach the minimum value. Here, \( \mathbf{H} \) can be regarded as a set of data about \( u \) and \( v \), and \( f(t) \) is the target density function. The specific expressions of MISE and AMISE are as follows:

\[
\text{MISE} \hat{f}(\cdot; \mathbf{H}) = \text{EISE} \hat{f}(\cdot; \mathbf{H}) = \mathbb{E} \int_{\mathbb{R}^2} [\hat{f}(x, \mathbf{H}) - f(x)]^2 \, dx
\]

\[
\text{AMISE} \hat{f}(\cdot; \mathbf{H}) = n^{-1} |\mathbf{H}|^{-1/2} R(K) + \frac{1}{4} \mu_2(K)^2 \left( \text{vech}^T \mathbf{H} \right) \Psi_4 \left( \text{vech} \mathbf{H} \right)
\]

\[
\mathbf{H}_{\text{AMISE}} = \arg \min_{\mathbf{H} \in \mathbb{H}} \text{AMISE} \hat{f}(\cdot; \mathbf{H}),
\]

where \( R(K) = \int_{\mathbb{R}^2} K(x)^2 \, dx \), \( \mu_2(K) \mathbf{1} = \int_{\mathbb{R}^2} xx^T K(x) \, dx \), \( \mu_2(K) < \infty \), \( \text{vech} \) represents a special form of column vector, for example, \( \text{vech} \left( \begin{array}{ccc} a & c & d \\ c & d & 0 \end{array} \right) = \left( \begin{array}{ccc} a \\ c \\ d \end{array} \right) \).

Then \( \Psi_4 \) is a 3 \times 3 matrix, the specific form is as follows:

\[
\Psi_4 = \int_{\mathbb{R}^2} \text{vech}[2D^2 f(x) - dg[D^2 f(x)]] \text{vech}^T [2D^2 f(x) - dg[D^2 f(x)]] \, dx.
\]

\( D^2 f(x) \) is the Hessian matrix of \( f \), and \( dg[A] \) denotes the diagonal matrix formed by replacing all off-diagonal entries of \( A \) by zeroes. \( \Psi_4 \) are integrals of products of density derivatives of \( f(\cdot) \).

To obtain the optimal bandwidth matrix \( \mathbf{H} \), the key is to estimate \( \Psi_4 \). The elements in \( \Psi_4 \) are \( \psi_{ij} = \int_{\mathbb{R}^2} f^{(r)}(x) f^{(s)}(x) \, dx \), \( r = (r_1, r_2) \), \( |r| = r_1 + r_2 \), \( r_1, r_2 \) are non-negative integers, \( f^{(r)}(x) \) is the \( r \)-order partial derivative with respect to \( (l_1, l_2) \). Firstly, the original data are standardized, and the test bandwidth matrix \( \mathbf{G} \) is used to replace \( \mathbf{H} \) to estimate each element \( \Psi_r \) in \( \Psi_4 \), separately:

\[
\Psi_r(\mathbf{G}) = n^{-1} \sum_{i=1}^{n} f^{(r)}(T_i; \mathbf{G}) = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} K_r^{(r)}(T_i - T_j).
\]

Calculate the variance and deviation of related \( \mathbf{G} \). Let \( \mathbf{G} = g^2 \mathbf{I} \). Since \( g \) is also unknown, calculate the variance and deviation of \( g \) to obtain the AMSE related to \( g \):

\[
\text{AMSE} \hat{g}(g) = 2n^{-2} s^{-2} \sum_{i=1}^{n} \psi_{0} R(K^{(r)}) + \left[ n^{-1} s^{-2} \sum_{i=1}^{n} \psi_{0} R(K^{(r)}) \right] (0) + \frac{1}{2} s^2 \mu_2(K) \sum_{i=1}^{n} \psi_{0} + 2e_i^2.
\]

\( e_i \) represents the unit vector whose \( i \)-th value is 1, and sum over all \( \text{AMSE} \hat{g}(g) \) to get \( \text{SAME} \Psi_4 \):

\[
\text{SAME} \Psi_4 = \sum_{r \geq 4} \text{AMSE} \hat{g}(g).
\]

To take the \( g \) which minimizes \( \text{SAME} \Psi_4 \). Here, the final bandwidth matrix \( \mathbf{H} \) is generally obtained by the two-stage method. The selection of the optimal bandwidth matrix can be implemented in the “ks” package of the R language.

As for the selection of the adaptive bandwidth matrix, the Bayesian method proposed by Zougab et al. (2014) [20] was adopted in this paper. Under the Bayesian framework, the prior distribution is selected as the conjugate prior distribution, and the explicit expression of the adaptive window width matrix is obtained by combining with
the square loss function. Here, \( t \) can be viewed as a set of data about \( u \) and \( v \). The specific calculation steps are as follows: the kernel estimation is carried out by leave-one-out method estimation.

\[
\hat{f}_{-j}(t_i | \{ t_{-j} \}, H_i) = \frac{1}{n - 1} \sum_{j \neq i}^{n} K_{H_i}(t_i - t_j).
\]

The kernel function generally uses Gaussian kernel:

\[
\hat{f}(t) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{(2\pi)^{d/2} |H_i|^{1/2}} \exp \left\{ -\frac{1}{2} (t - t_i)^T H_i^{-1} (t - t_i) \right\}.
\]

At this time, its prior distribution can take its conjugate prior distribution, that is, the inverse Wishard prior distribution,

\[
\pi(H_i) = \frac{|Q|^{r/2} |H_i|^{-(r+d+1)/2}}{2^{d/2} \Gamma_d(r/2)} \exp \left\{ -\frac{1}{2} \text{tr}(QH_i^{-1}) \right\},
\]

where \( \Gamma_d(q) = 2^{dq} \pi^{d(d-1)/4} \prod_{k=0}^{d-1} \Gamma(q - \frac{k}{2}) \) is the classic Gamma function.

Under the square loss function, the explicit expression of a posteriori estimate can be given, which can avoid the consumption of a lot of operation time during sampling calculation. For each bandwidth matrix \( H_i \), the form of a posteriori probability is as follows:

\[
\pi(H_i | t_i) = \frac{\hat{f}(t_i | \{ t_{-j} \}, H_i) \pi(H_i)}{\int_H \hat{f}(t_i | \{ t_{-j} \}, H) \pi(H) dH}.
\]

The posterior distribution \( \pi(H_i | x_i) \) can be expressed as follows:

\[
\pi(H_i | t_i) = \frac{\sum_{j=1}^{n} |H_i|^{-(r+d+2)/2} \exp \left\{ -\frac{1}{2} \text{tr} \left\{ (t_i - t_j)^T (t_i - t_j) + Q \right\} H_i^{-1} \right\} \times m}{2^{(r+1)d/2} \Gamma_d((r + 1)/2) \sum_{j=1}^{n} |(t_i - t_j)^T (t_i - t_j) + Q|^{-(r+1)/2}}.
\]

For each \( j \) term in the posterior distribution, where \( j \neq i \) is multiplied by something with unit 1, where the quantity with unit 1 can be expressed as:

\[
m = \frac{|(t_i - t_j)^T (t_i - t_j) + Q|^{-(r+1)/2}}{|(t_i - t_j)^T (t_i - t_j) + Q|^{-(r+1)/2}}.
\]

Then, we can get:

\[
\pi(H_i | t_i) = \frac{\sum_{j=1}^{n} |H_i|^{-(r+d+2)/2} \exp \left\{ -\frac{1}{2} \text{tr} \left\{ (t_i - t_j)^T (t_i - t_j) + Q \right\} H_i^{-1} \right\} \times m}{2^{(r+1)d/2} \Gamma_d((r + 1)/2) \sum_{j=1}^{n} |(t_i - t_j)^T (t_i - t_j) + Q|^{-(r+1)/2}} = \frac{n}{\sum_{j=1}^{n} |(t_i - t_j)^T (t_i - t_j) + Q|^{-(r+1)/2}} \times \omega(r + 1, (t_i - t_j)^T (t_i - t_j) + Q)
\]

\[
H_i = \frac{1}{r - d} \sum_{j=1}^{n} \frac{|(t_i - t_j)^T (t_i - t_j) + Q|^{-(r+1)/2}}{\sum_{j=1}^{n} |(t_i - t_j)^T (t_i - t_j) + Q|^{-(r+1)/2}} \times \left\{ (t_i - t_j)^T (t_i - t_j) + Q \right\},
\]
where \( r \) is the degree of freedom, given \( r = n^{2/(d+4)} + d \), \( Q \) is the covariance matrix of the sample. For the adaptive bandwidth matrix, the specific form of the two-dimensional Gauss kernel function is as follows:

\[
K_{H_i}(t) = |H_i|^{-\frac{1}{2}} K_{H_i}^{-\frac{1}{2}} = |H_i|^{-\frac{1}{2}} \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{1}{2} (t - t_i)^T H_i^{-1} (t - t_i) \right).
\]

The set of weights at \((u_1, v_1)\) is \( w_i(u_1, v_1) = K(H_i^{-\frac{1}{2}} (t_i - t_1)), i = 1, 2, \cdots, n \), here \( d = 2, t_i \) is a binary vector about \((u, v)\), \( H_i \) is a binary bandwidth matrix about \((u, v)\). The corresponding hat matrix \( L \) has the following forms:

\[
L_i = \begin{pmatrix}
(X^T_1, 0_{(1 \times 2p)})(X^T(u_1, v_1)X(u_1, v_1))^{-1}X^T(u_1, v_1)W(u_1, v_1) \\
(X^T_2, 0_{(1 \times 2p)})(X^T(u_2, v_2)X(u_2, v_2))^{-1}X^T(u_2, v_2)W(u_2, v_2) \\
\vdots \\
(X^T_n, 0_{(1 \times 2p)})(X^T(u_n, v_n)X(u_n, v_n))^{-1}X^T(u_n, v_n)W(u_n, v_n).
\end{pmatrix}
\]

3. Simulation Experiment

In order to reflect the spatial scale difference and interdependence between \((u, v)\), both \( u_1 \) and \( v_1 \) are taken as n sample points on uniformly distributed of \( U(10, 20) \), and set the matrix \( A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \) and \([u, v] = [u_1, v_1] * A\). By rotating \((u_1, v_1)\), a set of geographic locations \((u, v)\) with relevant relationships is obtained. The rotated \((u, v)\) was used in the experiment to reflect the spatial scale difference and the dependent relationship between the geographical positions of the spatial data.

The model used to generate spatial data is as follows:

\[
Y_i = \beta_1(u_i, v_i)X_{1i} + \beta_2(u_i, v_i)X_{2i} + \epsilon_i, i = 1, 2, \cdots, n,
\]

where the independent variable \( X_1 \) is a column vector of all 1, that is, the observation values \( X_{11}, X_{21}, \cdots, X_{n1} \) are all 1, and the observation values \( X_{12}, X_{22}, \cdots, X_{n2} \) are independent values generated from the uniform distribution \( U(0, 2) \), the error \( \epsilon_1, \epsilon_2, \cdots, \epsilon_n \) are independent and identically distributed random numbers that obey the normal distribution \( N(0, 2^2) \), where the sample size is taken as \( n = 100 \), let \( \beta_1(u, v) = \frac{1}{2}(u + v), \beta_2(u, v) = \frac{1}{2}u \), and the real surfaces are shown in Figure 1.

![Coefficient function \( \beta_1(u, v) \) (left) and \( \beta_2(u, v) \) (right).](image)

Figure 1. The real surface of the coefficient function \( \beta_1(u, v) \) (left) and \( \beta_2(u, v) \) (right).

The coefficient functions in Equation (19) are calibrated by the Formula (2), and the optimal bandwidth in the kernel function is selected in three ways: the GCV-based method in Equation (8), the Plug-in-based matrix \( H \) method in Equation (11) and the adaptive bandwidth matrix \( H \) method in Equation (16). We get the GCV-based op-
timal bandwidth \( h_0 = 35.09 \), and the Plug-in-based optimal bandwidth matrix \( \mathbf{H} \) is
\[
\mathbf{H}_0 = \begin{bmatrix} 7.9068 & 17.1485 \\ 17.1485 & 38.2639 \end{bmatrix}
\]
The accuracy of parameter estimation is evaluated by the mean square error (MSE). Its definition is \( \text{MSE}(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} (\theta - \hat{\theta})^2 \); the smaller the value of the MSE, the better the fitting effect. In order to compare the performance of the above three estimation methods, keep other quantities fixed, re-sample data with a sample size of \( n \) from \( N(0, 2^2) \), do \( N = 200 \) repeated experiments, calculate the MSE of \( Y, \beta_1, \beta_2 \), and take the average of the N replications. The numerical results are shown in the Table 1.

Additionally, Figures 2 and 3 display the estimated coefficients and the estimation residuals of \( \hat{\beta}_1(u, v) \) and \( \hat{\beta}_2(u, v) \) with the GCV-based bandwidth, the bandwidth matrix and the adaptive bandwidth matrix method, respectively.

**Table 1.** MSE results of the GCV-based bandwidth \( h \), the bandwidth matrix \( \mathbf{H} \) and the adaptive bandwidth matrix for \( N = 200 \) replications.

| Bandwidth Type | GCV-Based | Matrix H-Based | Adaptive Matrix H-Based |
|----------------|-----------|----------------|------------------------|
| MSE(\( Y \))  | 3.0298    | 2.3928         | 0.6633                 |
| MSE(\( \beta_1 \)) | 0.0048    | 0.0136         | 0.0011                 |
| MSE(\( \beta_2 \)) | 0.0105    | 0.0137         | 0.0012                 |
| Running time   | 14.24 s   | 2.38 s         | 5.67 s                 |

It can be found from the simulation results in Table 1 that the MSE of regression coefficients estimation obtained by using the bandwidth matrix \( \mathbf{H} \) and the adaptive bandwidth matrix are smaller than that obtained by using GCV-based bandwidth \( h \) for 200 replications. The parameter estimation MSE of \( \beta_1(u, v) \) and \( \beta_2(u, v) \) obtained by the adaptive bandwidth matrix is less than that of the parameter estimation of \( \beta_1(u, v) \) and \( \beta_2(u, v) \) obtained by the direct use of the bandwidth matrix. In other words, the estimation result obtained by the adaptive bandwidth matrix is better than that obtained by the use of the bandwidth matrix. That means the performance of two dimensions’ unconstrained bandwidth matrix for geographical location \((u, v)\) is better than that of selecting one bandwidth for two variables \((u, v)\) in the spatially varying coefficient model. In addition, when the bandwidth matrix \( \mathbf{H} \) obtained by the Plug-in method is used for parameter estimation, the running time of the program is shorter than that obtained by the GCV-based method.

From the residual graph of the coefficient function in Figures 2 and 3, the estimated residuals are dispersed by selecting the same bandwidth \( h \) for \( u \) and \( v \), and the estimated residuals are relatively concentrated around zero by selecting the bandwidth matrix \( \mathbf{H} \), although there are a few points with large differences. The residuals boxplots of \( \hat{\beta}_1(u, v) \) in Figure 2 and \( \hat{\beta}_2(u, v) \) in Figure 3 show that the estimated residuals of \( \hat{\beta}(u, v) \) are more concentrated by using the adaptive bandwidth matrix than the global bandwidth matrix. The reason for the occurrence of large bias points may be related to the non-diagonal elements of the bandwidth matrix \( \mathbf{H} \), or an adaptive bandwidth matrix can be introduced. The introduction of the adaptive bandwidth matrix can reduce the outliers. It further proves the effectiveness of introducing a two-dimensional adaptive bandwidth matrix into the spatial varying coefficient model.
Figure 2. The estimated surface of $\beta_1(u,v)$ (top row) and residuals boxplots (bottom row) with the GCV-based method (left panel), the bandwidth matrix $H$ (middle panel) and the adaptive bandwidth matrix $H$ method (right panel).

The above results show that, in the spatial varying coefficient model, considering the spatial anisotropy and correlation of spatial coordinates $(u,v)$, the estimated coefficient surfaces obtained by selecting bandwidth matrix $H$ is more accurate than that obtained by selecting the same bandwidth $h$ for the two spatial coordinates $(u,v)$. In addition, compared with GCV-based method, using the Plug-in-based method to select the bandwidth matrix can reduce the number of iterations in model fitting and improve computational efficiency.
4. Empirical Research

The Fractional Vegetation Cover (FVC) quantifies the denseness of vegetation and reflects the growth trend of vegetation. FVC plays an important role in reflecting the process of climate change and is an important indicator of ecological environmental change and is a useful parameter for many environmental and climate-related applications. In this research, the datasets of variables FVC, temperature and precipitation in Yili, Xinjiang were collected to study the spatial varying relationships between FVC and climate factors.

4.1. Data Sources

Yili Kazakh Autonomous Prefecture was the selected research area, located in northwest China’s Xinjiang Uygur Autonomous Region, surrounded by mountains, and vulnerable to influence of airflow around the Tianshan mountains, the Wusun mountain, Kequin and the Nalati mountain. Inside the Yili river near the Yili river valley, there is a humid climate and abundant rainfall; it is arguably the wettest region in Xinjiang. In this paper,
the data on air temperature, precipitation and FVC in Yili, Xinjiang in 2011 adopted by Feng et al. (2016) [28] were used, with a total of 835 corresponding data points.

4.2. Modelling

The specific form of the spatial varying coefficient model in this section is as follows:

\[ Y_i = \beta_1(u_i, v_i)X_{i1} + \beta_2(u_i, v_i)X_{i2} + \beta_3(u_i, v_i)X_{i3} + \epsilon_i, \quad (20) \]

where \( Y_i \) represents the observation value of the ith vegetation coverage index, \( X_{i1} \) is always equal to 1, \( X_{i2} \) represents the observation value of the ith annual average precipitation, and \( X_{i3} \) represents the observation value of the ith annual average temperature, \( \beta_1(u_i, v_i) \) represents the benchmark vegetation coverage index, \( \beta_2(u_i, v_i) \) represents the influence intensity of the annual average precipitation at the ith point on the vegetation coverage index, \( \beta_3(u_i, v_i) \) represents the intensity of influence of the annual average temperature at the ith point on the vegetation coverage index, \( \epsilon_i \) represents the residual. For the estimation of the coefficient function, the local linear GWR method is used for model fitting and parameter estimation.

4.3. Results and Analysis

4.3.1. Results of Unary Bandwidth and Bandwidth Matrix

According to the method of selecting the bandwidth matrix introduced in Section 2.2, the bandwidth \( h = 12,938.16 \) and the bandwidth matrix \( H = \begin{pmatrix} 614198610 & 2083252 \\ 2083252 & 253867658 \end{pmatrix} \), according to formulas in Equations (3) and (4), a one-dimensional kernel function and a two-dimensional kernel function, are constructed to fit the model respectively and to obtain the MSE of the estimated value of the dependent variable and the running time of the program. The results in Table 2 show that the MSE(\( Y \)) obtained by the two-dimensional kernel function with the bandwidth matrix is smaller, the estimation is more accurate, the program running time is greatly shortened, and the running efficiency is greatly improved.

Table 2. Selected bandwidth and results of MSE.

| Bandwidth Type | Bandwidth \( h \) | Bandwidth Matrix \( H \) |
|---------------|-------------------|-----------------------|
| MSE(\( Y \))  | 0.074             | 0.039                 |
| Running time  | 33.69 min         | 1.32 min              |

The estimated values obtained from the bandwidth matrix \( H \) in Table 3 show that: the coefficient function \( \beta_2(u, v) \) and the coefficient function \( \beta_3(u, v) \) of annual mean temperature and precipitation both show obvious spatial and regional differences. There was a positive correlation between precipitation and FVC, and the regions with more precipitation had higher vegetation coverage. There was also a positive correlation between temperature and FVC.

Table 3. Regression coefficients estimation with bandwidth matrix \( H \).

| Regression Coefficients | Minimum     | 1/4 Quantile | Median | 3/4 Quantile | Maximum |
|-------------------------|-------------|--------------|--------|--------------|---------|
| \( \beta_1(u, v) \)  | -10.5274    | -0.0085      | 0.8041 | 1.3285       | 12.5262 |
| \( \beta_2(u, v) \)  | -0.0349     | -0.0017      | 0.0003 | 0.0021       | 0.0358  |
| \( \beta_3(u, v) \)  | -0.0788     | 0.0013       | 0.0155 | 0.0566       | 0.3834  |

The results in Figure 4 show that, in the spatial varying coefficient model, the estimated value obtained by selecting the binary bandwidth matrix \( H \) for the kernel function is roughly consistent with the real vegetation coverage index, and the residual is small. The vegetation coverage index in most areas of Yili is relatively high, while the vegetation coverage index in the border area and the northwest is slightly lower, which is consistent
with the actual situation. The reasons for this phenomenon may be related to temperature, precipitation, terrain, and so forth. It should be noted that, in addition to intuitively exploring spatial varying regression relationships by summary tables or images of the estimated coefficient functions, the statistical inference methods can be used to test the significance of spatial non-stationary relationships, including the residual sum of square based statistics with an F-distribution (Leung et al., 2000 [29]; Mei et al., 2006 [30]) and the bootstrap methods (Wei and Qi 2012 [31]; Mei et al., 2016 [32]).

According to the adaptive bandwidth matrix \( H \) method discussed in Section 2, Equation (16), the estimated result of the coefficient function can be found in Table 4, and Figure 5 shows the prediction of the FVC and the residuals with the adaptive bandwidth matrix.

### Table 4. Regression coefficients estimation with the adaptive bandwidth matrix \( H \).

| Regression Coefficients | Minimum | 1/4 Quantile | Median | 3/4 Quantile | Maximum |
|-------------------------|---------|--------------|--------|--------------|---------|
| \( \beta_1(u, v) \)     | -5.3196 | 0.3858       | 0.8049 | 1.1690       | 6.6727  |
| \( \beta_2(u, v) \)     | -0.0207 | -0.0011      | 0.0002 | 0.0013       | 0.0196  |
| \( \beta_3(u, v) \)     | -0.0288 | -0.0008      | 0.0141 | 0.0520       | 0.2599  |

By observing the estimated values of the coefficient functions in Table 4, it is easy to see that the values of \( \hat{\beta}(u, v) \) are mostly positive numbers. For \( \beta_2(u, v) \) and \( \beta_3(u, v) \), most of the values are positive, that is, precipitation and temperature have a positive effect on the FVC. By comparing Tables 3 and 4 we can find that, compared to the range of parameter estimates obtained with a fixed bandwidth matrix, the range of parameter estimates obtained by the adaptive bandwidth matrix is smaller.
5. Summary and Prospect

(1) When there is a scale problem and a correlation between two variables, that is, the smoothness between the two variables is different and there is anisotropy in space, it is more reasonable to consider the two variables in selecting the binary bandwidth matrix than to select the same bandwidth fitting value between the two variables. At the same time, as an intermediary between a one-dimensional univariate and a high-dimensional multivariate, a bivariate has a smoother direction than a univariate kernel, and is easy to visualize. This also makes it possible to generalize the spatio-temporal variable coefficient model to the unconstrained bandwidth matrix of the spatio-temporal variable coefficient model or an even higher dimensional bandwidth matrix.

(2) Compared with the GCV method, using the Plug-in method to select the bandwidth matrix can reduce the number of iterations in the process of selecting the optimal bandwidth \( h \), and can shorten the iteration time of program operation and improve the efficiency of calculation.

(3) By observing the residuals of the parameter estimates, we find that the residual range of the parameter estimates obtained under the bandwidth is small, while most of the residual values of the parameter estimates obtained under the bandwidth matrix are very concentrated. It will have points with larger residual values, which may be related to non-diagonal elements. We can also try to add an adaptive bandwidth matrix, which is worth continuing to study.

Author Contributions: Conceptualization, X.H. and H.Z.; methodology, Y.L.; software, Y.L.; validation, X.H. and H.Z.; formal analysis, H.Z. and Y.L.; investigation, Y.L. and X.H.; resources, Q.S.; data curation, Q.S.; writing—original draft preparation, Y.L.; writing—review and editing, H.Z.; visualization, Y.L. and H.Z.; supervision, X.H. and H.Z.; project administration, H.J.; funding acquisition, X.H. All authors have read and agreed to the published version of the manuscript.

Funding: This study is supported by the National Natural Science Foundation of China grand number 11961065 and U1703237). Zhang’s work is also partially supported by the Natural Science Foundation of Xinjiang grant number 2019D01C045 and The Ministry of education of Humanities and Social Science project grant number 19YJA910007.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data in the paper can be requested by email to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.
References

1. Brunsdon, C.; Fotheringham, A.S.; Charlton, M.E. Geographically weighted regression: A method for exploring spatial non-stationarity. *Geogr. Anal.* 1996, 28, 281–298. [CrossRef]
2. Brunsdon, C.; Fotheringham, S.; Charlton, M.E. Geographically weighted regression—Modelling spatial non-stationarity. *J. R. Stat. Soc. Ser. D Stat.* 1998, 47, 431–443. [CrossRef]
3. Fotheringham, A.S.; Charlton, M.E.; Brunsdon, C. Geographically Weighted Regression: A Natural Evolution of The Expansion Method for Spatial Data Analysis. *Environ. Plan. A* 1998, 30, 1905–1927. [CrossRef]
4. Wang, N.; Mei, C.L.; Yan, X.D. Local linear estimation of spatial varying coefficient models: An improvement on the geographically weighted regression technique. *Environ. Plan. A* 2008, 40, 986–1005. [CrossRef]
5. Fotheringham, A.S.; Yang, W.; Kang, W. Multiscale Geographically Weighted Regression (MGWR). *Ann. Am. Assoc. Geogr.* 2017, 107, 1247–1265. [CrossRef]
6. Leong, Y.Y.; Yue, J.C. A Modification to Geographically Weighted Regression. *Int. J. Health Geogr.* 2017, 16, 11. [CrossRef] [PubMed]
7. Yu, H.; Fotheringham, A.S.; Li, Z.; Oshan, T.; Kang, W.; Wolf, L.J. Inference in multiscale geographically weighted regression. *Geogr. Anal.* 2020, 52, 87–106. [CrossRef]
8. Loftsgaarden, D.O.; Quesenberry, C.P. A Nonparametric Estimate of a Multivariate Density Function. *Ann. Math. Stat.* 1965, 36, 1049–1051. [CrossRef]
9. Breiman, L.; Meisel, W.; Purcell, E. Variable kernel estimates of multivariate densities. *Technometrics* 1977, 19, 135–144. [CrossRef]
10. Abramson, I.S. On bandwidth variation in kernel estimates—a square root law. *Ann. Stat.* 1982, 10, 1217–1223. [CrossRef]
11. Fan, J.; Gijbels, I. Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *J. R. Stat. Soc. Ser. B Methodol.* 1995, 57, 371–394. [CrossRef]
12. Tian, M.Z. Selection and Application of Bandwidth in Modern Nonparametric Statistics; Science Press: Beijing, China, 2019; pp. 206–260.
13. Gramacki, A. Nonparametric Kernel Density Estimation and Its Computational Aspects; Springer International Publishing: Cham, Switzerland, 2018; pp. 63–83.
14. Chacón, J.E.; Duong T. Multivariate Kernel Smoothing and Its Applications; CRC Press: Boca Raton, FL, USA, 2018; pp. 43–66.
15. Wand, M.P.; Jones, M.C. Multivariate plug-in bandwidth selection. *Comput. Stat.* 1994, 9, 97–116.
16. Duong, T.; Hazelton, M. Plug-in bandwidth matrices for bivariate kernel density estimation. *J. Nonparametr Stat.* 2003, 15, 17–30. [CrossRef]
17. Chacón, J.E.; Duong, T. Multivariate plug-in bandwidth selection with unconstrained pilot bandwidth matrices. *Test* 2010, 19, 375–398. [CrossRef]
18. Brewer, M. Bayesian model for local smoothing in kernel density estimation. *Stat. Comput.* 2000, 10, 299–309. [CrossRef]
19. Zhang, X.; King, M.L.; Hyndman, R.J. A Bayesian approach to bandwidth selection for multivariate kernel density estimation. *Comput. Stat. Data Anal.* 2006, 50, 3009–3031. [CrossRef]
20. Zougab, N.; Adjabi, S.; Kokonendji, C. Bayesian estimation of adaptive bandwidth matrices in multivariate kernel density estimation. *Comput. Stat. Data Anal.* 2014, 75, 28–38. [CrossRef]
21. Belaid, N.; Adjabi, S.; Kokonendji, C.C.; Zougab, N. Bayesian adaptive bandwidth selector for multivariate discrete kernel estimator. *Commun. Stat.-Theory Methods* 2018, 47, 2988–3001. [CrossRef]
22. Gramacki, A.; Gramacki, J. FFT-based fast computation of multivariate kernel density estimators with unconstrained bandwidth matrices. *J. Comput. Graph. Stat.* 2017, 26, 459–462. [CrossRef]
23. Sain, S.R.; Baggerly, K.A.; Scott, D.W. Cross-validation of multivariate densities. *J. Am. Stat. Assoc.* 1994, 89, 807–817. [CrossRef]
24. Wu, T.; Chen, C.; Chen, H. A variable bandwidth selector in multivariate kernel density estimation. *Stat. Probab. Lett.* 2007, 77, 462–467. [CrossRef]
25. Xu, J. Generalized Spatial Variable Coefficient Model. Ph.D. Thesis, Hunan Normal University, Changsha, China, 2009.
26. Wand, M.P.; Jones, M.C. Comparison of smoothing parameterizations in bivariate kernel density estimation. *J. Am. Stat. Assoc.* 1993, 88, 520–528. [CrossRef]
27. Mei, C.L.; Wang, N. Modern Regression Analysis Method; Science Press: Beijing, China, 2012; pp. 163–190.
28. Feng, J.; Zhang, H.; Hu, X. Spatial non-stationary characteristics of the impacts of precipitation and temperature on FVC: A case study of Yili River Valley, Xinjiang. *Chin. J. Ecol.* 2016, 36, 4626–4634.
29. Leung, Y.; Mei, C.L.; Zhang, W.X. Statistical Tests for Spatial Nonstationarity Based on the Geographically Weighted Regression Model. *Environ. Plan. A* 2000, 32, 9–32. [CrossRef]
30. Mei, C.L.; Wang, N.; Zhang, W.X. Testing the importance of the explanatory variables in a mixed geographically weighted regression model. *Environ. Plan. A Econ. Space* 2006, 38, 587–598. [CrossRef]
31. Mei, C.L.; Xie, Q. On the estimation and testing of mixed geographically weighted regression models. *Econ. Model.* 2012, 29, 2615–2620. [CrossRef]
32. Mei, C.L.; Xu, M.; Wang, N. A bootstrap test for constant coefficients in geographically weighted regression models. *Int. J. Geogr. Inf. Sci.* 2016, 30, 1622–1643. [CrossRef]