

Title of the paper: A Modification to Geographically Weighted Regression

Authors: Yin-Yee Leong¹ and Jack C. Yue*¹

¹Department of Statistics, National Chengchi University, Taipei, Taiwan, R.O.C. 11605

*Corresponding author:

Email: Yin-Yee Leong - 100354501@nccu.edu.tw ; Jack C. Yue* - csyue@nccu.edu.tw

Abstract

Background

Geographically weighted regression (GWR) is a modelling technique designed to deal with spatial non-stationarity, e.g., the mean values vary by locations. It has been widely used as a visualization tool to explore the patterns of spatial data. However, the GWR tends to produce unsmooth surfaces when the mean parameters have considerable variations, partly due to that all parameter estimates are derived from a fixed range (bandwidth) of observations. In order to deal with the varying bandwidth problem, this paper proposes an alternative approach, namely Conditional geographically weighted regression (CGWR)

Methods

The estimation of CGWR is based on an iterative procedure, analogy to the numerical optimization problem. Computer simulation, under realistic settings, is used to compare the performance between the traditional GWR, CGWR, and a local linear modification of GWR. Furthermore, this study also applies the CGWR to two empirical datasets for evaluating the model performance. The first dataset consists of disability status of Taiwan's elderly, along with some social-economic variables and the other is Ohio's crime dataset.

Results

Under the positively correlated scenario, we found that the CGWR produces a better fit for the response surface. Both the computer simulation and empirical analysis support the proposed approach since it significantly reduces the bias and variance of data fitting. In addition, the response surface from the CGWR reviews local spatial characteristics according to the corresponded variables.

Conclusions

As an explanatory tool for spatial data, producing accurate surface is essential in order to provide a first look at the data. Any distorted outcomes would likely mislead the following analysis. Since the CGWR can generate more accurate surface, it is more appropriate to use it exploring data that contain suspicious variables with varying characteristics.

Keywords: Geographically weighted regression, Modifiable areal unit problem (MAUP),

Generalized additive model, Computer simulation, Cross validation

1. Background

The data collected nowadays are diversified and many of them possess the records of locations, namely spatial data. Spatial regression is a popular tool for analysing the spatial data [2, 35] and the first-order stationarity is a common assumption, which means that the expected (mean) values are fixed at different locations. The error terms of spatial regression are usually not independent and, like in time series analysis, their covariance is assumed to follow some spatial models, such as the simultaneous autoregressive (SAR) and moving average (MA) models [12, 30, 34]. However, the first-order stationarity is a questionable assumption in practice and the modifiable areal unit problem (MAUP) often occurs [5, 13, 22]. The MAUP is a spatial version of the Simpson paradox, where the trends appearing in individual groups of data are different to those in the aggregate data. The biased estimates might be a consequence of the parameter values not being identical in the study area and the inclusion of data with different attributes.

Because the parameter values are not identical at different locations, estimation via the ordinary least squares (OLS) with all observations would likely distort the local distinctness. One possible solution is to include only the locations of data with similar attributes (i.e., homogeneity). However, it is difficult to decide the number of groups with different attributes and identify the locations of data in each group. Moreover, the mean value of a non-stationary process is usually a step function [8] or is continuous across space, and it is difficult to find the exact boundary of appropriate locations. The other possibility is to use the varying coefficient model [10], allowing the coefficient terms to vary according to locations. Then, the model is a form of local linear models [15] and can be used to explore the dynamic property of spatial data. Based on the concept of the varying coefficient model, geographically weighted regression (GWR) is modified to solve the MAUP [6].

The GWR allows the regression coefficients to vary across space, and the coefficient estimates of all variables are obtained from a moving data window, which is analogous to kernel regression for obtaining a smoothing estimate. It is also a popular tool for exploratory data analysis (EDA) on spatial data [19, 32]. In particular, the GWR is often a popular visualization tool in geographical

information system, to explore possible patterns of a study region and acquire valuable information for further data analysis (such as clusters detection) [11, 36]. Note that the optimal width (or bandwidth) of the moving windows in a GWR is determined by cross-validation (CV) or Akaike's information criterion (AIC) [16]. The OLS can be treated as a special case of the GWR with a window of infinite width (although the local distinctness is likely to be lost by averaging all observations).

Most of the modifications to the GWR are on the selection and testing of the bandwidth. For example, using the CV and AIC for the bandwidth selection is a data-driven method, similar to the kernel regression method, wherein the estimates are sensitive to outliers [16]. In addition, the data variations are not necessarily the same and a fixed bandwidth is likely to create discrepancy in parameters' estimates at different locations. On the other hand, the hypothesis testing of the parameters depends on the bandwidth as well. For example, Leung et al. [21] proposed goodness-of-fit tests and found that the degree of freedom of GWR residuals is a function of the bandwidth, and this makes bandwidth selection somewhat subjective.

Determining the bandwidth probably is the focus of modifying GWR over the years. Brunson et al. [7] introduced a mixed GWR model with vector bandwidths, allowing the coefficients having different bandwidths (via a backfitting algorithm) and the bandwidths being functions of data density. Shi et al. [27] suggested the weight of data determined by their attributes, rather than by the distance between observations. Furthermore, Farber and Páez [16] found that it is possible to reduce the bias by modifying the CV procedure. Subsequently, Wang et al. [31] introduced local linear estimation, or a polynomial fitting technique, to reduce the bias in parameters' estimates.

The reason for considering different bandwidth is that the GWR tends to produce ragged surfaces, in addition to biased estimates. Suppose the true surfaces are linear or ridged. As shown in Figure 1, there are (false) hot spots and (false) cold spots in the GWR estimation, and they are especially obvious at edges and corners, where the true values and the estimated surfaces are in the first and second rows, respectively. The bias of GWR becomes larger for ridge surfaces (and other

non-linear surfaces) and the GWR seems to provide misleading interpretations. The detailed discussions of GWR estimates are given later in this manuscript.

[Insert Figure 1 about here]

In this study, our focus is also on the bandwidth selection for each variable, using the correlations between independent variables. The idea of the proposed approach is to use the correlations to improve the estimation via an iteration algorithm, similar to the method of control variate in variance reduction [24]. The empirical analyses of GWR showed that correlations often exist between GWR coefficients. For example, Bivand and Brunstad [4] found the coefficients to be highly correlated in a case study. We found the rough coefficient surfaces can be smoother (Figure 1) if the coefficients are positively correlated.

For the rest of this manuscript, we first introduce the GWR and the proposed modification of GWR, conditional GWR (CGWR), and its theoretical results. Then, we use simulation to evaluate the proposed method and compare it with the basic GWR and the local linear method proposed by Wang et al. [31]. In addition to simulations, we also apply the proposed method to two data sets for empirical study. Finally, we conclude with discussions on the limitations and future applicability of the proposed method.

2. Methods

The GWR models a dependent variable y via a linear function of a set of p independent variables, x_1, x_2, \dots, x_p , or

$$Y_i = \beta_{i0} + \sum_{k=1}^p \beta_{ik} x_{ik} + \varepsilon_i \quad (1)$$

where β_{ik} and x_{ik} are the parameters and observed values of the independent variable k ($k = 1, \dots, p$) for observation i ; ε_i is the error term for observation i , which is generally assumed to be from a

normal distribution with zero mean and constant variance σ^2 (i.e., $\varepsilon_i \sim N(0, \sigma^2)$). The subscript i represents the spatial location of the observation i ($i=1, \dots, n$). In other words, each location has its own regression model in the GWR model. The idea behind Equation (1) is that nearby data of each location usually possess similar attributes. Thus, choosing an appropriate range (which is referred to as “bandwidth” in this study) is plausible to obtain a fine local regression.

The parameter set β_i of observation i is derived by matrix algebra, or

$$\hat{\beta}_i = (\mathbf{X}^T \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_i \mathbf{Y} \quad (2)$$

where $\hat{\beta}_i = (\hat{\beta}_{i0}, \hat{\beta}_{i1}, \dots, \hat{\beta}_{ip})^T$, $\mathbf{X} = (\mathbf{1}, \mathbf{x}_1, \dots, \mathbf{x}_p)^T$, $\mathbf{Y} = (Y_1, \dots, Y_n)^T$, and \mathbf{W}_i is (diagonal) weight matrix with its weight w_{ij} of row i and column j defined as:

$$w_{ij} = \exp \left[-\frac{1}{2} (d_{ij}/h)^2 \right] \quad (3)$$

As mentioned earlier, the bandwidth selection is generally calibrated by minimizing the CV score or the AIC. However, if the data locations are sparse in the study area, the distance-weighted kernel might not be appropriate due to insufficient information. Brunson et al. [8] introduced rank-based and k-nearest neighbourhood methods to deal with sparse data. In addition to the GWR, we also consider one of its modifications by Wang et al. [31]. It is a local linear approach, or a Taylor expansion version of the GWR, and is expected to have better fitting if the geographical surface is linear-shaped.

Using a single bandwidth in GWR is likely to create unsatisfactory estimates if the attributes of independent variables are not similar. For example, independent variables with larger variations require a larger bandwidth (and more observations). **On the other hand, locally sampling (i.e. narrower bandwidth) is preferable over global sampling (i.e. wider bandwidth) for the areas with larger gradient changes. The concept is similar to widely recognized importance sampling [25], which assigns more sampling weight on informative area.** Therefore, allowing different bandwidths for each independent variable seems to be a desirable modification to the GWR. Unfortunately, the varying bandwidths cannot be managed by the weighted least squares or Equation (2). Brunson et

al. [7] proposed a backfitting algorithm for selecting different bandwidths, but the selection of bandwidths is somewhat objective and it usually require lots of computation time.

In this study, we introduce an approach (Conditional GWR; CGWR) to determine the bandwidth for each independent variable by iteration, which is inspired by the vector bandwidth method of Brunson et al. [7] and the kernel smoothing method in the varying-coefficient model of Wu and Chiang [33]. For the proposed method, we adapt the ideas of the Generalized Additive Model (GAM) and Jacobi iteration [17, 20, 23, 28] to determine the appropriate bandwidths. Using the format of the GAM, the GWR model can be re-expressed as

$$Y_i = f_{i1} + \dots + f_{ik} + \dots + f_{ip} + \varepsilon_i \quad (4)$$

where $f_{ik} = \beta_{ik} \times x_{ik}$ and β_{ik} is the parameter coefficient of the variable k at location i . If f_{ik} is the intercept, then x_{ik} is set to 1. Then, we can use the Jacobi iteration to solve the equations (4), one-by-one, for parameter f_{ik} . Again, we assume that $f_k\{l\}$ denotes the l^{th} iteration vector of f_k , and $f_k\{l\}$ denotes an $n \times 1$ vector composed of f_k . Then, the proposed method can be summarized iteratively as follows:

Step 1. Set the initial solution of f_k to be zero, i.e. $f_k\{0\} = \mathbf{0}$, where $k = 1, \dots, p$, and let

$$l = 1.$$

Step 2. For each element $f_k\{l\}$, apply the basic GWR model with only one independent

variable, x_k . The value of the dependent variable is $y^* = y - \sum_{\substack{j=1 \\ j \neq k}}^p f_j\{l-1\}$, i.e. we regress

$\left(y - \sum_{\substack{j=1 \\ j \neq k}}^p f_j\{l-1\} \right)$ on the variable x_k without a fitting intercept. The bandwidth is obtained

by minimizing the cross-validated sum of squares (CVSS) or the AIC.

Step 3. Repeat Step 2 until the given stopping criterion is reached.

There are at least two reasons for finding optimal bandwidth solutions individually using the Jacobi iteration. First, although more complex numerical methods (such as the quasi-Newton

method) could be used, the Jacobi iteration usually requires less computation time. Second, although there are algorithms that converge faster than the Jacobi iteration, they are likely to produce biased estimates. For example, in the Gauss-Seidel iteration process, the estimate of one variable is updated based on the simultaneous estimates of other variables. If the estimates of some variables have severe biases, it might contaminate the estimates of other variables.

We think that the proposed estimation process can guarantee the convergence of the CGWR. In particular, if the bandwidth is predetermined during the iteration, then the GWR coefficients will converge to a constant for each location. We should use the case of two coefficients to demonstrate the convergence and the outline proof is given in Appendix A. Note that the method of Brunson et al. [7] can be treated as a special case of the CGWR method when the bandwidths are never updated. In the next section, we will use computer simulation to evaluate the stability of the CGWR and compare it with basic GWR and its local linear modification by Wang et al. [31].

3. Results and Discussion

3.1. Simulated Data

The computer simulation is separated into two parts: scenarios without clusters and with a cluster. For the latter scenario, a cluster is added into the intercept to exhibit the mean shift intervention. The cluster scenario is to evaluate the performance of estimation methods under the influence of a systematic change (or hot spots) in space, such as sources of pollution. Moreover, the coefficients are assumed to follow one of the following four surfaces: linear, quadratic, ridge, or hillside, and these settings are to check which would cause raggedness in the estimated surfaces. For the former scenario, we also examine two types of surfaces: single-type and mixed-type. The difference between these two types of surfaces is whether the coefficients follow same type of surfaces (single-type) or different types of surfaces (mixed-type). We want to know if the coefficients follow different types of surface would cause biased estimation of coefficients.

To simplify the discussion, suppose there are only two coefficients, i.e. one intercept and one independent variable in the spatial regression, or

$$Y_i = \beta_{i0} + \beta_{i1}x_i + \varepsilon_i, \quad (5)$$

where i is a natural number which indicates the location of the observation. Next, we define the signal vs. noise ratio, i.e. the S/N ratio, where the signal represents the variations on the surface of the coefficients and the noise is the random fluctuation of observations. Larger S/N ratios are associated with larger variations in the coefficient surfaces, in which case the coefficient's pattern is

easier to be detected. In particular, we assume that $\text{signal} = 3 \times \left(\frac{\sum_i (\beta_{ik} - \bar{\beta}_k)^2}{n-1} \right)^{\frac{1}{2}}$ and the noise is

equal to the standard deviation of the error term. Here, it is 0.5. The bias and variance of the estimates can be used together to evaluate the accuracy of the proposed CGWR. We define the average discount rate as follows:

$$\frac{n^{-1} \sum_i (MSE_{method})}{n^{-1} \sum_i (MSE_{OLS})}, \quad (6)$$

where MSE refers to the mean square error, i.e. the sum of variance and the squared bias. Note that the MSE of the OLS estimate is used as a benchmark for comparison in Equation (6) and it can be computed for all locations by using the weighted mean given by:

$$\frac{\sum_i \left[(MSE_{OLS})_{location\ i} \times (MSE_{method} / MSE_{OLS})_{location\ i} \right]}{\sum_i (MSE_{OLS})_{location\ i}}. \quad (7)$$

There are four types of surfaces for the coefficients, as shown in Figure 2. Surfaces 1 and 2 are polynomial functions (related to linear functions) of the independent variables, and surfaces 3 and 4 are non-linear. Similar settings also appear in previous studies on GWR [30], and these have practical implications. For example, the quadratic surface (surface 2) often occurs in situations involving housing prices, where prices are significantly higher for locations near a town centre or a transportation centre [14, 31]. **In addition, the relationship between the environmental factor and real**

estate price can be different in urban areas than in the countryside [9]. Furthermore, the relationship between disease and environmental factors might appear as non-linear on geographic surface. For instance, incidence rate of Dengue disease are highly correlated with population density but the relationship seems to fade out if there is proper disease prevention policy [26]. As a result, the surface of coefficients can be non-linear. For every non-stationary surface, we assume that there are 10×10 regular lattice points (i.e. 100 locations).

For a scenario without clusters, several cases are tested under different S/N ratios. The first case is the single-type surface, where both intercept and independent variable follow the same type of surface. The second case, namely mixed-type surface, assumes that the intercept and the independent variable follow different types of surfaces. For the second case, we only examine two combinations: linear-quadratic (a polynomial surface) and ridge-hillside (a non-polynomial surface).

For the scenario with clusters, we add two clusters in an intercept. The clusters are circular and occupy 18% of space in the study area. The circular assumption is quite usual in geographical studies, and literature proves that 10% ~ 20% of clustered area is a common phenomenon [29]. Two levels of mean shifts, such as 1σ and 2σ , are also added at the cluster locations. The simulation settings of both scenarios are in Table 1. For all scenarios, the errors are drawn from a normal distribution with a mean of 0 and a standard deviation of 0.5. The reason for choosing the standard deviation 0.5 is to incorporate the values of the S/N ratio. All results are based on 100 simulation runs.

[Insert Figure 2 about here]

[Insert Table 1 about here]

For CGWR, the Gaussian kernel is chosen, and the optimal bandwidth is the one with the minimum CVSS. Furthermore, we require a reasonable range of bandwidth to prevent the estimation

from being too localized or globalized for extremely small or large bandwidths, respectively. The upper bound of the range is the maximum length on the map, and the lower bound has to have at least five data points each of 1/5th weight. The preceding setting is also used in the ‘spgwr’ package [3] (version 0.5–4) of R, a free statistical software.

We will first show the simulation results of the scenario without clusters. In particular, we will compare three GWR estimations with the smoothness of the mean surface, the average discount rate, average bandwidth, average variance, and average bias of estimates. The stopping criterion for the CGWR is reached when the average absolute relative change rate of β_0 and β_1 is less than 0.005% of the previous step. The simulation results are similar if adopting smaller stopping criteria.

To simplify the notation, we will use β_0 and β_1 to denote the coefficients of the intercept and slope of the independent variable x , respectively. In the case of single-type surfaces and mixed-type surfaces, the two coefficients are assumed to be perfectly positively correlated and close to uncorrelated, respectively. We have not considered the negative correlation because the proposed algorithm does not work when the coefficients are not positively correlated. Nonetheless, we will consider a two-stage modification for CGWR when the coefficients are not positively correlated.

3.1.1 Single-type Surfaces

We first compare the smoothness of the three different GWR methods. For instance, the mean surfaces from 100 simulation runs for quadratic and hillside surfaces with the S/N ratio of $\beta_0 = \beta_1 = 5$ are shown in Figure 3. Obviously, the CGWR produces the best fit, and the mean surfaces are almost identical to the true surfaces as shown in Figures 2 and 3. The GWR tends to produce bumpy surfaces, which are more ragged for the β_1 surfaces. The edge effect of the GWR is obvious; this may be because fewer observations were used in the estimation. On the other hand, the local linear method tends to produce linear-like surfaces and provides distorted information for the non-linear

surfaces and for the β_1 surfaces. In contrast, the CGWR produces a remarkable fit even in complex surfaces and provides valuable information for further data analysis.

[Insert Figure 3 about here]

Table 2-1 shows the results of the discount rates in the case where β_0 and β_1 follow a linear surface. We can see that both the proposed CGWR and the local linear method have significant improvements over the basic GWR method. Interestingly, the local linear method is better (with respect to smaller discount rates) than the GWR when the S/N ratio is large, but the basic GWR is better when the S/N is small. The reason might be that larger noises produce larger fluctuations, and thus the average tangent line in the local linear method is inaccurate or unstable. Similar results are also found for the other three surfaces, as evidenced in Tables 2-2, 2-3, and 2-4. This suggests that the local linear method might not be very stable if the S/N ratio is small.

The CGWR and the local linear method again outperform the basic GWR method in the case of a quadratic surface. However, the CGWR appears to be the best, and the advantage is more obvious when the S/N is increased. For the non-linear surfaces, the CGWR continues to work satisfactorily, whereas the local linear model does not. In fact, the local linear model might even produce worse results than the basic GWR. The CGWR is still reliable for the non-linear surfaces, and it performs much better than the other two methods.

[Insert Tables 2-1 to 2-4 about here]

[Insert Table 3 about here]

Intuitively, we expect that the bandwidth to be small if the S/N is large because distant observations can be very different and cause biased estimations. In general, all three GWR methods

have significant drops in bandwidth when the S/N ratio increases from one to three. Moreover, the bandwidths for a linear surface should be larger than those for a non-linear surface under the same S/N ratio because the surface change is quite homogenous in any direction.

The bandwidth results can also be used to explain why the CGWR outperforms the other two methods. We will choose two surfaces (linear and hillside) to discuss these results. Table 3 shows the average bandwidths. The local linear method often yields larger bandwidths. If the true surface is close to linear, we can rely on observations within a larger bandwidth and, thus, have smaller variances than those for non-linear surfaces. Since the shape of a hillside is close to linear, the bandwidths in the case of the hillside are very similar to those in the linear case. They are also much larger than those of the quadratic and ridge cases. For more details, [see Appendix B](#).

The bandwidths of the CGWR seem to be related to the signal strength. For example, if the S/N ratio is small, the bandwidth is expected to be large in order to provide a stable estimate. If we fix the S/N ratio of β_1 , then the β_0 bandwidth of the CGWR decreases as the S/N ratio of β_0 increases for all four surfaces (Appendix B). Similar results hold for the β_1 bandwidths if we fix the S/N ratio of β_0 . The simulation results of CGWR match our expectations.

[Insert Tables 4-1 to 4-2 about here]

The variances and biases of the estimates from the three GWR methods can also be used for making comparisons. Again, we will use the cases of linear and ridge surfaces for a detailed discussion. Further, because there are many combinations for the S/N ratios of β_0 and β_1 , we have only shown the results when the S/N ratio equals one and five. The results are shown in Tables 4-1 and 4-2. Unlike the previous comparisons, we also provide the variances and biases of the OLS estimates. In general, a larger S/N ratio tends to produce a larger bias. Moreover, the OLS estimates fail to capture the spatial trend causing the largest bias, but it uses all the observations in the

estimation (i.e. infinite bandwidth) and thus has the smallest variance. As for the three GWR estimations, the variances of the estimators are generally larger than the biases.

The results of the linear surface are in Table 4-1. As mentioned earlier, the average bandwidths of the local linear method are the largest, which possibly indicates the smallest variances. In addition, the local linear method has the smallest bias and the smallest discount rates for linear surfaces (Table 2-1). Although the CGWR has a larger bias than the local linear method in the case of linear surface, it dominates the basic GWR method with respect to both the variance and bias. The CGWR performs the best with ridge surfaces, outperforming the basic GWR and the local linear method with respect to both variance and bias.

3.1.2. Mixed-type Surfaces

Next, we repeat the same comparisons for the three GWR estimation methods with the mixed-type surfaces. The results are similar to those in the single-type surfaces, and thus we will only show the results of the discount rates. As mentioned earlier, there are two cases in this scenario: linear-quadratic (a polynomial surface) and ridge-hillside (a non-polynomial surface). In the first case, the underlying surface of intercept is linear, and the slope is quadratic. In the second case, all surfaces are of non-polynomial type, and it is more complex than the first one.

[Insert Tables 5-1 to 5-2 about here]

Basically, the CGWR also has smaller discount rates than the basic GWR for the mixed-type surfaces (Tables 5-1 and 5-2). We will focus on the results that differ from those of the single-type surfaces. Although the local linear estimation is better than the GWR for linear-quadratic surfaces, it performs adversely for the ridge-hillside surfaces. It is inadequate to use the linear fitting method to approximate non-linear surfaces, such as the ridge-hillside case. On the contrary, (similar to the single-type cases) the CGWR dominates the other two methods in both cases.

3.1.3. Two-stage Fitting Procedure

We found that the CGWR works well when there is a positive correlation. However, in reality, there is a great likelihood of variables not being positively correlated. To overcome this difficulty, the CGWR can be modified into a two-stage process. In the first stage, we divide the variables into two groups. In both groups, the variables are non-negatively (or positively) correlated within the group. Any two variables are non-positively (or negatively) correlated if they are from different groups. We choose one group of variables and apply the basic GWR method to this group. In the second stage, we apply the CGWR method to the other group of variables by treating the first group of variables (chosen in the first stage) as constants.

We use an example to demonstrate the two-stage fitting. Let us assume there are two independent variables and an intercept. Let the coefficients of the two independent variables be negatively correlated. In other words, let the coefficients of variables x_1 and x_2 be negatively correlated, and the coefficients between x_1 and intercepts be positively correlated. We first apply the basic GWR on x_2 in the first stage, and then apply the CGWR on the intercept and x_1 in the second stage. We will use a simulation to evaluate the two-stage modification and show the results of the linear and hillside surfaces in Figure 4. Similar to the previous simulation, the two-stage CGWR seems to work well even when the variables are not all positively correlated.

[Insert Figure 4 about here]

[Insert Figure 5 about here]

3.1.4. Single-type Surfaces with clusters

The purpose of considering the scenario with clusters is to investigate whether the estimated surface would be influenced by the cluster intervention on β_0 . Figure 5 illustrates the cluster

location and the mean shift level. Single-type surfaces are assumed to evaluate the performance under cluster intervention. The mean smoothness and average discount rate of β_0 are in Figure 6 and Table 6. The CGWR again has the best performance and provides the most accurate information pertaining to the location and size of the clusters. Although the GWR seems to reveal the true cluster locations, it suffers from bumpy fitting and gives rise to ‘false clusters.’ The local linear method seems to oversmooth the surface and blur the local pattern, although this might suggest a possible cluster on the edges.

[Insert Figure 6 about here]

[Insert Table 6 about here]

From the preceding computer simulations studies, we found that the proposed CGWR method makes a significant improvement over the basic GWR method. Although the local-linear method behaves well in the linear surface, if the coefficient surfaces are non-linear, the CGWR also outperforms the local linear method. In the following discussion, we will use two real-world datasets to compare the CGWR and other two methods and provide further evidence in support of the CGWR.

3.2. Empirical Data

We apply the CGWR to two empirical data sets: the first is from the 2000 Taiwan Census and the other is the Ohio crime data provided by Anselin [1]. These two examples are designed to demonstrate that the CGWR would yield better estimation results. For the Taiwan data, our goal is to explore the relationship between the proportion of elderly disability and social factors. The elderly population in Taiwan have been increasing rapidly all around the country, while the medical resources still concentrate in the metropolitan areas (or northern Taiwan). Hu and Yue [18] applied

spatial regression model to the elderly disability data of township level and found they are spatially auto-correlated. Brunson [7] argued that the spatial autocorrelation seems to be caused by spatial non-stationarity (i.e., identifiability). His claim motivates us to re-examine the data using the GWR-based model.

3.2.1. Taiwan Data

Taiwan 2000 Census includes data of 350 townships and their proportions of disabled elderly are set as the dependent variable. Since this variable appears to be right-skewed, a log transformation (i.e. $y_i^* = \log(y_i + 1)$) is applied. Four independent variables are selected: the population density (POP), proportion of elderly (ELD), elderly mortality rate (EMR), and education level (EDU). These independent variables are standardized into the [0, 1] interval. Before applying the GWR, we first test spatial non-stationarity with the F-test suggested by Leung et al., [21]. The F-test shows that the model is spatially non-stationary with p-value < 0.001. This confirms the hypothesis of Brunson [7] and creates incentive for plugging the GWR-type analysis.

[Insert Table 7 about here]

The correlation of the intercept and the variable POP is 0.463 (Table 7), and they are placed in the same group. Similarly, the variables ELD, EMR, and EDU are in the other group since they are positively correlated pairwise. Thus, we use the two-stage modification and apply the CGWR on the group of positively correlated variables (i.e. the intercept and the variable POP). First, we treat the variables ELD, EMR, and EDU as constants after obtaining estimates from the basic GWR method. Then, we apply the CGWR to the intercept and the variable POP as $(y_i^* - \hat{\beta}_{i2}^{GWR} ELD_i - \hat{\beta}_{i3}^{GWR} HMR_i - \hat{\beta}_{i4}^{GWR} EDU_i) = \hat{\beta}_{i0}^{CGWR} + \hat{\beta}_{i1}^{CGWR} POP_i + r_i$. After fitting the CGWR, the calibrated bandwidths vary across the variables. Also, we set the lower and upper bounds of the bandwidth as 1 km and 400 km, respectively.

There is a noticeable difference between the estimates from the CGWR and those from other methods (Figure 7). The coefficient surfaces of the local linear method appear to spread in the north-south or east-west direction with linear boundaries. Similarly, the surfaces of the basic GWR also show descending (or ascending) patterns but with curved boundaries. However, the intercept of the CGWR shows clusters (or concentration) of high disability rate on inland (mountainous areas). For the variable POP, the number of coefficient levels varies among different models and their spreading directions are not identical. The GWR has the fewest levels and the local linear method has the largest; the spreading is in the east-west direction for the local linear method, different to other methods.

We also use the pseudo R-square values and the residual plots for model evaluation (Figure 8). The pseudo R-square is the Pearson product moment correlation coefficient of the fitted value and the observed value; a large value usually indicates a better fit. The pseudo R-square value of the CGWR is 0.894, largest among three methods. Moreover, the residual plots are also in favour of the CGWR because there are fewer outliers, and the CGWR appears to have smaller variance. Except for one observation (standardized residual larger than 3), **the residuals histogram of CGWR (350 observations) looks more symmetric and less skewed to the right than those of the basic GWR and local linear method.** It should be noted that either one of the variable sets can be chosen as a constant. If we apply the CGWR procedure to the other group of variables (i.e. ELD, EMR, and EDU), then the CGWR has a better fit (although the pseudo R-square is slightly low at 0.874).

[Insert Figure 7 about here]

[Insert Figure 8 about here]

3.2.2. Ohio Data

The Ohio data is the Ohio crime data (which can be found on 'spgwr' package) with the information of 49 neighborhoods including the crime per inhabitant, average income values, and

average housing costs. In this study, we define the crime per inhabitant as the dependent variable and the rest of them as the predictors. First, we fit the data with the GWR model. By cross validation criterion, the optimal bandwidth is 2.27 (Table 8). And yet, Leung et al. **F-test** [21] suggests none of the variable is non-stationary. Therefore, the OLS analysis is applied and one observation is considered as outlier and removed accordingly.

Next, we re-fit the data with the CGWR and compare the estimation result with those of the OLS, the GWR, and the local linear models. Table 9 lists the pseudo R-square and the p-value of normality test (Kolmogorov-Smirnov test) for residuals, and Figure 9 shows the residuals plot. Overall, the CGWR yields the best performance in estimation and produces a more reliable result. For other methods, none of them gives satisfactory estimates. For example, despite the GWR produces a large pseudo R-square, its residuals are not normally distributed and its variance is likely not constant. The OLS is also not a feasible model, judging from the information of normality test and constant variance.

[Insert Table 8 about here]

[Insert Table 9 about here]

[Insert Figure 9 about here]

4. Conclusions

The GWR has become a popular tool for explanatory data analysis and detecting spatial non-stationarity ever since its introduction. The GWR provides useful information for data analysis, especially helpful in deciding important explanatory variables. This technique allows regression coefficients to vary across space and obtains their estimates from a bandwidth of observations according the data attribute. However, the GWR tends to produce ragged surfaces (as shown in Figure 1), and a fixed bandwidth may not be appropriate since the independent variables are necessary to be homogeneous (e.g., their variations can be quite different). In this study, we proposed

a modification to the GWR, namely CGWR, which allows the group of positively correlated independent variables to have its own bandwidth via an iterative calibration process.

We used computer simulated and empirical data to compare the proposed method with the GWR and its local linear modification by Wang et al. [31]. Based on the simulation results, we found that the CGWR outperforms other two methods, with respect to the bias and variance, when the regression coefficients are positively correlated. The advantage is especially noticeable in the case of non-linear surfaces. In particular, the clusters have little influences on the estimation of CGWR. The results of empirical studies also support the CGWR and it generally has larger R-square and has fewer extreme outliers (e.g., the absolute value of standardized residual larger than 2 or 3) than the GWR and the local linear method.

However, the proposed method has its limitations. First, probably the most critical limitation, the current setting of CGWR only works for the case if there are independent variables with positive correlation. Although not shown here, we found that the CGWR does not work well in the case of independent variables with negative correlation. It is like the antithetic variate in variance reduction of Monte Carlo Integration. Antithetic variate is one of the popular variance reduction methods but it only works when two variables are negatively correlated [25]. Thus, we suggest first calculating the correlation coefficients between independent variables. Then, form a group of variables which are pairwise positively correlated and apply the CGWR only to this group of variables. Another possibility is that the independent variables often can be separated into two groups and variables within/between groups are positively/negatively correlated, as seen in Taiwan 2000 Census data. We can apply the two-stage CGWR to two groups of variables.

Second, the CGWR is a computer-intensive method and its computing time increases rapidly as the number of variables increases, although the convergence of coefficients can be speed up by using the moving average method. Third, the CGWR is not guaranteed to work if there are many variables, and so far it is effective for the case up to four variables. A possible modification to the case with more variables would be to separate the variables into two groups and use double iteration. Then, the CGWR can be applied to each group of variables forming the inner loop and the process

re-iterated between the two groups forming the outer loop until both groups of variables converge. To demonstrate the feasibility of this idea, we also conducted an experiment with six variables, separating them into two groups of three variables each. We found that the estimation did converge and produce satisfactory estimates.

In addition to the fixed bandwidth, it seems that there is still room for improvement about the GWR. In particular, when the S/N ratio is small, the estimated coefficient surfaces would be non-linear (i.e., ragged surfaces), even when the true surfaces are linear. In addition, the variance reduction of the CGWR over the GWR is more obvious than that for bias reduction. This indicates that the GWR estimates have large variance when the S/N ratio is small. In other words, if the variances of GWR estimates are reduced, the bias can also be further reduced, producing more stable estimates.

Author's contributions

Leong developed the estimation algorithm, conducted simulation studies, as well as empirical data analysis, and drafted the manuscript. Yue proposed the general methodology, designed the simulation, interpreted the results and helped to draft the manuscript. All authors read and approved the final manuscript.

Availability of data and materials

The Columbus Ohio crime data is prepared by Luc Anselin, Spatial Analysis Laboratory, Department of Agricultural and Consumer Economics, University of Illinois, Urbana-Champaign. The dataset can be downloaded at following link: <https://spatial.uchicago.edu/sample-data>, or available at the 'spgwr' package in R statistical software.

References

1. Anselin L. *Spatial Econometrics: Methods and Models*. Dordrecht, The Netherlands: Kluwer Academic Publishers; 1988.
2. Auchincloss AH, Gebreab SY, Mair C, Diez Roux AV. A Review of Spatial Methods in Epidemiology, 2000–2010. *Annual Review of Public Health*. 2012; 33: 107–122.
doi: 10.1146/annurev-publhealth-031811-124655
3. Bivand R. and Danlin Y. R Manual for Package ‘spgwr,’. 2010.
<http://cran.r-project.org/web/packages/spgwr/spgwr.pdf> Accessed 01 Mar 2017.
4. Bivand R, Brunstad R. Further Explorations of Interactions between Agricultural Policy and Regional Growth in Western Europe: Approaches to Nonstationarity in Spatial Econometrics. 2005. http://www.feweb.vu.nl/ersa2005/final_papers/671.pdf
Accessed 01 March 2017
5. Bratti M, De Benedictis L, Santoni G. On the Pro-trade Effects of Immigrants. *Review of World Economics/Weltwirtschaftliches Archiv*. 2014; 150(3), 557–594.
doi:10.1007/s10290-014-0191-8
6. Brunsdon C, Fotheringham AS, Charlton M. Geographically Weighted Regression: Modelling Spatial Nonstationarity, *The Statistician*, 1998; 47: 431–443.
doi: 10.1111/1467-9884.00145
7. Brunsdon C, Fotheringham AS, and Charlton M. Some Notes on Parametric Significance Tests for Geographically Weighted Regression. *Journal of Regional Science*. 1999; 39: 497–524.
doi: 10.1111/0022-4146.00146
8. Brunsdon C, Fotheringham AS, and Charlton M. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. UK: Wiley & Sons; 2002.
9. Cho S, Lambert DM, Kim SG, Jung S. Extreme Coefficients in Geographically Weighted Regression and Their Effects on Mapping. *GIScience and Remote Sensing*. 2009; 46(3): 273–

288.

doi: 10.2747/1548-1603.46.3.273

10. Cleveland WS, Grosse E, Shyu WM. Local Regression Models. in: JM Chambers and TJ Hastie (eds.), *Statistical Models in S*. Pacific Grove: Wadsworth & Brooks; 1991: 309–376.
11. Comber A, Fisher P, Brunson C, Khmag A. Spatial Analysis of Remote Sensing Image Classification Accuracy. *Remote Sensing of Environment*. 2012; 127: 237–246.
doi: 10.1016/j.rse.2012.09.005
12. Cressie N. *Statistics for Spatial Data*. New York: Wiley & Sons; 1993
13. Dark, SJ, Bram D. The Modifiable Areal Unit Problem (MAUP) in Physical Geography. *Progress in Physical Geography*. 2007; 31: 471–479.
doi: 10.1177/0309133307083294
14. Du H, Mulley C. Understanding Spatial Variations in The Impact of Accessibility on Land Value Using Geographically Weighted Regression. *The Journal of Transport and Land Use*. 2012; 5(2): 46–59. doi: 10.5198/jtlu.v5i2.225
15. Fan JQ, Zhang WY. Statistical Methods with Varying Coefficient Models. *Statistics and its Interface*. 2008; 1: 179–195.
16. Farber S, Páez A. A Systematic Investigation of Cross-validation in GWR Model Estimation: Empirical Analysis and Monte Carlo Simulations. *Journal of Geographical Systems*. 2007; 9: 371–396. doi: 10.1007/s10109-007-0051-3
17. Hastie TJ, Tibshirani RJ. *Generalized Additive Models*. London, New York: Chapman and Hall; 1990.
18. Hu YW, Yue JC. *Spatial Analysis of Taiwan Elderly Disability*. Technical Report, Department of Statistics, National Chengchi University, Taipei, Taiwan R.O.C ; 2002.
<http://csyue.nccu.edu.tw/ch/2000Taiwan%20Old%20Age%20Census.pdf>
Accessed 01 March 2017

19. Kauh B, Schweikart J, Krafft T, Keste A, Moskwyn M. Do the Risk Factors for Type 2 Diabetes Mellitus Vary by Location? A Spatial Analysis of Health Insurance Claims in Northeastern Germany Using Kernel Density Estimation and Geographically Weighted Regression. *International Journal of Health Geographics*. 2016; 15: 68.
doi: 10.1186/s12942-016-0068-2
20. Lou Y, Caruana R, Gehrke J. Intelligent Models for Classification and Regression. in *Proceedings of the 18th ACM SIGKDD international conference on KDD*. New York: ACM, 2012; 150–158.
21. Leung Y, Mei CL, Zhang WX. Statistical Tests for Spatial Nonstationarity based on the Geographically Weighted Regression Model. *Environment and Planning A*. 2000; 32: 9–32.
doi: 10.1111/j.1538-4632.1996.tb00936.x
22. Ma YZ. Simpson's Paradox in Natural Resource Evaluation. *Mathematical Geosciences*. 2009; 41(2): 193–213.
23. Manoranjan VS, Olmos GM. A Two-Step Jacobi-Type Iterative Method. *Computers Math. Application* 1997; 34, 1–9.
doi: 10.1016/S0898-1221(97)00093-X
24. Ripley BD. *Stochastic Simulation*. Wiley Interscience Paperback Series; 1987.
25. Rubinstein RY, Kroese DP. *Simulation and the Monte Carlo Method*. John Wiley & Sons; 2011.
26. Seng SB, Chong AK, Moore A. Geostatistical Modelling, Analysis and Mapping of Epidemiology of Dengue Fever in Johor State, Malaysia, (pp. 109–123). Presented at the 17th Annual Colloquium of the Spatial Information Research Centre (SIRC 2005: A Spatio-temporal Workshop); 2005.
27. Shi HJ, Zhang LJ, Liu JG. A New Spatial-Attribute Weighting Function for Geographically Weighted Regression. *Canadian Journal of Forest Research*. 2006; 36: 996–1005.
28. Thisted R. *Elements of Statistical Computing*. Chapman-Hall, London; 1988.

29. Tango T, Takahashi K. A Flexibly Shaped Spatial Scan Statistic for Detecting Clusters. *International Journal of Health Geographics*. 2005; 4: 11.
doi : 10.1186/1476-072X-4-11
30. Ward MD, Gleditsch KS. *Spatial Regression Models*. Thousand Oaks, CA: Sage; 2008.
31. Wang N, Mei CL, Yan XD. Local Linear Estimation of Spatially Varying Coefficient Models: An Improvement on the Geographically Weighted Regression Technique. *Environment and Planning A*. 2008; 40: 986–1005.
32. Wheeler DC. Geographically Weighted Regression. in: MM Fischer and P Nijkamp, (eds.), *Handbook of Regional Science*. Springer, Berlin, Heidelberg; 2014: 1435–1459.
33. Wu CO, Chiang CT. Kernel Smoothing on Varying Coefficient Models with Longitudinal Dependent Variable. *Statistica Sinica*. 2000; 10: 433–456.
34. Wang TC, Yue JC. Spatial Clusters in a Global-dependence Model. *Spatial and Spatio-temporal Epidemiology*. 2013; 5: 39–50.
35. Wei Y, Ye X. Urbanization, Urban Land Expansion and Environmental Change in China. *Stochastic Environmental Research and Risk Assessment*. 2014; 28: 757–765.
36. Zou B, Pu Q, Muhammad B, Weng Q, Zhai L, Nichol J. High-resolution Satellite Mapping of Fine Particulates based on geographically weighted regression. *IEEE Geosci. Remote Sens. Lett*. 2016; 13: 495–499.
doi: 10.1109/LGRS.2016.2520480

Figures

Figure 1. True Surfaces and GWR Mean Surface

The figure compares the true surfaces and the estimated GWR respond surface. (Linear and Ridge Surface)

Figure 2. Four Coefficient Surfaces

The figure illustrates different kinds of surfaces in simulation settings. (1 & 2 are Polynomial, and 3 & 4 are Non-polynomial)

Figure 3-1. β_0 Mean Surface

Estimated intercept respond surface derived from different method. (Quadratic and Hillside Surfaces, S/N ratio of $\beta_0 = \beta_1 = 5$).

Figure 3-2. β_1 Mean Surface

Estimated intercept respond surface derived from different method. (Quadratic and Hillside Surfaces, S/N ratio of $\beta_0 = \beta_1 = 5$).

Figure 4. Average Discount Rates by x_2

The average discount rates of additional explanatory variables during the simulation study. (Note: The baseline is GWR.)

Figure 5. Cluster Location and Mean Shift Level.

The area shows cluster intervention in simulation study. The shade depth in the first graph represents the mean shift level.

Figure 6. β_0 Mean Surface in Cluster Settings

Estimated intercept response surface derived from different method. (Quadratic and Hillside Surfaces, S/N ratio of $\beta_0 = \beta_1 = 5$).

Figure 7. Surface of Intercept and Variable POP for Different GWR Methods

Comparison of different method by using Taiwan disability data

Figure 8. Residual Plots for Different GWR Methods (Taiwan Data)

These plots illustrate the residual analysis after fitting with different model. The residuals are generated from Taiwan disability data.

Figure 9. Residual Plots of Various Models (Ohio Data)

Different residual plot from Ohio crime data. Note that the data are cut into half according to fitted values before conducting F-test.