# Local Linear Estimation of Spatially Varying Coefficient Models: An Improvement on the Geographically Weighted Regression Technique

**Article** *in* Environment and Planning A · April 2008

DOI: 10.1068/a3941 · Source: RePEc



**3 authors**, including:

Changlin Mei
Xi'an Polytechnic University
**86** PUBLICATIONS **2,138** CITATIONS

SEE PROFILE

Xiaodong Yan
Beijing Normal University
**169** PUBLICATIONS **2,336** CITATIONS

SEE PROFILE

*Environment and Planning A* 2008, volume 40, pages 986 – 1005

doi:10.1068/a3941

# Local linear estimation of spatially varying coefficient models: an improvement on the geographically weighted regression technique

**Ning Wang, Chang-Lin Mei**
School of Science, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, People's Republic of China; e-mail: wning@mail.xjtu.edu.cn; clmei@mail.xjtu.edu.cn

**Xiao-Dong Yan**
Key Laboratory of Regional Climate-Environment Research for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, People's Republic of China; e-mail: yxd@tea.ac.cn
Received 4 February 2006; in revised form 3 July 2006; published online 11 February 2008

**Abstract.** Geographically weighted regression (GWR), as a useful method for exploring spatial nonstationarity of a regression relationship, has been applied to a variety of areas. In this method a spatially varying coefficient model is locally calibrated and the spatial-variation patterns of the locally estimated regression coefficients are taken as the main evidence of spatial nonstationarity for the underlying data-generating processes. Therefore, the validity of the analysis results drawn by GWR is closely dependent on the accuracy between the underlying coefficients and their estimates. Motivated by the local polynomial-modelling technique in statistics, we propose a local linear-based GWR for the spatially varying coefficient models, in which the coefficients are locally expanded as linear functions of the spatial coordinates and then estimated by the weighted least-squares procedure. Some theoretical and numerical comparisons with GWR are conducted and the results demonstrate that the proposed method can significantly improve GWR, not only in goodness-of-fit of the whole regression function but also in reducing bias of the coefficient estimates.

## 1 Introduction

The technique of geographically weighted regression (GWR), proposed by Fotheringham and his colleagues in 1996, is of great appeal for the analysis of spatial data in a number of areas such as geography, econometrics, epidemiology, and environmental science. In GWR the spatial nonstationarity of a regression relationship is explored by locally calibrating a spatially varying coefficient model of the form

$$y_i = \sum_{j=1}^{p} \beta_j(u_i, v_i)x_{ij} + \varepsilon_i, \qquad i = 1, 2, \ldots, n, \tag{1}$$

with the weighted least-squares procedure, where $(y_i; x_{i1}, \ldots, x_{ip})$ are observations of the response $y$ and explanatory variables $x_1, \ldots, x_p$ at a location $(u_i, v_i)$ in the studied geographical region, the regression coefficients $\beta_j(u_i, v_i)$ $(j = 1, 2, \ldots, p)$ are unknown functions of the location coordinates, and $\varepsilon_i$ $(i = 1, 2, \ldots, n)$ are error terms with mean zero and common variance $\sigma^2$. The model allows a spatially varying intercept term by selecting $x_{i1} \equiv 1\,(i = 1, 2, \ldots, n)$.

In addition to the pioneering work by Brunsdon et al (1996; 1998; 1999a) and Fotheringham et al (1997a; 1997b), GWR has been extensively studied in various ways. For example, Fotheringham et al (1998) compared GWR with the expansion method and Brunsdon et al (1999b) compared GWR with random coefficient modelling by use of a case study. Leung et al (2000a; 2000b) discussed the statistical tests for nonstationarity of the coefficients and spatial autocorrelation among the residuals of GWR calibration. Páez et al (2002a; 2002b) proposed a maximum-likelihood-based framework of estimation and inference for GWR from the perspective of error-variance heterogeneity.

The seminal monograph by Fotheringham et al (2002) has systematically summarized the basic theory and statistical inference problems of GWR and addressed some extensions of the GWR technique to generalized linear models, principal component analysis, and so on. The multicollinearity and correlation among the GWR estimates of the coefficients has been investigated by Wheeler and Tiefelsdorf (2005). Furthermore, Páez (2005) conducted some simulations to compare GWR with the expansion method in the validity of the coefficient estimates and found that, on average, GWR is flexible enough to capture the true patterns of the coefficient variation, at least in some simple cases.

On the other hand, there has been a large body of literature on applications of GWR in a variety of fields. For example, with the GWR technique, Fotheringham et al (2001) analyzed spatial variation in school performance; Brunsdon et al (2001) investigated the spatial variations of the average rainfall – altitude relationship in Great Britain. The GWR technique was used by Calvo and Escolar (2003) to solve the problems of spatial aggregation bias and spatial autocorrelation in ecological inference. Foody (2003; 2004) explored, with GWR, the spatial relationships between the normalized-difference vegetation index and rainfall, as well as the spatial relationships between species richness and environmental determinants. GWR was also employed by Zhang et al (2004) and Zhang and Shi (2004) to analyze the spatial characteristics of tree growth. Recently, Wang et al (2005) used GWR to estimate the net primary production of Chinese forest ecosystems. For other applications of GWR see Atkinson et al (2003), Huang and Leung (2002), Longley and Tobón (2004) McMillen (2004), Mittal et al (2004), Nakaya (2001), Nakaya et al (2005), among others.

It may be concluded that GWR is being established as a standard tool in exploratory spatial-data analysis due to its effectiveness and wide applications. In this technique the spatial-variation patterns of the locally estimated regression coefficients are taken as the main evidence to infer whether the underlying data-generating process exhibits a spatial nonstationarity or local deviations from a global regression model. Therefore, the validity of the analysis results drawn by GWR largely depends on the accuracy between the estimates of the coefficients and their actual values. Methodologically, however, GWR is essentially a kind of Nadaraya – Watson kernel smoother (for example, see Härdle, 1990; Wand and Jones, 1995) in the sense that the coefficients in the model in equation (1) are locally taken to be constant in order to obtain their estimates by the weighted least-squares procedure. As is well known, the Nadaraya – Watson kernel-smoothing approach generally leads to less-accurate estimates for the regression function, and suffers from the so-called boundary effect, which means that bias of the regression-function estimate is of a larger order at the boundary than in the interior of the domain of the exploratory variable (see Fan and Gijbels, 1996). The boundary-effect problem for GWR has been addressed by, for example, Brunsdon et al (1996) and Fotheringham et al (1997a), and may lead to misleading estimates of the coefficients and induce unsupported spatial patterns onto the analysis results. Therefore, it is important to develop some methods to tackle the bias-reduction problem in GWR. Some efforts have been devoted to this aspect. For instance, the mixed GWR model (Brunsdon et al, 1999a; Fotheringham et al, 2002; Mei et al, 2004; 2006) and the semiparametric geographically weighted Poisson regression (Nakaya et al, 2005) can be considered as a trial for reduction estimation biases of GWR, because some coefficients are treated as constants to be estimated, which will reduce bias of the GWR estimation if these coefficients are indeed spatially nonvarying. The use of improved bandwidth-selection criteria such as the corrected Akaike information criterion ($AIC_c$) and adaptive bandwidths (see, among others, Brunsdon et al, 1999a; Fotheringham et al, 2002; Hurvich et al, 1998) is another effort to reduce the bias of GWR estimation.

Nevertheless, the emphasis of bandwidth-selection criteria mainly focus on the estimate of whole-regression functions rather than the estimates of the coefficients themselves. To the best of our knowledge, very little, except for the work by Páez (2005), has been reported in the literature about the ability of GWR to retrieve spatial-variation patterns of the coefficients.

It has been noted that the recently developed local polynomial-modelling method has gained much attention in statistics because of its many advantages including good asymptotic properties of bias and variance of the estimates, good behavior in terms of minimax efficiency, and, more importantly, the ability of automatically correcting for the boundary effect (see Fan and Gijbels, 1996, for details). Motivated by this methodology, we propose in this paper an improved calibrating procedure, called local linear-fitting-based GWR, for the spatially varying coefficient model in equation (1). In the proposed method the coefficients are locally expanded as linear functions of the spatial coordinates, which is a special case of the local polynomial-modelling methodology, and then they are locally estimated in each location by the weighted least-squares procedure. Given the aforementioned advantages of the local polynomial-modelling method, we may expect that the proposed method is better performing than GWR, especially in reducing bias of the coefficient estimates and in truly retrieving spatial-variation patterns of the coefficients.

The remainder of this paper is organized as follows: in section 2 we briefly describe the GWR technique to facilitate our discussion; the proposed local linear-fitting based GWR is derived in section 3; some theoretical and numerical comparisons of the proposed method and GWR are performed, respectively, in sections 4 and 5; the paper is then concluded with a brief summary and discussion.

## 2 Brief description of the GWR technique

In order to facilitate the derivation of the proposed method and the comparison of the proposed method and GWR, we briefly describe the GWR technique in this section.

For a given location or focal point $(u_0, v_0)$ in the studied geographical region, let $d_{0i}$ be the distance between $(u_0, v_0)$ and the $i$th designed location $(u_i, v_i)$. The regression coefficients in the model in equation (1) at $(u_0, v_0)$ are locally estimated by formulating the weighted least-squares problem. Namely, minimize

$$\sum_{i=1}^{n}\left[y_i - \sum_{j=1}^{p}\beta_j(u_0, v_0)x_{ij}\right]^2 K_h(d_{0i}) \ , \tag{2}$$

with respect to $\beta_j(u_0, v_0)$ $(j = 1, 2, \ldots, p)$, where $K_h(\cdot) = K(\cdot/h)$ with $K(\cdot)$ being a given kernel function and $h$ being the bandwidth. Solving this weighted least-squares problem leads to the estimates $\hat{\beta}_j(u_0, v_0)$ of $\beta_j(u_0, v_0)$ $(j = 1, 2, \ldots, p)$ at $(u_0, v_0)$, and they can be expressed with the matrix notation as

$$\begin{aligned}\hat{\boldsymbol{\beta}}(u_0, v_0) &= \left[\hat{\beta}_1(u_0, v_0), \hat{\beta}_2(u_0, v_0), \ldots, \hat{\beta}_p(u_0, v_0)\right]^{\mathrm{T}} \\ &= \left[\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_0, v_0)\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_0, v_0)\boldsymbol{Y} \ ,\end{aligned} \tag{3}$$

where

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \qquad \boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \ , \tag{4}$$

and

$$\mathbf{W}(u_0, v_0) = \mathrm{diag}[K_h(d_{01}), K_h(d_{02}), \ldots, K_h(d_{0n})] \ , \tag{5}$$

is a diagonal weight matrix. By taking $(u_0, v_0)$ to be each of the designed locations $(u_i, v_i)$ $(i = 1, 2, \ldots, n)$, respectively, we can obtain the vector of the fitted values for the response $y$ at $n$ designed locations $(u_i, v_i)$ $(i = 1, 2, \ldots, n)$ as

$$\hat{\boldsymbol{Y}} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)^{\mathrm{T}} = \mathbf{C}\boldsymbol{Y} ,$$

where

$$\mathbf{C} = \begin{pmatrix} \boldsymbol{x}_1^{\mathrm{T}}[\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_1, v_1)\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_1, v_1) \\ \boldsymbol{x}_2^{\mathrm{T}}[\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_2, v_2)\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_2, v_2) \\ \vdots \\ \boldsymbol{x}_n^{\mathrm{T}}[\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_n, v_n)\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_n, v_n) \end{pmatrix} \tag{6}$$

is called a hat matrix in terms of regression analysis and $\boldsymbol{x}_i^{\mathrm{T}} = (x_{i1}, x_{i2}, \ldots, x_{ip})$ is the $i$th row of the matrix $\mathbf{X}$ in equation (4). The resultant residual vector and residual sum of squares (RSS) under the hat matrix $\mathbf{C}$ are, respectively,

$$\hat{\boldsymbol{\varepsilon}}_c = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = (\mathbf{I} - \mathbf{C})\boldsymbol{Y} ,$$

and

$$\mathrm{RSS}_c = \hat{\boldsymbol{\varepsilon}}_c^{\mathrm{T}}\boldsymbol{\varepsilon}_c = \boldsymbol{Y}^{\mathrm{T}}(\mathbf{I} - \mathbf{C})^{\mathrm{T}}(\mathbf{I} - \mathbf{C})\boldsymbol{Y} ,$$

where $\mathbf{I}$ is the identity matrix of order $n$.

The kernel function $K(\cdot)$ is commonly chosen as the Gaussian kernel:

$$K(t) = \exp\left(-\frac{1}{2}t^2\right) , \tag{7}$$

or the bi-square function:

$$K(t) = \begin{cases} (1 - t^2)^2 , & \text{if } |t| \leqslant 1 , \\ 0 , & \text{otherwise} . \end{cases} \tag{8}$$

The cross-validation (CV) procedure is generally employed to select the optimal value of bandwidth $h$. In this procedure, the score:

$$\mathrm{CV}(h) = \sum_{i=1}^{n} [y_i - \hat{y}_{(i)}(h)]^2 , \tag{9}$$

is used as an objective function where $\hat{y}_{(i)}(h)$ is the fitted value of $y_i$ under bandwidth $h$ with the observation at location $(u_i, v_i)$ omitted from the process of calibration. Select $h_0$ as the optimal bandwidth value, such that:

$$\mathrm{CV}(h_0) = \min_{h>0} \mathrm{CV}(h) .$$

Other approaches, such as the generalized cross-validation (GCV) criterion and the $\mathrm{AIC}_c$ (see Hurvich et al, 1998; Fotheringham et al, 2002, chapter 2), are also applicable to GWR for bandwidth selection. In order to distinguish GWR from the proposed method that will be derived in the next section, we henceforth call GWR 'the basic GWR'.

## 3 Local linear-fitting-based GWR

In local polynomial modelling, the regression function is locally expanded as a $p$-order polynomial of the explanatory variable by Taylor's formula, and the weighted least-squares procedure is used to locally estimate the regression function and its derivatives (up to $p$-order) at each given point in the domain of the explanatory variable. Motivated by this methodology, we can locally expand each of the coefficients $\beta_j(u, v)$

($j = 1, 2, \ldots, p$) in the model in equation (1) as a linear function of the spatial coordinates $u$ and $v$ at a given location $(u_0, v_0)$ in the studied region, and then a similar calibration procedure to the basic GWR is used to obtain the local estimates of each coefficient and its two partial derivatives at $(u_0, v_0)$.

Suppose that each of the coefficients $\beta_j(u, v)$ ($j = 1, 2, \ldots, p$) in the spatially varying coefficient model in equation (1) has second continuous partial derivatives with respect to the geographical location coordinates $u$ and $v$. For a given location $(u_0, v_0)$ and according to Taylor's expansion of a bivariate function, $\beta_j(u, v)$ ($j = 1, 2, \ldots, p$) can be locally approximated as

$$\begin{aligned} \beta_j(u, v) &\approx \beta_j(u_0, v_0) + \beta_j^{(u)}(u_0, v_0)(u - u_0) + \beta_j^{(v)}(u_0, v_0)(v - v_0) \ , \\ j &= 1, 2, \ldots, p \ , \end{aligned} \tag{10}$$

in a neighbourhood of $(u_0, v_0)$, where $\beta_j^{(u)}(u_0, v_0)$ and $\beta_j^{(v)}(u_0, v_0)$ are the partial derivatives of $\beta_j(u, v)$ at $(u_0, v_0)$ with respect to $u$ and $v$, respectively. With the approximate expressions of $\beta_j(u, v)$ ($j = 1, 2, \ldots, p$) in equation (10) and according to the principle of the local polynomial modelling, the following weighted least-squares problem is formulated. That is, minimize

$$\sum_{i=1}^{n} \left\{ y_i - \sum_{j=1}^{p} [\beta_j(u_0, v_0) + \beta_j^{(u)}(u_0, v_0)(u_i - u_0) + \beta_j^{(v)}(u_0, v_0)(v_i - v_0)] x_{ij} \right\}^2 K_h(d_{0i}), \tag{11}$$

with respect to $\beta_j(u_0, v_0)$, $\beta_j^{(u)}(u_0, v_0)$, and $\beta_j^{(v)}(u_0, v_0)$ ($j = 1, 2, \ldots, p$), where $K_h(d_{0i})$ ($i = 1, 2, \ldots, n$) and those in equation (2) are equal.

For this weighted least-squares problem, an explicit expression of the solution can be, as usual, obtained with the matrix notation as follows. Let

$$\mathbf{X}(u_0, v_0) = \begin{pmatrix} x_{11} & x_{11}(u_1 - u_0) & x_{11}(v_1 - v_0) & \cdots & x_{1p} & x_{1p}(u_1 - u_0) & x_{1p}(v_1 - v_0) \\ x_{21} & x_{21}(u_2 - u_0) & x_{21}(v_2 - v_0) & \cdots & x_{2p} & x_{2p}(u_2 - u_0) & x_{2p}(v_2 - v_0) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n1}(u_n - u_0) & x_{n1}(v_n - v_0) & \cdots & x_{np} & x_{np}(u_n - u_0) & x_{np}(v_n - v_0) \end{pmatrix}, \tag{12}$$

and

$$\boldsymbol{P}(u_0, v_0) = [\beta_1(u_0, v_0), \beta_1^{(u)}(u_0, v_0), \beta_1^{(v)}(u_0, v_0), \ldots, \beta_p(u_0, v_0), \beta_p^{(u)}(u_0, v_0), \beta_p^{(v)}(u_0, v_0)]^{\mathrm{T}} \ .$$

Then the solution of the weighted least-squares problem in equation (11) or the estimate of parameter vector $\boldsymbol{P}(u_0, v_0)$ at location $(u_0, v_0)$ can be expressed by:

$$\begin{aligned} \hat{\boldsymbol{P}}(u_0, v_0) &= [\hat{\beta}_1(u_0, v_0), \hat{\beta}_1^{(u)}(u_0, v_0), \hat{\beta}_1^{(v)}(u_0, v_0), \ldots, \hat{\beta}_p(u_0, v_0), \hat{\beta}_p^{(u)}(u_0, v_0), \hat{\beta}_p^{(v)}(u_0, v_0)]^{\mathrm{T}} \\ &= [\mathbf{X}^{\mathrm{T}}(u_0, v_0)\mathbf{W}(u_0, v_0)\mathbf{X}(u_0, v_0)]^{-1}\mathbf{X}^{\mathrm{T}}(u_0, v_0)\mathbf{W}(u_0, v_0)\boldsymbol{Y} \ , \end{aligned} \tag{13}$$

where $\boldsymbol{Y}$ and $\mathbf{W}(u_0, v_0)$ are the same matrices as those in equations (4) and (5)—that is, the $n$-dimensional vector consisting of the observations of the response $y$ at $n$ locations and the $n \times n$ weight matrix at $(u_0, v_0)$.

For each of $j = 1, 2, \ldots, p$, we then take $\hat{\beta}_j(u_0, v_0)$ in $\hat{\boldsymbol{P}}(u_0, v_0)$ as an estimate of $\beta_j(u_0, v_0)$ at $(u_0, v_0)$. Furthermore, two partial derivatives—$\beta_j^{(u)}(u_0, v_0)$ and $\beta_j^{(v)}(u_0, v_0)$ of $\beta_j(u, v)$ at $(u_0, v_0)$—can be estimated by $\hat{\beta}_j^{(u)}(u_0, v_0)$ and $\hat{\beta}_j^{(v)}(u_0, v_0)$, which reflect

the local varying ratios of $\beta_j(u, v)$ along the two axis directions of the spatial region, respectively.

Furthermore, let $\boldsymbol{e}_{k,3p}$ be a $3p$-dimensional column vector with the $k$th element being 1 and the other elements being 0. Then the estimates of $\beta_j(u, v)$ and its two partial derivatives at $(u_0, v_0)$ can be closely expressed for each of $j = 1, 2, \ldots, p$ as

$$\begin{cases} \hat{\beta}_j(u_0, v_0) = \boldsymbol{e}_{3j-2,3p}^{\mathrm{T}} \hat{\boldsymbol{P}}(u_0, v_0), \\ \hat{\beta}_j^{(u)}(u_0, v_0) = \boldsymbol{e}_{3j-1,3p}^{\mathrm{T}} \hat{\boldsymbol{P}}(u_0, v_0), \\ \hat{\beta}_j^{(v)}(u_0, v_0) = \boldsymbol{e}_{3j,3p}^{\mathrm{T}} \hat{\boldsymbol{P}}(u_0, v_0), \end{cases} \tag{14}$$

where $\hat{\boldsymbol{P}}(u_0, v_0)$ is shown in equation (13).

Taking $(u_0, v_0) = (u_i, v_i)$ $(i = 1, 2, \ldots, n)$, we can obtain the estimates of each regression coefficient $\beta_j(u, v)$ as well as its partial derivatives at all of the designed geographical locations, respectively. Once the estimates $\hat{\beta}_j(u_i, v_i)$ $(j = 1, 2, \ldots, p)$ are obtained, the fitted values of the response $y$ at $n$ designed locations can be computed by

$$\hat{y}_i = \sum_{j=1}^{p} \hat{\beta}_j(u_i, v_i) x_{ij} = \boldsymbol{x}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}(u_i v_i), \qquad i = 1, 2, \ldots, n, \tag{15}$$

where $\hat{\boldsymbol{\beta}}(u_i, v_i) = [\hat{\beta}_1(u_i, v_i), \hat{\beta}_2(u_i, v_i), \ldots, \hat{\beta}_p(u_i, v_i)]^{\mathrm{T}}$ is the vector of the estimates of $\beta_j(u_i, v_i)$ $(j = 1, 2, \ldots, p)$ at $(u_i, v_i)$ and $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^{\mathrm{T}}$ is the vector of observations of the explanatory variables $x_1, x_2, \ldots, x_p$ at $(u_i, v_i)$.

We henceforth refer the above method for calibrating the spatially varying coefficient model in equation (1) to local linear-fitting-based GWR.

Furthermore, some commonly used quantities in the local linear-fitting-based GWR can be expressed by matrix notation in closer forms, which is useful for the purposes of practical computation. Firstly, according to the first equation in (14), the vector of the estimated coefficients at each designed location $(u_i, v_i)$ can be expressed as

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = \begin{pmatrix} \hat{\beta}_1(u_i, v_i) \\ \hat{\beta}_2(u_i, v_i) \\ \vdots \\ \hat{\beta}_p(u_i, v_i) \end{pmatrix} = \begin{pmatrix} \boldsymbol{e}_{1,3p}^{\mathrm{T}} \hat{\boldsymbol{P}}(u_i, v_i) \\ \boldsymbol{e}_{4,3p}^{\mathrm{T}} \hat{\boldsymbol{P}}(u_i, v_i) \\ \vdots \\ \boldsymbol{e}_{3p-2,3p}^{\mathrm{T}} \hat{\boldsymbol{P}}(u_i, v_i) \end{pmatrix} = \mathbf{Q} \hat{\boldsymbol{P}}(u_i, v_i), \tag{16}$$

where

$$\mathbf{Q} = \begin{pmatrix} \boldsymbol{e}_{1,3p}^{\mathrm{T}} \\ \boldsymbol{e}_{4,3p}^{\mathrm{T}} \\ \vdots \\ \boldsymbol{e}_{3p-2,3p}^{\mathrm{T}} \end{pmatrix}$$

is a $p \times 3p$ matrix with the $i$th row being $\boldsymbol{e}_{3i-2,3p}^{\mathrm{T}}$ and $\hat{\boldsymbol{P}}(u_i, v_i)$ is computed from equation (13) by replacing $(u_0, v_0)$ with $(u_i, v_i)$. Then, with equations (14) and (15), the vector of the fitted values of the response $y$ at $n$ designed locations is

$$\hat{\boldsymbol{Y}} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1^{\mathrm{T}} \mathbf{Q} \hat{\boldsymbol{P}}(u_1, v_1) \\ \boldsymbol{x}_2^{\mathrm{T}} \mathbf{Q} \hat{\boldsymbol{P}}(u_2, v_2) \\ \vdots \\ \boldsymbol{x}_n^{\mathrm{T}} \mathbf{Q} \hat{\boldsymbol{P}}(u_n, v_n) \end{pmatrix} = \mathbf{L} \boldsymbol{Y}, \tag{17}$$

where

$$\mathbf{L} = \begin{pmatrix} \boldsymbol{x}_1^{\mathrm{T}}\mathbf{Q}[\mathbf{X}^{\mathrm{T}}(u_1, v_1)\mathbf{W}(u_1, v_1)\mathbf{X}(u_1, v_1)]^{-1}\mathbf{X}^{\mathrm{T}}(u_1, v_1)\mathbf{W}(u_1, v_1) \\ \boldsymbol{x}_2^{\mathrm{T}}\mathbf{Q}[\mathbf{X}^{\mathrm{T}}(u_2, v_2)\mathbf{W}(u_2, v_2)\mathbf{X}(u_2, v_2)]^{-1}\mathbf{X}^{\mathrm{T}}(u_2, v_2)\mathbf{W}(u_2, v_2) \\ \vdots \\ \boldsymbol{x}_n^{\mathrm{T}}\mathbf{Q}[\mathbf{X}^{\mathrm{T}}(u_n, v_n)\mathbf{W}(u_n, v_n)\mathbf{X}(u_n, v_n)]^{-1}\mathbf{X}^{\mathrm{T}}(u_n, v_n)\mathbf{W}(u_n, v_n) \end{pmatrix} . \quad (18)$$

Therefore, the residual vector under the hat matrix $\mathbf{L}$ is

$$\hat{\boldsymbol{\varepsilon}}_L = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n)^{\mathrm{T}} = \boldsymbol{Y} - \mathbf{L}\boldsymbol{Y} = (\mathbf{I} - \mathbf{L})\boldsymbol{Y} ,$$

and the residual sum of squares is

$$\mathrm{RSS}_L = \hat{\boldsymbol{\varepsilon}}_L^{\mathrm{T}}\hat{\boldsymbol{\varepsilon}}_L = \boldsymbol{Y}^{\mathrm{T}}(\mathbf{I} - \mathbf{L})^{\mathrm{T}}(\mathbf{I} - \mathbf{L})\boldsymbol{Y} .$$

It can be observed from equation (17) that, like the basic GWR, the local linear-fitting-based GWR is still a linear smoother, in the sense that each fitted value $\hat{y}_i$ is a linear combination of the observations $y_1, y_2, \ldots, y_n$ of the response $y$. Furthermore, at each designed location the same weight matrix as that in equation (5) for the basic GWR is used here to derive the estimates of the coefficients and their partial derivatives. Therefore, the kernel functions such as Gaussian kernel and bi-square kernel in equations (7) and (8) are still applicable to the proposed method. Additionally, the cross-validation procedure used in the basic GWR for bandwidth selection can be borrowed to achieve the same task in the local linear-fitting-based GWR, where each value $\hat{y}_{(i)}(h)$ in the CV($h$) score in equation (9) is computed by equation (15) with the $i$th observation deleted while computing the estimate $\hat{\boldsymbol{\beta}}(u_i, v_i)$. Some alternative methods, such as GCV and $\mathrm{AIC}_c$ criteria, can also be used to determine the optimal bandwidth value in the local linear-fitting-based GWR. We omitted the specific descriptions of these bandwidth-selection procedures for the proposed method, as we considered them to be essentially the same as those in the basic GWR.

## 4 Theoretical comparison with the basic GWR

In the basic GWR the regression coefficients $\beta_j(u, v)$ ($j = 1, 2, \ldots, p$) in the model in equation (1) are locally taken to be constant at each focal point $(u_0, v_0)$ when formulating the weighted least-squares problem in equation (2). Methodologically, the basic GWR is comparable to the Nadaraya – Watson kernel-smoothing approach in the context of statistical nonparametric regression (for example, see Härdle, 1990; Wand and Jones, 1995), although the whole regression function, rather than the regression coefficients, is locally treated as a constant in the Nadaraya – Watson kernel-smoothing method. It has been theoretically proven that the Nadaraya – Watson kernel-smoothing method suffers from the boundary effect (for example, see Fan and Gijbels, 1996; Wand and Jones, 1995). With this argument, we infer that the estimates of the coefficients with the basic GWR may suffer from a more serious boundary effect because the related spatial region is 2-dimensional and more locations are on the boundary, although the theoretical confirmation remains to be investigated for GWR.

In contrast to the basic GWR, however, the local linear-fitting-based GWR formulates the weighted least-squares problem in equation (11) by locally taking the regression coefficients $\beta_j(u, v)$ ($j = 1, 2, \ldots, p$) to be linear functions of the spatial coordinates at each focal point $(u_0, v_0)$, which is comparable to the local-linear smoother—a special but commonly used case of local polynomial modelling (see Fan and Gijbels, 1996). As mentioned in the introduction, the local polynomial-modelling technique outperforms the Nadaraya – Watson method in many aspects, of which the

most important one is the ability to automatically correct for the boundary effect. That is, the order of bias of the estimate for the regression function on the boundary remains the same as that in the interior of the domain of the explanatory variable. Furthermore, the estimation bias of the local polynomial regression is smaller than that of the Nadaraya–Watson kernel smoother. Given the advantages of the local polynomial regression, we may expect that the proposed local linear-fitting-based GWR can improve the performance of the basic GWR.

The following theoretical result can further demonstrate, to some extent, that the local linear-fitting-based GWR is a refinement of the basic GWR.

**Theorem** *When all the coefficients* $\beta_j(u, v)$ $(j = 1, 2, \ldots, p)$ *in the model in equation* (1) *are exactly linear functions of the spatial coordinates* $(u, v)$, *the estimates of both the coefficients and their partial derivatives derived by the local linear-fitting-based GWR are unbiased estimators of the corresponding coefficients and their constant partial derivatives at each focal point* $(u_0, v_0)$ *for any bandwidth h. In contrast, when all of the coefficients* $\beta_j(u, v)$ $(j = 1, 2, \ldots, p)$ *are constant, their estimates obtained by the basic GWR are unbiased at each focal point* $(u_0, v_0)$ *for any bandwidth h.*

The proof is given in the appendix.

It has been described by Fotheringham et al (1998) that the basic GWR is a natural evolution of the expansion method (see Casetti, 1972; Jones and Casetti, 1992). As a further refinement of the basic GWR, the proposed local linear-fitting-based GWR can be considered as an improvement of the expansion method in the sense that each regression coefficient is locally taken as a linear function rather than as a global-specific function of the spatial coordinates. This improvement largely enhances the flexibility and suitability of the expansion method and may capture more complex variation patterns of the coefficients.

## 5 Numerical comparison with the basic GWR

In this section we conduct further some simulation experiments to give a numerical comparison of the proposed method and the basic GWR, not only in goodness-of-fit of the whole regression function but also in reducing bias of the coefficient estimates.

### 5.1 Spatial layout for simulation and model for generating datasets

The spatial layout for simulation is designed as a two-dimensional square region with the length of each side being 12 units. A Cartesian coordinate system is built in such a way that its origin is located at the bottom-left corner of this square region. The locations where the observations are collected consist of $m \times m$ lattice points with a distance $l = 12/(m-1)$ between any two neighboring points along the horizontal and vertical axes. These locations lead to a sample size of $n = m^2$ for the observations.

Take $m = 7$, 13, and 25 leading to small, moderate, and large sample sizes of $n = 49$, 169, and 625, respectively. Under the above specific Cartesian coordinate system, and for these three sample sizes, the spatial coordinates of locations $(u_i, v_i)$ for collecting the observations can respectively be computed by

(i) $(u_i, v_i) = \left[2 \bmod\left(\frac{i-1}{7}\right), 2 \operatorname{int}\left(\frac{i-1}{7}\right)\right], \qquad i = 1, 2, \ldots, 49$ ;

(ii) $(u_i, v_i) = \left[\bmod\left(\frac{i-1}{13}\right), \operatorname{int}\left(\frac{i-1}{13}\right)\right], \qquad i = 1, 2, \ldots, 169$ ;

(iii) $(u_i, v_i) = \left[0.5 \bmod\left(\frac{i-1}{25}\right), 0.5 \operatorname{int}\left(\frac{i-1}{25}\right)\right], \qquad i = 1, 2, \ldots, 625$ ;

where $\bmod(i-1/m)$ is the remainder of $i-1$ divided by $m$, and $\operatorname{int}(i-1/m)$ is the integer part of number $i-1/m$ for $m = 7$, 13, and 25, respectively. In fact, each latter case can be a refinement of the former case by letting the distance between any two

neighboring points along the horizontal and vertical axes be half of the corresponding distance of the former case.

The spatially varying coefficient model for generating datasets is

$$y_i = \beta_1(u_i, v_i) + \beta_2(u_i, v_i)x_i + \varepsilon_i, \qquad i = 1, 2, \ldots, n ,$$

where the observations $x_i, x_2, \ldots, x_n$ of the explanatory variable $x$ are randomly drawn from the uniform distribution on interval $(0, 2)$ and the error terms $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independently drawn from the normal distribution $N(0, 0.25)$. The following two cases for the regression coefficients are considered:

case 1 $\beta_1(u, v) = \frac{1}{6}(u + v)$, $\beta_2(u, v) = \frac{1}{3}u$ ;

case 2 $\beta_1(u, v) = 4\sin\left(\frac{1}{12}\pi u\right)$, $\beta_2(u, v) = \frac{1}{324}\left[36 - (6 - u)^2\right]\left[36 - (6 - v)^2\right]$ ,

where $(u, v)$ are spatial coordinates of any location in the square region, and each regression coefficient is adjusted by an appropriate constant so that its value in the region is between 0 and 4. We depict these regression coefficients in figures 1 and 2, respectively, to give a visual explanation.

The motivation for considering the regression coefficients in case 1 comes from the theorem in section 4: if all of the regression coefficients are linear functions of the spatial coordinates, the local linear-fitting-based GWR will produce unbiased estimates of the coefficients. Therefore, we take case 1 into account to testify this fact with a number of replications. In addition to the purpose of conducting simulation under

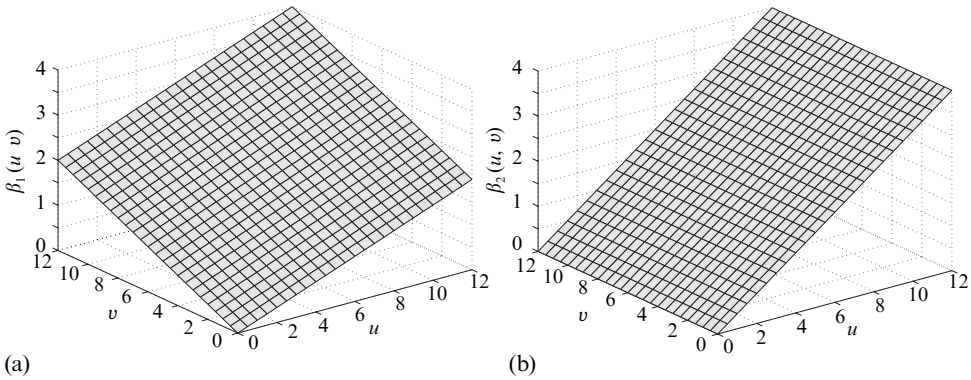


**Figure 1.** True surfaces of the regression coefficients in case 1. (a) $\beta_1(u, v)$ and (b) $\beta_2(u, v)$.

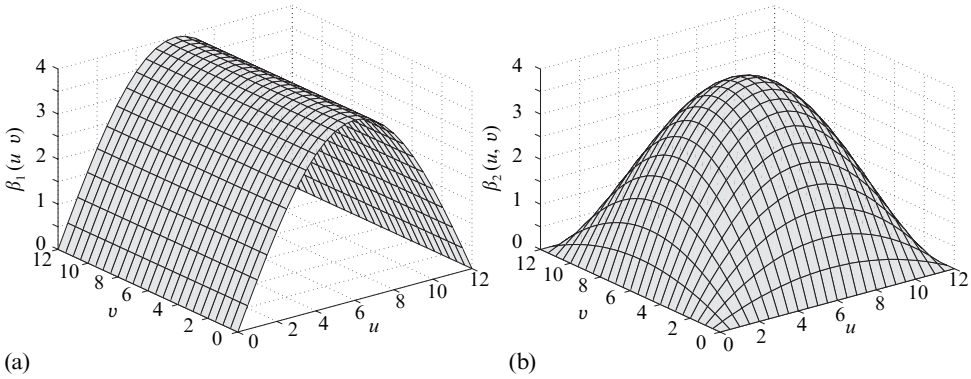


**Figure 2.** True surfaces of the regression coefficients in case 2. (a) $\beta_1(u, v)$ and (b) $\beta_2(u, v)$.

more complicated spatial variation patterns of the coefficients, case 2 is partially motivated by some application backgrounds. For example, when we consider the relationship between house price and house area in a city where a main avenue passes through the city north to south, the price of land for building an average sized house along the two sides of the avenue may be approximately described by $\beta_1(u, v)$, and the house price per unit area around the center of the city may approximately follow the variation pattern of $\beta_2(u, v)$ in case 2.

**5.2 Statistics for evaluating model-fitting effects**

For the purpose of comparing the local linear-fitting-based GWR with the basic GWR numerically, the following statistics are introduced to evaluate the model-fitting effects of the two approaches.

5.2.1 *Statistic for evaluating goodness-of-fit of regression function*

It is recalled that, for a linear smoother of the form

$$\hat{\boldsymbol{Y}} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)^{\mathrm{T}} = \mathbf{S}\boldsymbol{Y} ,$$

the standard residual sum of squares:

$$\hat{\sigma}^2 = \frac{1}{\mathrm{tr}[(\mathbf{I}-\mathbf{S})^{\mathrm{T}}(\mathbf{I}-\mathbf{S})]}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{\boldsymbol{Y}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})^{\mathrm{T}}(\mathbf{I}-\mathbf{S})\boldsymbol{Y}}{\mathrm{tr}[(\mathbf{I}-\mathbf{S})^{\mathrm{T}}(\mathbf{I}-\mathbf{S})]} , \tag{19}$$

is generally taken as an estimator of variance $\sigma^2$ of the error terms in the model, where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. The quantity $\hat{\sigma}^2$ measures goodness-of-fit of the regression function for the linear smoother in the sense that, if the model is well fitted, $\hat{\sigma}^2$ should be close to $\sigma^2$, at least when the sample size $n$ is sufficiently large. We therefore use this quantity in our simulation to evaluate goodness-of-fit of the two methods by replacing $\mathbf{S}$ with the matrix $\mathbf{C}$ in equation (6) for the basic GWR and with $\mathbf{L}$ in equation (18) for the local linear-fitting-based GWR, respectively. Furthermore, considering the sampling error, we shall run $N$ replications for the error terms of the model and taken the mean of $\hat{\sigma}^2$ under $N$ replications:

$$M(\hat{\sigma}^2) = \frac{1}{N}\sum_{k=1}^{N}\hat{\sigma}^2(k) ,$$

as a final measure of goodness of fit for each of the two calibrating methods, where $\hat{\sigma}^2(k)$ stands for the value of $\hat{\sigma}^2$ in the $k$th replication.

5.2.2 *Statistic for evaluating bias of coefficient estimates*

For the $j$th regression coefficient $\beta_j(u, v)$ and a given designed location $(u_i, v_i)$, let

$$M[\hat{\beta}_j(u_i, v_i)] = \frac{1}{N}\sum_{k=1}^{N}\hat{\beta}_{j(k)}(u_i, v_i) \tag{20}$$

be the mean value of the estimates of $\beta_j(u_i, v_i)$ under $N$ replications, where $\hat{\beta}_{j(k)}(u_i, v_i)$ is the estimate of $\beta_j(u_i, v_i)$ in the $k$th replication. If $N$ is sufficiently large, $M[\hat{\beta}_j(u_i, v_i)]$ should be close to $\mathrm{E}[\hat{\beta}_j(u_i, v_i)]$. Therefore, the quantity $\beta_j(u_i, v_i) - M[\hat{\beta}_j(u_i, v_i)]$ approximately measures the bias of the estimate of $\beta_j(u, v)$ at $(u_i, v_i)$. We then define

$$B(j) = \frac{1}{n}\sum_{i=1}^{n}\left\{\beta_j(u_i, v_i) - M[\hat{\beta}_j(u_i, v_i)]\right\}^2 , \tag{21}$$

to evaluate the mean bias of the estimates of $\beta_j(u, v)$ at all of the $n$ designed locations. In the following simulation, $\beta_j(u_i, v_i)$ $(i = 1, 2, \ldots, n)$ for $j = 1, 2$ are computed by their respective expressions in cases 1 and 2 and $\hat{\beta}_j(u_i, v_i)$ $(i = 1, 2, \ldots, n)$

are obtained from equation (3) for the basic GWR and from equation (16) for the local linear-fitting-based GWR.

Although the quantity $B(j)$ is not computable in practical applications, in that the real values of each regression coefficient $\beta_j(u_i, v_i)$ are certainly unknown, it does make sense in simulation for comparing bias of the estimated coefficients by different calibration methods.

**5.3 Simulation results with analysis**

Throughout the simulation, the Gaussian kernel in equation (7) was used to generate the weights at each location $(u_i, v_i)$, and the cross-validation procedure was employed in each replication to select the optimal bandwidth value for both the basic GWR and the local linear-fitting-based GWR. However, some preliminary computation showed that the original form of Gaussian weights at each $(u_i, v_i)$

$$K_h(d_{ik}) = \exp\left[-\frac{1}{2}\left(\frac{d_{ik}}{h}\right)^2\right], \qquad k = 1, 2, \ldots, n,$$

makes it inconvenient to select the optimal value of $h$ for the local linear-fitting-based GWR in case 1 because the optimal value of $h$ in this case is generally very large and even approaches infinity in some replications. It is observed that, when all the coefficients are exactly linear functions of spatial coordinates, say, $\beta_j(u, v) = a_j + b_j u + c_j v$ $(j = 1, 2, \ldots, p)$, the spatially varying coefficient model in equation (1) becomes a global model of the form

$$y_i = \sum_{j=1}^{p}(a_j + b_j u_i + c_j v_i)x_{ij} + \varepsilon_i, \qquad i = 1, 2, \ldots, n. \tag{22}$$

It has been known that, in the basic GWR, an infinitely large bandwidth size leads to an identity-weight matrix at each focal point $(u_0, v_0)$, which makes the basic GWR estimates of the coefficients coincide with the ordinary least-squares estimates of the corresponding global-linear model (that is, all of the coefficients are constant). For the local linear-fitting-based GWR, although the theoretical investigation remains to be studied, we surmise that an infinitely large bandwidth size may imply that the global model in equation (22) is suitable for the data and that the ordinary least-squares estimates of the parameters $a_j$, $b_j$, and $c_j$ $(j = 1, 2, \ldots, p)$ may have some close relation to the estimates of the linear coefficients obtained by the proposed method with an identity-weight matrix (that is, an infinitely large bandwidth size). To make the weight matrix be an identity matrix at any focal point $(u_0, v_0)$, an infinitely large bandwidth size should be needed. Because of the aforementioned phenomenon for the proposed method, the transformed Gaussian weights

$$K_\theta(d_{ik}) = \exp(-\theta d_{ik}^2), \qquad k = 1, 2, \ldots, n,$$

at each $(u_i, v_i)$ were used in the following simulation for both methods. Here, the parameter for the weights is taken to be $\theta = 1/2h^2$ in order to avoid searching for its optimal value in too large a range and make it possible to include $\theta = 0$, which corresponds to an infinitely large value of $h$.

For cases 1 and 2 as well as sample sizes $n = 49$ and 169, we used fifty draws of the explanatory variable $x$, each with $N = 100$ replications of the error terms. The values of the statistics in equations (19) and (21) were computed for each method, respectively, to evaluate the model-fitting effects. The computation was implemented by writing the code under the environment of interactive matrix language procedure of SAS software (SAS Institute Inc., 1988). The related results are depicted in figure 3–6, respectively.

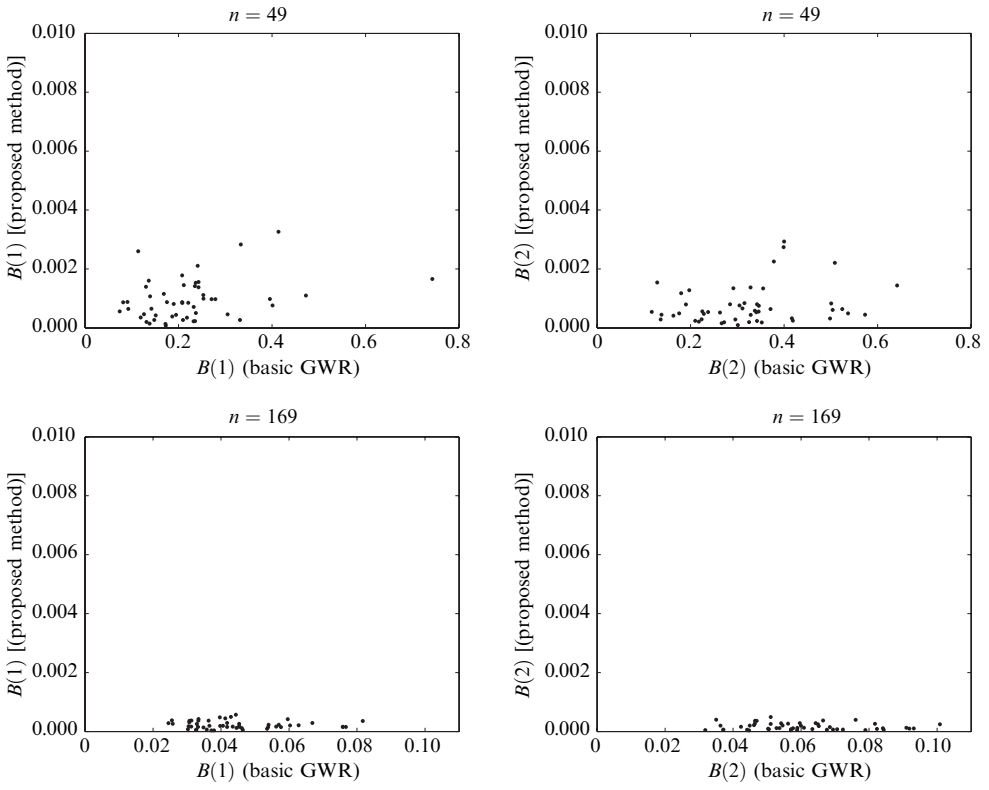


**Figure 3.** $B(j)$ of the basic geographically weighted regression (GWR) versus that of the proposed method for fifty draws of $x$, each with 100 replications of the error terms, for case 1.

Each point in figures 3 and 4 corresponds to a pair of $B(j)$ values obtained by the basic GWR and the proposed method, respectively, under the same draw of $x$. As the figures show, the proposed method substantially and consistently outperforms the basic GWR in terms of the mean bias of the coefficient estimates.

The proposed method performs especially well in case 1, where each of the coefficients is a linear function of spatial coordinates $(u, v)$. Noting the very different scales on the horizontal and vertical axes, we see that, for each draw of $x$, a nearly perfect fit for two coefficients is obtained by the proposed method, in the sense that the mean biases of the estimated coefficients under the 100 replications of the error terms are almost zero for each draw of $x$. This coincides with the theoretical result in section 4 that the local linear-fitting-based GWR will produce unbiased estimates of the coefficients $\beta_1(u, v)$ and $\beta_2(u, v)$ at any location $(u_0, v_0)$ in this case. Although the corresponding results obtained by the basic GWR are quite satisfactory under these simple forms of the coefficients, they are evidently not as good as those derived by the proposed method, especially for the smaller sample size.

For case 2, where the coefficient variation patterns are more complex than those in case 1, the local linear-fitting-based GWR still consistently outperforms the basic GWR for all of the fifty draws of $x$. Furthermore, with the sample size increasing, the values of $B(j)$ $(j = 1, 2)$ for the proposed method decrease more significantly than those for the basic GWR.

Figures 5 and 6 show the values of $M(\hat{\sigma}^2)$ for all of the fifty draws of $x$ with sample sizes $n = 49$ and 169 for cases 1 and 2, respectively. It is observed that, although the values of $M(\hat{\sigma}^2)$ for both methods all tend to $\sigma^2 = 0.25$ with sample sizes

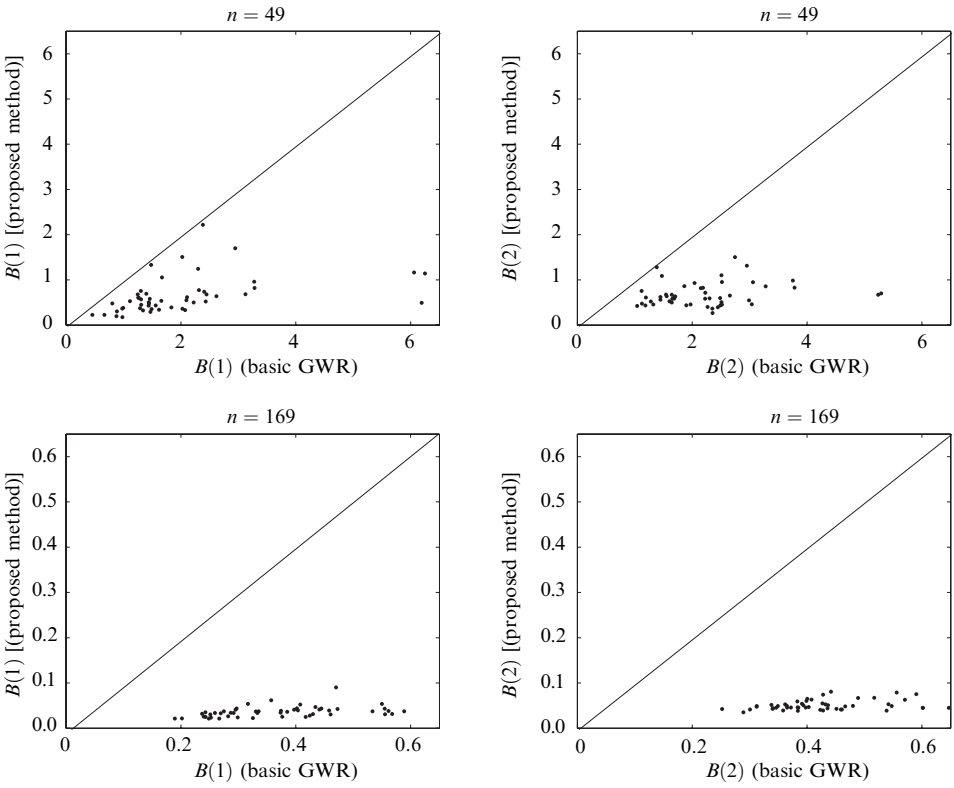


**Figure 4.** $B(j)$ of the basic geographically weighted regression (GWR) versus that of the proposed method for fifty draws of $x$, each with 100 replications of the error terms, for case 2.

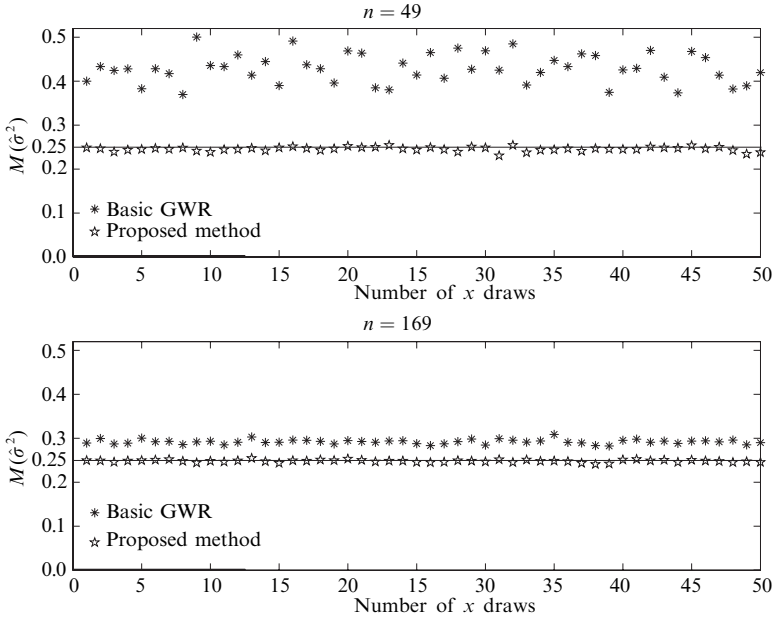


**Figure 5.** $M(\hat{\sigma}^2)$ of the basic geographically weighted regression (GWR) and the proposed method for fifty draws of $x$, each with 100 replications of the error terms, for case 1.

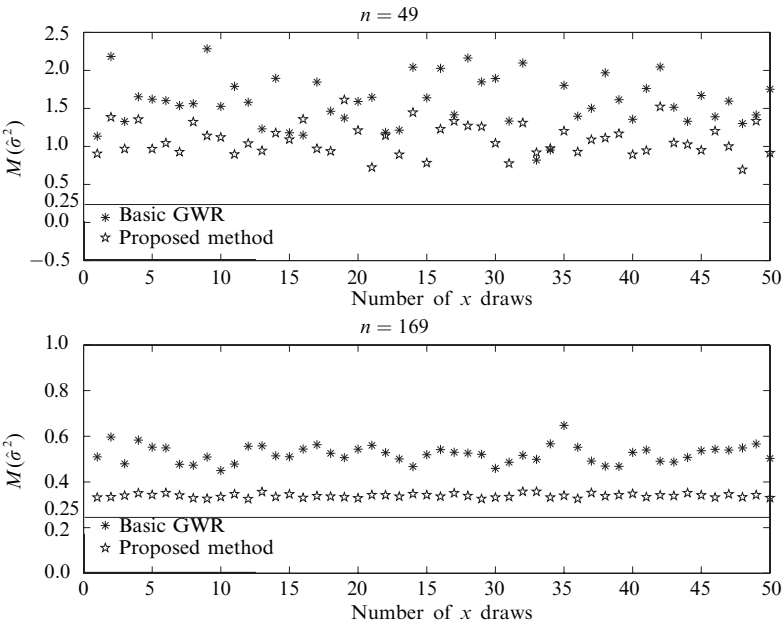


**Figure 6.** $M(\hat{\sigma}^2)$ of the basic geographically weighted regression (GWR) and the proposed method for fifty draws of $x$, each with 100 replications of the error terms, for case 1.

increasing, those for the proposed method are consistently closer to $\sigma^2 = 0.25$ for each of the fifty draws of $x$, except for four draws of $x$ in case 2 with the sample size of 49. This implies that the proposed method also outperforms the basic GWR in fitting the whole regression function. Especially in case 1, the values of $M(\hat{\sigma}^2)$ for the proposed method are almost equal to $\sigma^2 = 0.25$ for all of the fifty draws of $x$ under both sample sizes, but those for the basic GWR are systematically and significantly larger than $\sigma^2 = 0.25$. This can be explained by the theoretical result in section 4 that the proposed method produces an unbiased estimate for each of the coefficients in this case, and therefore generates a more accurate estimator of $\sigma^2$.

For both cases, carrying out replications of the explanatory variable $x$ for sample size $n = 625$ is very time consuming. We only generated one draw of $x$ with $N = 100$ replications of the error terms for this sample size to run the simulation. For each case, the values of $M[\hat{\beta}_j(u_i, v_i)]$ $(j = 1, 2)$ defined in equation (20) were outputted at all of the 625 designed locations for both methods, and they are depicted in figures 7 – 10 to give a visual comparison of the basic GWR and the proposed method in retrieving the spatial-variation patterns of the coefficients. Comparing these figures with their respective true surfaces shown in figures 1 and 2, we can observe that the local linear-fitting-based GWR retrieves the spatial-variation patterns of the coefficients much better than the basic GWR, not only on the boundary but also in the interior of the spatial region. This shows that the proposed method maintains the advantage of the local polynomial regression in automatically correcting for the boundary effect.

It has been known in the context of nonparametric regression that the cross-validation procedure leads to a highly variable choice of the bandwidth and has a tendency towards undersmoothing (for example, see Loader, 1999). However, $\text{AIC}_c$ can avoid the large variability and tendency to undersmooth (see Hurvich et al, 1998) and is one of commonly used criteria for bandwidth selection. With this fact in mind, we reconducted all of the simulations except for that with sample size $n = 625$ by using $\text{AIC}_c$ for bandwidth selection and found that, for both cases and both methods, there

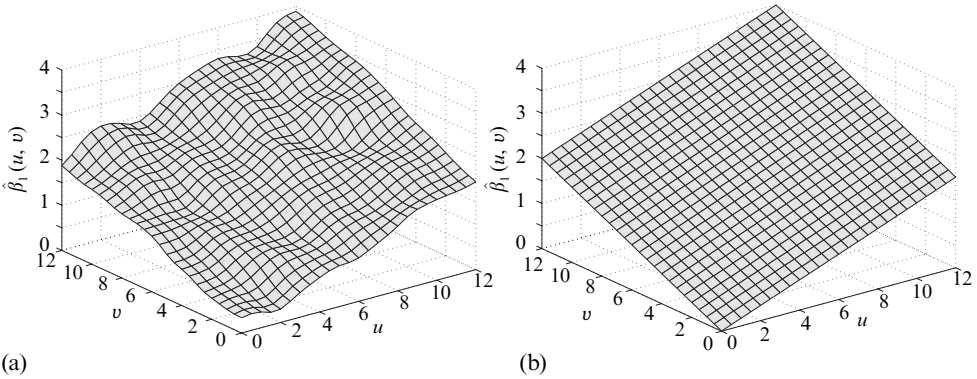


**Figure 7.** Surfaces of $M[\hat{\beta}_1(u_i, v_i)]$ obtained by (a) the basic geographically weighted regression (GWR) and (b) the local linear-fitting-based GWR in case 1 under 100 replications of the error terms.

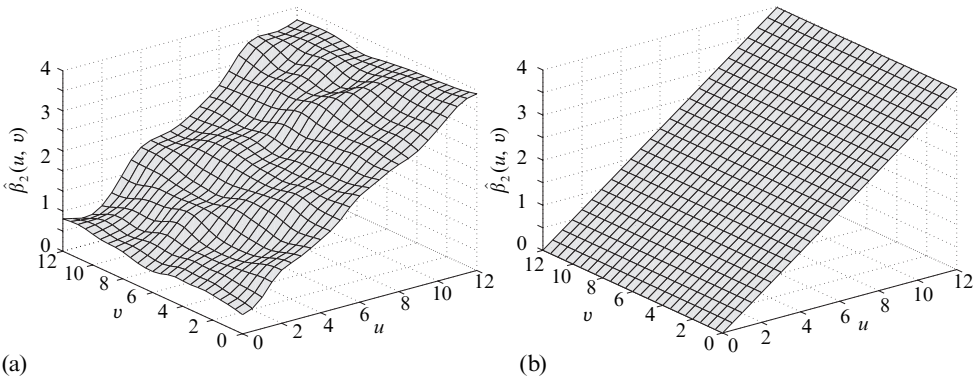


**Figure 8.** Surfaces of $M[\hat{\beta}_2(u_i, v_i)]$ obtained by (a) the basic geographically weighted regression (GWR) and (b) the local linear-fitting-based GWR in case 1 under 100 replications of the error terms.

is no evident difference between the mean biases of the coefficient estimates [that is, the value of $B(j)$, $j = 1, 2$] obtained by the CV and $AIC_c$ criteria respectively, although $AIC_c$ does select more stable and slightly larger values of bandwidth $h$ (that is, smaller values of $\theta$). However, the values of $M(\hat{\sigma}^2)$ with $AIC_c$ is generally larger than those with CV, especially for the smaller sample size of forty nine. The reason for this is that a large bandwidth size leads to a smoother estimate of the whole regression function, which results in a larger difference between the observed and fitted values of the response $y$, which therefore results in larger values of $M(\hat{\sigma}^2)$ for both methods. The above findings may demonstrate that bandwidth variation could exert less influence on the coefficient estimates than on the residual sum of squares. The specific simulation results with $AIC_c$ are not reported here (they are available from the authors) in consideration of the limited space available.

## 6 Summary and discussion

The basic GWR has been extensively studied and widely applied to a variety of areas for spatial data analysis. In this technique the validity of the estimates for the regression coefficients in the model in equation (1) is extremely important, because the spatial nonstationarity of the regression relationship is explored and interpreted based

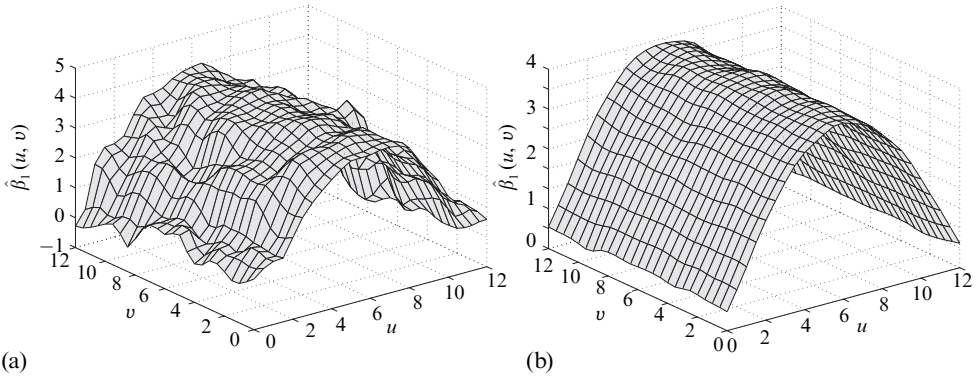


**Figure 9.** Surfaces of $M[\hat{\beta}_1(u_i, v_i)]$ obtained by (a) the basic geographically weighted regression (GWR) and (b) the local linear-fitting-based GWR in case 2 under 100 replications of the error terms.

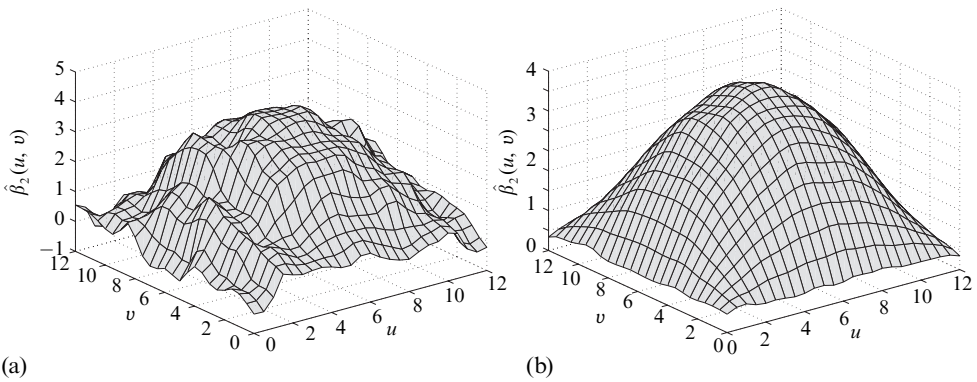


**Figure 10.** Surfaces of $M[\hat{\beta}_2(u_i, v_i)]$ obtained by (a) the basic geographically weighted regression (GWR) and (b) the local linear-fitting-based GWR in case 2 under 100 replications of the error terms.

mainly on the variation patterns of the estimated coefficients. Motivated by the local polynomial modelling technique in statistics, we propose an improved calibration approach for the spatially varying coefficient model, called local linear-fitting-based GWR, to achieve the task of reducing bias and the boundary effect of the coefficient estimates. With the proposed method we can simultaneously obtain the estimates of not only the coefficients themselves but also their partial derivatives. The estimated partial derivatives may provide some more comprehensive information on the spatial variation of the coefficients, although this issue is not discussed further for this paper. Some theoretical and numerical comparisons of the proposed method and the basic GWR are performed, and the results demonstrate that the proposed method substantially improves the performance of the basic GWR both in goodness-of-fit of the whole regression function and in reducing the bias of the coefficient estimates.

In this paper we consider only local linear expansion of the coefficients, which is a special but commonly used case of the local polynomial modelling technique. Theoretically, provided that the coefficients $\beta_j(u, v)$ $(j = 1, 2, \ldots, p)$ are sufficiently smooth, the proposed approach can be extended directly to a more-general situation where $\beta_j(u, v)$ $(j = 1, 2, \ldots, p)$ is locally expanded as a higher order of bivariate polynomial to calibrate locally the spatially varying coefficients model in equation (1).

However, because the coefficients $\beta_j(u, v)$ $(j = 1, 2, \ldots, p)$ are bivariate functions of the spatial coordinates, a higher order of local polynomial expansion will greatly increase the number of parameters that need to be estimated locally at each location, and therefore a large data size is needed to make these parameters estimable. The gain may be relatively small compared with the increase on the complexity of model calibration and the overhead of computation. From the viewpoints of simplicity and effectiveness in methodology, and feasibility in application, local linear expansion of the coefficients in the proposed method may be an appropriate choice in producing less-biased estimates of the coefficients.

Based on the proposed method, we may further expect that some more elaborate hypotheses on the spatially varying coefficient model will be statistically tested with the estimates of the coefficients and their partial derivatives. In addition to the typical hypothesis-testing problems that have been addressed in the context of the basic GWR (for example, see Brunsdon et al, 1999a; Fotheringham et al, 2002; Leung et al, 2000a; 2000b), some other hypotheses on the functional forms of the coefficients $\beta_j(u, v)$ $(j = 1, 2, \ldots, p)$ may be further tested, based on the estimates of partial derivatives of the coefficients. For example, the hypothesis that a certain coefficient is a linear function of the spatial coordinates $(u, v)$ can be tested by simultaneously examining whether the estimates of two partial derivatives of that coefficient at each location $(u, v)$ are equal to some constants; the hypothesis that a certain coefficient is unrelated to one of the spatial coordinates $u$ or $v$ can also be tested by evaluating whether the estimates of the partial derivatives with respect to that coordinate are equal to zero at all locations. The study of these types of hypothesis-testing problems may give deeper insight into the characteristics of spatial variation of the regression relationship. Another important issue for both the basic GWR and the proposed method is to investigate theoretically the orders of the asymptotic bias and the variance of the estimated coefficients. This problem may be difficult, but it deserves to be studied in the future.

**Acknowledgements.** This research is supported by the open project of Key Laboratory of Regional Climate – Environment Research for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, and the National Natural Science Foundation of China (No. 10531030). The authors would like to thank the reviewers for their valuable comments and suggestions especially on the simulations, which led to significant improvements on the paper.

**References**

Atkinson P M, German S E, Sear D A, Clark M J, 2003, "Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression" *Geographical Analysis* **35** 58 – 82

Brunsdon C, Fotheringham A S, Charlton M, 1996, "Geographically weighted regression: a method for exploring spatial nonstationarity" *Geographical Analysis* **28** 281 – 298

Brunsdon C, Fotheringham A S, Charlton M, 1998, "Geographically weighted regression: modelling spatial nonstationarity" *The Statistician* **47** 431 – 443

Brunsdon C, Fotheringham A S, Charlton M, 1999a, "Some notes on parametric significance tests for geographically weighted regression" *Journal of Regional Science* **39** 497 – 524

Brunsdon C, Aitkin M, Fotheringham A S, Charlton M, 1999b, "A comparison of random coefficient modelling and geographically weighted regression for spatially non-stationary regression problems" *Geographical and Environmental Modelling* **3** 47 – 62

Brunsdon C, McClatchey J, Unwin D J, 2001, "Spatial variations in the average rainfall – altitude relationship in Great Britain: an approach using geographically weighted regression" *International Journal of Climatology* **21** 455 – 466

Calvo E, Escolar M, 2003, "The local voter: a geographically weighted approach to ecological inference" *American Journal of Political Science* **47** 189 – 204

Casetti E, 1972, "Generating models by the expansion method: applications to geographic research" *Geographical Analysis* **4** 81 – 91

Fan J, Gijbels I, 1996 *Local Polynomial Modelling and Its Applications* (Chapman and Hall, London)

Foody G M, 2003, "Geographical weighting as a further refinement to regression modelling: an example focused on the NDVI – rainfall relationship" *Remote Sensing of Environment* **88** 283 – 293

Foody G M, 2004, "Spatial nonstationarity and scale-dependency in the relationship between species and environmental determinants for the sub-Saharan endemic avifauna" *Global Ecology and Biogeography* **13** 315 – 320

Fotheringham A S, Charlton M, Brunsdon C, 1997a, "Measuring spatial variations in relationships with geographically weighted regression", in *Recent Developments in Spatial Analysis* Eds M M Fischer, A Getis (Springer, London) pp 60 – 82

Fotheringham A S, Charlton M, Brunsdon C, 1997b, "Two techniques for exploring non-stationarity in geographical data" *Geographical System* **4** 59 – 82

Fotheringham A S, Charlton M, Brunsdon C, 1998, "Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis" *Environment and Planning A* **30** 1905 – 1927

Fotheringham A S, Charlton M, Brunsdon C, 2001, "Spatial variations in school performance: a local analysis using geographically weighted regression" *Geographical and Environmental Modelling* **5** 43 – 66

Fotheringham A S, Brunsdon C, Charlton M, 2002 *Geographically Weighted Regression—The Analysis of Spatial Varying Relationships* (John Wiley, Chichester, Sussex)

Härdle W, 1990 *Applied Nonparametric Regression* (Cambridge University Press, Boston, MA)

Huang Y, Leung Y, 2002, "Analysing regional industrialisation in Jiangsu province using geographically weighted regression" *Journal of Geographical Systems* **4** 233 – 249

Hurvich C M, Simonoff J S, Tsai C L, 1998, "Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion" *Journal of the Royal Statistical Society, Series B* **60** 271 – 293

Jones III J P, Casetti E, 1992 *Applications of the Expansion Method* (Routledge, London)

Leung Y, Mei C L, Zhang W X, 2000a, "Statistical tests for spatial nonstationarity based on the geographically weighted regression model" *Environment and Planning A* **32** 9 – 32

Leung Y, Mei C L, Zhang W X, 2000b, "Testing for spatial autocorrelation among the residuals of the geographically weighted regression" *Environment and Planning A* **32** 871 – 890

Loader C, 1999 *Local Regression and Likelihood* (Springer, New York)

Longley P A, Tobón C, 2004, "Spatial dependence and heterogeneity in patterns of hardship: an intra-urban analysis" *Annals of the Association of American Geographers* **94** 503 – 519

McMillen D P, 2004, "Employment densities, spatial autocorrelation, and subcenters in large metropolitan areas" *Journal of Regional Science* **44** 225 – 243

Mei C L, He S Y, Fang K T, 2004, "A note on the mixed geographically weighted regression model" *Journal of Regional Science* **44** 143 – 157

Mei C L, Wang N, Zhang W X, 2006, "Testing the importance of the explanatory variables in a mixed geographically weighted regression model" *Environment and Planning A* **38** 587 – 598

Mittal V, Kamakura W A, Govind R, 2004, "Geographic patterns in customer service and satisfaction: an empirical investigation" *Journal of Marketing* **68** 48 – 62

Nakaya T, 2001, "Local spatial interaction modelling based on the geographically weighted regression approach" *GeoJournal* **53** 347 – 358

Nakaya T, Fotheringham A S, Brunsdon C, Charlton M, 2005, "Geographically weighted Poisson regression for disease association mapping" *Statistics in Medicine* **24** 2695 – 2717

Páez A, 2005, "Local analysis of spatial relationships: a comparison of GWR and the expansion method", in *Lecture Notes in Computer Science, Volume 3482* Eds O Gervasi, M L Garrilova, V Kumar, A Lagana, H P Lee, Y Mun, D Taniar, C J K Tan (Springer, Berlin) pp 162 – 172

Páez A, Uchida T, Miyamoto K, 2002a, "A general framework for estimation and inference of geographically weighted regression models: 1. Location-specific kernel bandwidths and a test for locational heterogeneity" *Environment and Planning A* **34** 733 – 754

Páez A, Uchida T, Miyamoto K, 2002b, "A general framework for estimation and inference of geographically weighted regression models: 2. Spatial association and model specification tests" *Environment and Planning A* **34** 883 – 904

SAS Institute Inc., 1988 *SAS/IML, User's Guide, Release 6.03 Edition* (SAS Institute Inc., Cary, NC)

Wand M P, Jones M C, 1995 *Kernel Smoothing* (Chapman and Hall, London)

Wang Q, Ni J, Tenhunen J, 2005, "Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems" *Global Ecology and Biogeography* **14** 379 – 393

Wheeler D, Tiefelsdorf M, 2005, "Multicollinearity and correlation among local regression coefficients in geographically weighted regression" *Journal of Geographical Systems* **7** 161 – 187
Zhang L, Shi H, 2004, "Local modelling of tree growth by geographically weighted regression" *Forest Science* **50** 225 – 244
Zhang L, Bi H, Cheng P, Davis C J, 2004, "Modeling spatial variation in tree diameter-height relationships" *Forest Ecology and Management* **189** 317 – 329

**Appendix**

In the appendix we will examine the proof of the theorem introduced in section 4.

For the local linear-fitting-based GWR, suppose that

$$\beta_j(u, v) = a_j + b_j u + c_j v, \qquad j = 1, 2, \ldots, p ,$$

where $a_j$, $b_j$, and $c_j$ for each $j$ are constant. Then, for any focal point $(u_0, v_0)$ in the studied region, we have the exact expressions

$$\beta_j(u, v) = \beta_j(u_0, v_0) + \beta_j^{(u)}(u_0, v_0)(u - u_0) + \beta_j^{(v)}(u_0, v_0)(v - v_0),$$
$$j = 1, 2, \ldots, p ,$$

where $\beta_j(u_0, v_0) = a_j + b_j u_0 + c_j v_0$, $\beta_j^{(u)}(u_0, v_0) = b_j$, and $\beta_j^{(v)}(u_0, v_0) = c_j$ for $j = 1, 2, \ldots, p$. In this case, we have, for each $i$,

$$y_i = \sum_{j=1}^{p} \beta_j(u_i, v_i)x_{ij} + \varepsilon_i$$
$$= \sum_{j=1}^{p} \left[\beta_j(u_0, v_0) + \beta_j^{(u)}(u_0, v_0)(u_i - u_0) + \beta_j^{(v)}(u_0, v_0)(v_i - v_0)\right] x_{ij} + \varepsilon_i .$$

Therefore, the model in equation (1) can be expressed with the matrix notation as

$$\boldsymbol{Y} = (y_i, y_2, \ldots, y_n)^{\mathrm{T}} = \mathbf{X}(u_0, v_0)\boldsymbol{P}(u_0, v_0) + \boldsymbol{\varepsilon} ,$$

where $\mathbf{X}(u_0, v_0)$ is shown in equation (12), $\boldsymbol{P}(u_0, v_0) = [\beta_1(u_0, v_0), \beta_1^{(u)}(u_0, v_0), \beta_1^{(v)}(u_0, v_0), \ldots, \beta_p(u_0, v_0), \beta_p^{(u)}(u_0, v_0), \beta_p^{(v)}(u_0, v_0)]^{\mathrm{T}}$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^{\mathrm{T}}$ with $\mathrm{E}(\boldsymbol{\varepsilon}) = 0$. Substituting the above expression of $\boldsymbol{Y}$ into equation (13), we obtain the following for any bandwidth $h$:

$$\hat{\boldsymbol{P}}(u_0, v_0)$$
$$= [\hat{\beta}_1(u_0, v_0), \hat{\beta}_1^{(u)}(u_0, v_0), \hat{\beta}_1^{(v)}(u_0, v_0), \ldots, \hat{\beta}_p(u_0, v_0), \hat{\beta}_p^{(u)}(u_0, v_0), \hat{\beta}_p^{(v)}(u_0, v_0)]^{\mathrm{T}}$$
$$= [\mathbf{X}^{\mathrm{T}}(u_0, v_0)\mathbf{W}(u_0, v_0)\mathbf{X}(u_0, v_0)]^{-1}\mathbf{X}^{\mathrm{T}}(u_0, v_0)\mathbf{W}(u_0, v_0)[\mathbf{X}(u_0, v_0)\boldsymbol{P}(u_0, v_0) + \boldsymbol{\varepsilon}]$$
$$= \boldsymbol{P}(u_0, v_0) + [\mathbf{X}^{\mathrm{T}}(u_0, v_0)\mathbf{W}(u_0, v_0)\mathbf{X}(u_0, v_0)]^{-1}\mathbf{X}^{\mathrm{T}}(u_0, v_0)\mathbf{W}(u_0, v_0)\boldsymbol{\varepsilon} .$$

So, we get

$$\mathrm{E}[\hat{\boldsymbol{P}}(u_0, v_0)]$$
$$= \{\mathrm{E}[\hat{\beta}_1(u_0, v_0)], \mathrm{E}[\hat{\beta}_1^{(u)}(u_0, v_0)], \mathrm{E}[\hat{\beta}_1^{(v)}(u_0, v_0)], \ldots, \mathrm{E}[\hat{\beta}_p(u_0, v_0)] ,$$
$$\mathrm{E}[\hat{\beta}_p^{(u)}(u_0, v_0)], \mathrm{E}[\hat{\beta}_p^{(v)}(u_0, v_0)]\}^{\mathrm{T}}$$
$$= \boldsymbol{P}(u_0, v_0)$$
$$= [\beta_1(u_0, v_0), \beta_1^{(u)}(u_0, v_0), \beta_1^{(v)}(u_0, v_0), \ldots, \beta_p(u_0, v_0), \beta_p^{(u)}(u_0, v_0), \beta_p^{(v)}(u_0, v_0)]^{\mathrm{T}} .$$

This means that

$$\mathrm{E}[\hat{\beta}_j(u_0, v_0)] = \beta_j(u_0, v_0) = a_j + b_j u_0 + c_j v_0 , \qquad \mathrm{E}[\hat{\beta}_j^{(u)}(u_0, v_0)] = \beta_j(u_0, v_0) = b_j ,$$

and $\mathrm{E}[\hat{\beta}_j^{(v)}(u_0, v_0)] = \beta_j^{(v)}(u_0, v_0) = c_j$ for every $j = 1, 2, \ldots, p$ and any bandwidth $h$.

For the basic GWR, when all of the coefficients $\beta_j(u, v)$ $(j = 1, 2, \ldots, p)$ are constant, say

$$\beta_j(u, v) = \beta_j, \qquad j = 1, 2, \ldots, p,$$

the model in equation (1) becomes

$$\boldsymbol{Y} = (y_1, y_2, \ldots, y_n)^{\mathrm{T}} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^{\mathrm{T}}$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)^{\mathrm{T}}$, with $\mathrm{E}(\boldsymbol{\varepsilon}) = 0$. Substituting the above expression of $\boldsymbol{Y}$ into equation (3), we obtain at any focal point $(u_0, v_0)$

$$\begin{aligned}\hat{\boldsymbol{\beta}}(u_0, v_0) &= \left[\hat{\beta}_1(u_0, v_0), \hat{\beta}_2(u_0, v_0), \ldots, \hat{\beta}_p(u_0, v_0)\right]^{\mathrm{T}} \\ &= [\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_0, v_0)\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_0, v_0)[\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}] \\ &= \boldsymbol{\beta} + [\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_0, v_0)\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_0, v_0)\boldsymbol{\varepsilon}.\end{aligned}$$

Therefore,

$$\mathrm{E}[\hat{\boldsymbol{\beta}}(u_0, v_0)] = \boldsymbol{\beta} + [\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_0, v_0)\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_0, v_0)\mathrm{E}(\boldsymbol{\varepsilon}) = \boldsymbol{\beta},$$

which means that $\mathrm{E}[\hat{\beta}_j(u_0, v_0)] = \beta_j$ for every $j = 1, 2, \ldots, p$ and any bandwidth $h$.