INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE, 2017
VOL. 31, NO. 5, 982–998
http://dx.doi.org/10.1080/13658816.2016.1263731

Geographically weighted regression with parameter-specific distance metrics

Binbin Lu [a,b], Chris Brunsdon [c], Martin Charlton[c] and Paul Harris[d]

[a]School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China; [b]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China; [c]National Centre for Geocomputation, National University of Ireland Maynooth, Maynooth, Ireland; [d]Sustainable Soils and Grassland Systems, Rothamsted Research, North Wyke, Okehampton, UK

ABSTRACT

Geographically weighted regression (GWR) is an important local technique to model spatially varying relationships. A single distance metric (Euclidean or non-Euclidean) is generally used to calibrate a standard GWR model. However, variations in spatial relationships within a GWR model might also vary in intensity with respect to location and direction. This assertion has led to extensions of the standard GWR model to mixed (or semiparametric) GWR and to flexible bandwidth GWR models. In this article, we present a strongly related extension in fitting a GWR model with parameter-specific distance metrics (PSDM GWR). As with mixed and flexible bandwidth GWR models, a back-fitting algorithm is used for the calibration of the PSDM GWR model. The value of this new GWR model is demonstrated using a London house price data set as a case study. The results indicate that the PSDM GWR model can clearly improve the model calibration in terms of both goodness of fit and prediction accuracy, in contrast to the model fits when only one metric is singly used. Moreover, the PSDM GWR model provides added value in understanding how a regression model's relationships may vary at different spatial scales, according to the bandwidths and distance metrics selected. PSDM GWR deals with spatial heterogeneities in data relationships in a general way, although questions remain on its model diagnostics, distance metric specification, and computational efficiency, providing options for further research.

ARTICLE HISTORY
Received 23 June 2016
Accepted 16 November 2016

KEYWORDS
GWR; GWmodel; local regression; spatial heterogeneity; model anisotropy

1. Introduction

On consideration of Waldo Tober's first law of geography (Tobler 1970), Goodchild (2004) suggested a candidate second law of geography in the principle of spatial heterogeneity or nonstationarity. In this respect, there has been much interest in localized regression methods for spatial data analysis that produce spatially varying regression outputs instead of a 'one-size-fits-all' result of the usual global regression (Fotheringham and Brunsdon 1999). Notable localized regression techniques include the expansion method (Casetti 1972), the weighted spatial adaptive filtering model (Gorr

CONTACT Binbin Lu binbinlu@whu.edu.cn

and Olligschlaeger 1994), the geographically weighted regression (GWR) model (Brunsdon *et al*. 1996) and more recently, the Bayesian space-varying coefficient (SVC) model (Assunção 2003, Gelfand *et al*. 2003). Only the latter two still have merit, where the simplicity of GWR provides a valuable alternative to the sophisticated SVC model, whose inherent computational complexity tends to severely limit its widespread application. GWR has been recognized as an important and popular technique for modeling spatial heterogeneous processes across a wide range of domains (Fotheringham *et al*. 2015). Griffith (2008) provides a valuable critique of GWR with respect to its relationship to autoregressive and spatial filtering models.

The remit of a GWR model is to explore spatially varying relationships between the dependent and independent variables via a series of localized linear regression fits. Here at each local regression calibration point, a 'bump of influence' is produced where nearer observations have more influence in estimating the local set of regression parameters than do observations farther away (Fotheringham *et al*. 2002). This is achieved via some distance-decay kernel weighting scheme. Notably, research has tried to adapt or extend this weighting scheme in order to refine the GWR calibration and improve the associated output, specifically in terms of bandwidth and distance metric choices. In this respect, Farber and Páez (2007) propose two modified cross-validation (CV) approaches for optimal bandwidth selection that reduce the influence of outlying CV values. Brunsdon *et al*. (1999) introduce mixed GWR, that treats some dependent to independent data relationships as global (or fixed), while the rest as local (i.e. the usual case, but as in basic GWR, each at the same spatial scale). Yang (2014) extends mixed GWR to GWR with flexible bandwidths (FB GWR) that enables each data relationship to operate at its own (and commonly different) spatial scale via specifying its own relationship-specific bandwidth. Further refinements in the weighting scheme have been necessary in the GWR models of Huang *et al*. (2010) and Fotheringham *et al*. (2015) where the temporal dimension is incorporated; and of Harris *et al*. (2013), where hierarchical data structures are represented. Lu *et al*. (2011, 2014a) use non-Euclidean distance metrics in GWR, and found its fit could be improved by using a proper distance metric (i.e. network distance (ND) and travel time (TT)), instead of the usual Euclidean distance (ED). Further refinements in distance metric selection can be found in Lu *et al*. (2016), where a Minkowski approach is used to approximate the underlying 'optimum' metric. Páez (2004) provides an anisotropic version of GWR, allowing dependent/independent variable local relationships to vary in intensity with direction.

All such studies endorse a key principle in GWR in that its chosen distance-weighting scheme is crucial to its performance, and thus research to refine or improve this distance-weighting scheme is worthy. In this respect, we propose that the relationship between any specific dependent/independent variable pair adheres to its own specific spatial process and as such, should have its own, distinctive weighting computation. We do this by presenting a GWR model with parameter-specific distance metrics (PSDM GWR), extending the preliminary simulation work of this model from Lu *et al*. (2015), where a PSDM GWR model provided more accurate predictions and more accurate coefficient estimates, than those from a standard GWR calibration. The PSDM GWR model can be considered as an amalgamation of the non-ED metric GWR models of Lu *et al*. (2011), 2014a) with the FB GWR model of Yang (2014).

This article is organized as follows. Firstly, we introduce the methodology of PSDM GWR. Secondly, we present a London house price case study using PSDM GWR with ED and TT metrics. Thirdly, we investigate ways to improve the computational efficiency of the PSDM GWR back-fitting algorithm. Finally, we summarize and discuss future refinements to the methodology.

2. Methodology

2.1. *Geographically weighted regression*

GWR makes location-wise estimates to model spatially varying relationships. Generally, a basic form of GWR model can be expressed as:

$$y_i = \beta_{i0} + \sum_{k=1}^{m} \beta_{ik} x_{ik} + \varepsilon_i, \tag{1}$$

where y_i and $x_{ik}(k = 1, \cdots, m)$ are the observations of dependent variable and independent variable, respectively, at location i, $\beta_{ik}(k = 0, 1, \cdots, m)$ is the set of regression parameters estimated at location i; and ε_i is the random error term.

A standard GWR model is calibrated by a weighted least squares approach at each regression point, of which the matrix expression is:

$$\hat{\boldsymbol{\beta}}_{\mathbf{i}} = \left(\mathbf{X}^{\mathbf{T}}\mathbf{W}_{\mathbf{i}}\mathbf{X}\right)^{-\mathbf{1}}\mathbf{X}^{\mathbf{T}}\mathbf{W}_{\mathbf{i}}\mathbf{y}, \tag{2}$$

where **X** is the matrix of the independent variables with *m + 1* columns and a column of *1* s for the intercept (if there is one); **y** is the vector of the dependent variable; and $\mathbf{W_i}$ is a diagonal matrix denoting the geographical weightings for each observation data (sub)set at regression location *i*. Notably, $\mathbf{W_i}$ is calculated with a distance-decay kernel function, which is nonincreasing, real, and bounded from 0 to 1 (Cho *et al*. 2010). There are many kernel functions to choose from, for example, Gaussian, exponential, bi-square, tri-cube, and box-car (see Gollini *et al*. 2015). Here, the Gaussian kernel function is used and can be expressed as,

$$\text{Gaussian}: w_{ij} = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right], \tag{3}$$

where w_{ij} is the weight attributed to observation *j; d_{ij}* is the distance between observation *j* and regression point *i*; and *b* is the bandwidth, a key parameter to control the magnitude of distance-decay. The bandwidth can be either a fixed distance (i.e. a fixed distance bandwidth) or a fixed number of nearest neighbors (i.e. an adaptive distance bandwidth). It can be optimally found by minimizing the CV score (Cleveland 1979, Bowman 1984), or the Akaike information criterion (AIC) (Akaike 1973). AIC approaches are preferred as they account for model parsimony, that is, a trade-off between prediction accuracy and complexity. In particular, a corrected version of the AIC (AICc) (Hurvich *et al*. 1998) is used in this study, whose calculation is:

$$\text{AIC}_c = 2n\ln(\hat{\sigma}) + n\ln(2\pi) + n\left\{\frac{n + \text{tr}(\mathbf{S})}{n - 2 - \text{tr}(\mathbf{S})}\right\}, \tag{4}$$

where n is the number of observations; $\hat{\sigma}$ denotes the estimated standard deviation of the error term; and $\mathrm{tr}(\mathbf{S})$ denotes the trace of the hat matrix $\mathbf{S}$. The hat matrix is the projection matrix from the observed $\mathbf{y}$ to the fitted values $\widehat{\mathbf{y}}$ (Hoaglin and Welsch 1978). In a GWR calibration, each row of $\mathbf{S}$ can be expressed as:

$$r_i = X_i\left(X^T W(u_i, v_i)X\right)^{-1} X^T W(u_i, v_i), \tag{5}$$

where X_i is its ith row of the matrix of independent variables $\mathbf{X}$.

2.2. *GWR with parameter-specific distance metrics*

In the standard GWR technique, the ED metric is uniformly used when determining the geographically weighting for dependent/independent variable relationships. Even with non-ED metrics (Lu *et al*. 2014a, 2016), the proximities between the observations and each regression calibration location are similarly calculated in the same way, no matter how many different independent variable types are included in the regression model. Meanwhile, the key to control the spatial scale or magnitude of nonstationarities is the bandwidth: small bandwidths tend to reveal spatial pattern at a detailed microscopic scale, while large bandwidths are inclined to present spatial variations at a broad macroscopic scale (Fotheringham *et al*. 2002). However, it is likely that the scale or the intensity of the spatial relationships may differ among each dependent/independent variable relationship, and as such, each should have diverse responses to the weighting computation, even within the same regression model (Lu *et al*. 2015). The first GWR model to consider this was proposed by Brunsdon *et al*. (1999) with mixed GWR; a model that allows each regression relationship to be treated either as local or as global. This mixed or semiparametric GWR model has subsequently been refined by Mei *et al*. (2004, 2016) and by Nakaya *et al*. (2005). In the mixed GWR model, the bandwidth for all local relationships is taken to be the same, while the 'bandwidth' for all global relationships is also the same. The natural extension of the mixed GWR model is to allow each relationship to have its own specific bandwidth, that is, the flexible bandwidths of the FB GWR model proposed by Yang (2014) (see also, Yang *et al*., (2011, 2012)). The PSDM GWR of this study is similar in spirit to the FB GWR model, whereas FB GWR only used ED metrics, the new PSDM GWR model allows each relationship to have its own specifically 'optimized' distance metric for each regression relationship (i.e. each independent variable and the intercept).

Both mixed GWR and FB GWR employ back-fitting algorithms (Hastie and Tibshirani 1986), so it is natural to adopt a similar algorithm here. In particular, we extend the algorithm presented in the preliminary work on the PSDM GWR model (Lu *et al*. 2015), to now adopt varying bandwidths with correspondence to parameter-specific distance metrics. Suppose we have calculated distance matrices, $DM_0, DM_1, \cdots, DM_m$, and corresponding bandwidths $bw_0, bw_1, \cdots, bw_m$, specifically for each independent variable in the model (Equation (1)). The PSDM GWR model can be calibrated via a back-fitting procedure in the following steps:

(1) Make an initial guess of the coefficients $\widehat{\boldsymbol{\beta}}^{(0)} = \left(\widehat{\boldsymbol{\beta}}_0^{(0)}, \widehat{\boldsymbol{\beta}}_1^{(0)}, \cdots, \widehat{\boldsymbol{\beta}}_{\mathbf{m}}^{(0)}\right)$, calculate all the estimated terms $\widehat{\mathbf{y}}_0^{(0)} = \widehat{\boldsymbol{\beta}}_0^{(0)} \cdot \mathbf{X}_0, \cdots, \widehat{\mathbf{y}}_m^{(0)} = \widehat{\boldsymbol{\beta}}_{\mathbf{m}}^{(0)} \cdot \mathbf{X}_m$, where $\mathbf{X}_i$ is the (i + 1)th

column of $\mathbf{X}(i = 0, 1, \cdots, m)$, where the symbol $\cdot$ means the element-wise multiplication of the two vectors, and the residual sum of squares, $RSS^{(0)}$;

(2) Provide the maximum number of iterations $\mathbb{N}$ and the threshold value τ as criterions to terminate the back-fitting iterations;

(3) To start the back-fitting iterations, initialize the iteration-count number $k = 1$;

(4) For each independent variable $x_l(l = 0, 1, \cdots, m)$, do the following operations:

(a) Calculate $\xi_l^{(k)} = \mathbf{y} - \sum_{j \neq l}^{m} Latestyhat\left(\widehat{\mathbf{y}}_j^{(k-1)}, \widehat{\mathbf{y}}_j^{(k)}\right)$, where $Latestyhat$ is a conditional function:

$$Latestyhat\left(\widehat{\mathbf{y}}_j^{(k-1)}, \widehat{\mathbf{y}}_j^{(k)}\right) = \begin{cases} \widehat{\mathbf{y}}_j^{(k)}, & if\ \widehat{\mathbf{y}}_j^{(k)} exists \\ \widehat{\mathbf{y}}_j^{(k-1)}, & otherwise \end{cases}. \tag{6}$$

(b) Do the weighted regression between $\xi_l^{(k)}$ and $\mathbf{X}_l$ (see Equation (2)) and get a new set of coefficients $\widehat{\boldsymbol{\beta}}_{\mathbf{l}}^{(k)}$, where the weighting matrix is computed with the corresponding distance matrix DM_l and bandwidth bw_l;

(c) Update the estimated term $\widehat{\mathbf{y}}_l^{(k)} = \widehat{\boldsymbol{\beta}}_{\mathbf{l}}^{(k)} \cdot \mathbf{X}_l$;

(5) Calculate the predicted value $\widehat{\mathbf{y}}^{(k)}$ with the newly estimated coefficients $\widehat{\boldsymbol{\beta}}^{(\mathbf{k})} = \left(\widehat{\boldsymbol{\beta}}_0^{(\mathbf{k})}, \widehat{\boldsymbol{\beta}}_1^{(\mathbf{k})}, \cdots, \widehat{\boldsymbol{\beta}}_{\mathbf{m}}^{(\mathbf{k})}\right)$, and then update the residual sum of squares (RSS), $RSS^{(k)}$

(6) Calculate the changing value of RSS (CVR), as,

$$\mathrm{CVR}^{(k)} = RSS^{(k)} - RSS^{(k-1)} \tag{7}$$

or a differential version,

$$\mathrm{CVR}^{(k)} = \frac{RSS^{(k)} - RSS^{(k-1)}}{RSS^{(k-1)}}. \tag{8}$$

(7) Update the iteration-count number k = k + 1;

(8) If $\mathrm{CVR}^{(k)}$ is larger than τ or the number of iterations k exceeds $\mathbb{N}$, then terminate; otherwise, go to step 4.

In the algorithm above, we assume that all the bandwidths have been provided. However, they are usually unknown for a PSDM GWR model in practice. For the FB GWR model, Yang (2014) suggests two strategies to select parameter-specific bandwidths: (I) make brute-force searches from a broad collection of possible bandwidth values and choose the best performing set of bandwidths; (II) optimize the bandwidth for each independent variable within the back-fitting iterations, that is, select an optimum bw_l in step (4b) of the above algorithm. The former strategy is extremely computationally intensive, and its accuracy will rely on the candidate bandwidth values provided. Therefore, the latter strategy is adopted for selecting the multiple bandwidths here.

The back-fitting algorithm can still be computationally demanding however, even if the bandwidths are provided or known. The computational costs largely depend on the speed of convergence, that is, the eventual number of iterations. This depends on the following elements: (i) the initial guess of $\widehat{\boldsymbol{\beta}}^{(0)}$; (ii) bandwidths given or not given (if not given, the bandwidth in each iteration should be selected via a CV or AICc approach); (iii) the choice of CVR, and generally the differential version is recommended; (iv) the threshold value τ; and (v) the maximum number of iterations $\mathbb{N}$. Note that τ together with $\mathbb{N}$ construct the criterion to terminate the back-fitting iterations; τ could be a fairly small value to determine whether the back-fitting process has converged, and is vital for the fitting accuracy; while $\mathbb{N}$ is a sufficiently large integer to avoid running endlessly, and it is likely that the back-fitting procedure fails to converge if terminated by the iteration-count number k reaching $\mathbb{N}$. Computational efficiency of the back-fitting algorithm will be further discussed in context of the results from the following case study.

3. Case study with London house price data

In the short and preliminary study of Lu *et al*. (2015), a simulation experiment was conducted on a 25*25 square grid, where one regression parameter was set as stationary ($\beta_1 = 2$), while the other was set as nonstationary ($\beta_2 = \log(u + v)$, where (u, v) are the coordinates), and there was no intercept term specified. Accordingly, the former regression parameter was estimated using a zero (constant) distance metric (i.e. distance between any pair of locations is zero), while the latter regression parameter was estimated using the usual ED metric. The results from this very basic simulation experiment demonstrated the potential of PSDM GWR in providing more accurate parameter estimates and more accurate predictions than that found with standard GWR model. Moreover, the simulation study actually calibrated PSDM GWR as a special case of mixed GWR. In fact, PSDM GWR is a general form of many models: (a) the global regression, (b) basic GWR, (c) mixed GWR, and (d) FB GWR; where PSDM GWR allows the added flexibility of distance metric choice for each regression relationship. We now empirically demonstrate the potential of the PSDM GWR model using real data.

3.1. *Data and model*

As our case study data, we use the same London house price data set described in Lu *et al*. (2014a). It consists of 2108 properties sold during the 2001 calendar year, as visualized in Figure 1. For these data, the house sale price *PURCHASE* (the dependent variable) is combined with a series of hedonic characteristics (the independent variables), including those measuring: structural characteristics, construction time, property type, and local household income conditions. Following a forward selection procedure described on Lu *et al*. (2014a), we retain only the following three independent variables: *FLOORSZ* (the floor size of the property in square meters), *BATH2* (a dummy variable indicating if the property has two or more bathrooms), and *PROF* (the percentage of the workforce in professional occupations in the census enumeration district in which the property is located). Thus, the corresponding GWR model can be expressed as:

$$PURCHASE_i = \beta_{0i} + \beta_{1i}FLOORSZ_i + \beta_{2i}BATH2_i + \beta_{3i}PROF_i, \tag{9}$$

where the subscript i means a local regression calibration location (or regression point).

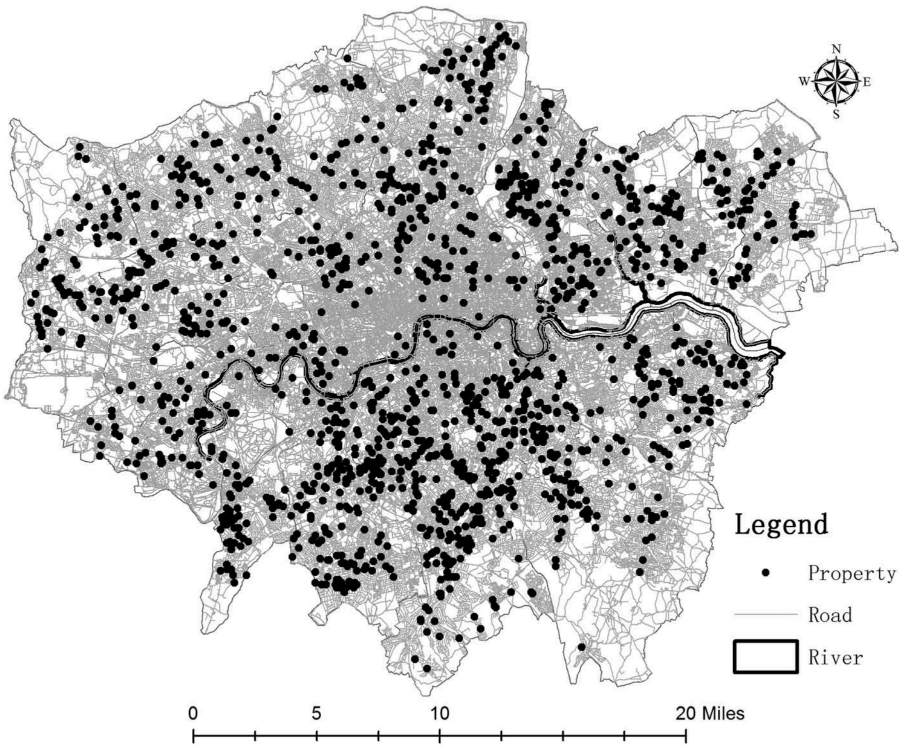


Figure 1. House price data and road network data in London.

3.2. *GWR calibrations with ED, TT, and PS distance metrics*

Lu *et al*. (2014a) calibrated the GWR model for Equation (9) with ED and TT metrics, respectively, where the TT metric outperformed that with the ED metric. In that study, a single distance metric together with a single bandwidth was used uniformly for estimating all the GWR parameters. However, the four sets of parameters (i.e. $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$) may themselves individually vary across different spatial scales and whose closeness in space may not be represented in the usual 'as the crow flies' distances. Here *PROF* represents spatial variation agreeing with simple proximities at the census enumeration district level and thus the usual ED metric is considered suitable. The determinants *FLOORSZ* and *BATH2*, however, are the structural attributes of a property, of which the spatial nonstationarities might be more reflected by the accessibilities between each regression calibration location and the observation locations. Thus, it is very possible (and following the results of Lu *et al*. 2014a), the TT metric is suitable in both cases. Therefore, in our PSDM GWR model fit, the TT metric is used to estimate *FLOORSZ* and *BATH2* (i.e. $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$), and the ED metric is used to estimate the *Intercept*[1] and *PROF* (i.e. $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_3$).

3.2.1. *Summary of the GWR calibrations with ED, TT, and PS distance metrics*

In order to be able to coherently compare the results of this study with those of Lu *et al*. (2014a), fixed bandwidths and Gaussian kernel function are again adopted for the

Table 1. GWR model calibration information and outputs using ED, TT, and PSDM metrics.

	GWR (ED)	GWR (TT)	GWR (PSDM)			
	All	All	β_0	β_1	β_2	β_3
Distance metric	ED	TT	ED	TT	TT	ED
Bandwidth	1914(m)	175(s)	51,137(m)	100(s)	58(s)	51,175(m)
AICc	37,382.97	37,293.96	36,471.88	37,056.98	36,626.56	36,471.86
R-squared	0.864	0.885	0.901			
RSS	1.002638e + 12	849078765203	7.316638e + 11			

following three GWR calibrations: (i) GWR with ED metric (i.e. the standard GWR calibration), (ii) GWR with the TT metric, and (iii) PSDM GWR using the ED metric for estimating $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_3$, and the TT metric for estimating $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, respectively. The model calibration information (distance metrics and bandwidths) and outputs (AICc, R-squared, and RSS) are given in Table 1.

Observe that a single bandwidth is used uniformly for both ED GWR and TT GWR calibrations, but that a distinct bandwidth is specified for each parameter estimate in the PSDM GWR calibration. Thus, the number of distance metrics used directly relates to the number of bandwidths needed. Specifically, the bandwidth in every back-fitting iteration is always optimized via the AIC approach, and the bandwidths for the PSDM GWR calibration in Table 1 are the ones used in the final iterations for all the independent variables. As shown in Lu et al. (2014a), there is only a moderate difference between the bandwidths of ED GWR (1914 meters) and TT GWR (175 s, which can be approximately recognized as a distance bandwidth at 2346 meters if we take the average speed is 30 miles/h) calibrations; noting that straight line distances will always be smaller than real-world travelling distances. However, they are very different from the bandwidths found in the PSDM GWR calibration.

In the PSDM GWR model, bandwidths under the ED metric for $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_3$ are very similar and are relatively large (at 51,137 and 51,175 meters, respectively); bandwidths under the TT metric for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are considerably different and relatively small (i.e. 100 and 58 s, respectively). Intuitively, these bandwidths are reasonable. *PROF* is sampled at the census enumeration district level and its spatial behavior is expected to be of an ED nature with a bandwidth to suit. Bandwidths for *FLOORSZ* and *BATH2* reflect houses that are geographically close together (in an ED sense) excluding those that have a significant barrier between them, such as houses directly opposite each other but on different sides of the river Thames. The use of the TT metric picks up on these subtle spatial effects for these particular hedonic variables.

Note also, there are four AICc values for the PSDM GWR calibration, which seems unconventional when compared with the ED GWR and TT GWR calibrations. The PSDM GWR model is fitted via the aforementioned back-fitting procedure, where a hat matrix can only be calculated from regressing one independent variable successively. Figure 2 presents the AICc values calculated from all the iterations. It is notable that the AICc for each independent variable converges, but to different values. The smallest AICc indicates the best level of goodness-of-fit (GoF) this PSDM GWR model could accomplish, while the largest one shows the bottom line for its performance. Hence, it is reasonable to evaluate the PSDM GWR model performance via the smallest of the four AICc values at 36,471.86, that is, a strong reduction of 822.1 compared with TT GWR and a strong

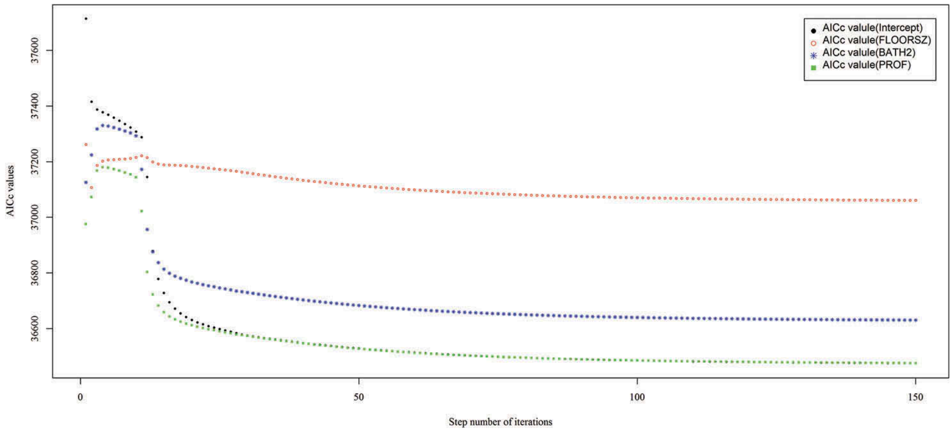


Figure 2. AICc values calculated from all the iterations in the PSDM GWR calibration.

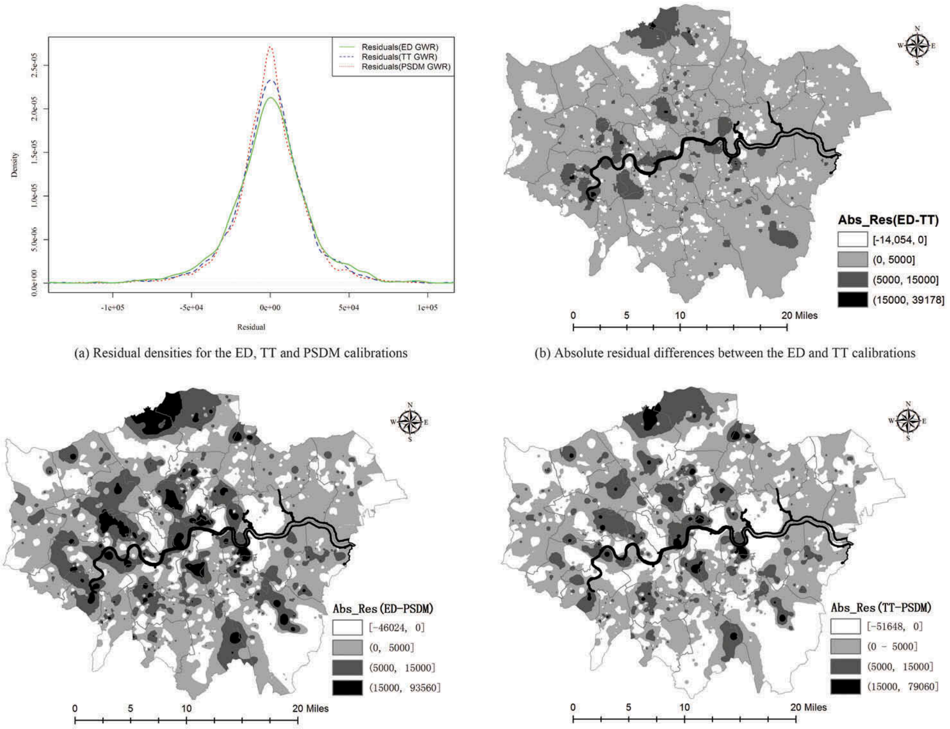
reduction of 911.11 compared with ED GWR. Such differences in AICc suggest great potential in the PSMD GWR model, at least in terms of GoF, over the two alternatives.

Furthermore, the R-squared from the PSDM GWR calibration improves by 1.6% in comparison with the TT GWR calibration, and by 3.7% compared with the ED GWR calibration. These reductions indicate better predictive performance from PSDM GWR over the TT and ED GWR calibrations. Note however, that it is unwise to compare the three GWR models by their R-squared only, since values are directly affected by the bandwidths, which are not comparable when different distance metrics are used (see Lu *et al*. 2014a for further discussions). Similar cautions should be taken with the RSS values.

3.2.2. *Residual comparisons for ED GWR, TT GWR, and PSDM GWR models*

In Figure 3(a), we plot the residual densities from the ED GWR, TT GWR, and PSDM GWR fits, where their shapes indicate normality, as would be expected. The red line in Figure 3(a) corresponds to the residual density for the PSDM GWR model, and shows the highest density of values around 0 together with the lowest density of values in tails (high positive and negative residuals). In this respect, the PSDM GWR model can be taken as most accurate predictor of the three study models.

To investigate model prediction accuracy spatially, we produce three discrepancy maps of the absolute value of a GWR predicted *PURCHASE* price minus the actual price, minus the same calculation from an alternative GWR calibration. The residual maps are smoothed using inverse distance weighting. Positive values in Figure 3(b–d) indicate better predictions from the latter calibration (i.e. TT GWR in Figure 3(b) and PSDM GWR in Figure 3(c–d)), while negative values imply better predictions from the former calibration (i.e. ED GWR in Figure 3(b–c) and TT GWR in Figure 3(d)). Figure 3(b) clearly shows that TT GWR predicts better than ED GWR in most regions, especially along the west and middle parts of the River Thames. From Figure 3(c–d), the PSDM GWR model appears to perform both the best and the worst locally, with no discernable spatial pattern evident. The relative improvement in prediction accuracy of PSDM GWR appears strongest in the northern and western parts of London when compared with ED GWR. However, TT GWR outperforms PSDM GWR along the River Thames, which highlights the



(a) Residual densities for the ED, TT and PSDM calibrations

(b) Absolute residual differences between the ED and TT calibrations

(c) Absolute residual differences between the ED and PSDM calibrations

(d) Absolute residual differences between the TT and PSDM calibrations

Figure 3. Residual comparisons for the ED GWR, TT GWR, and PSDM GWR models. The legend titles mean the absolute residuals from the first GWR model minus the absolute residuals from the second GWR model.

effectiveness of the TT metric when an obvious barrier is present. Although PSDM GWR is known to be the most accurate predictor overall or in a global sense (from Table 1), locally it outperforms ED GWR at 62% of the 1601 regression locations; and outperforms TT GWR at 56.8% of the locations. However, the improvements are difficult to fully appreciate in any detail. This requires more involved investigations of the TT metric, overlaid onto the London road network (see also Lu *et al.* 2014a), together with the discrepancy maps shown.

3.2.3. *Parameter comparisons for ED GWR, TT GWR, and PSDM GWR models*

In Figure 4, we compare the four sets of parameters from the three calibrations via 12 pair-wise scatterplots. From the first column of Figure 4, it is clear that the ED GWR and TT GWR models produce relatively similar and thus highly correlated parameter estimates (with correlation coefficients around 0.95). The PSDM GWR model however, produces very different parameter estimates to the other two models, as presented in the second and third columns of Figure 4. This behavior is entirely expected, as different distance metrics, and furthermore, different bandwidths are used for its calibration (as presented in Table 1). Estimates of the *Intercept* and *PROF* are fairly stable within the study area, ranging only from −18,094.47 to −17,994.24, from 1080.33 to 1082.59, respectively. This is not surprizing since the corresponding bandwidths (at 51,137 and 51,175 meters) are tending toward a

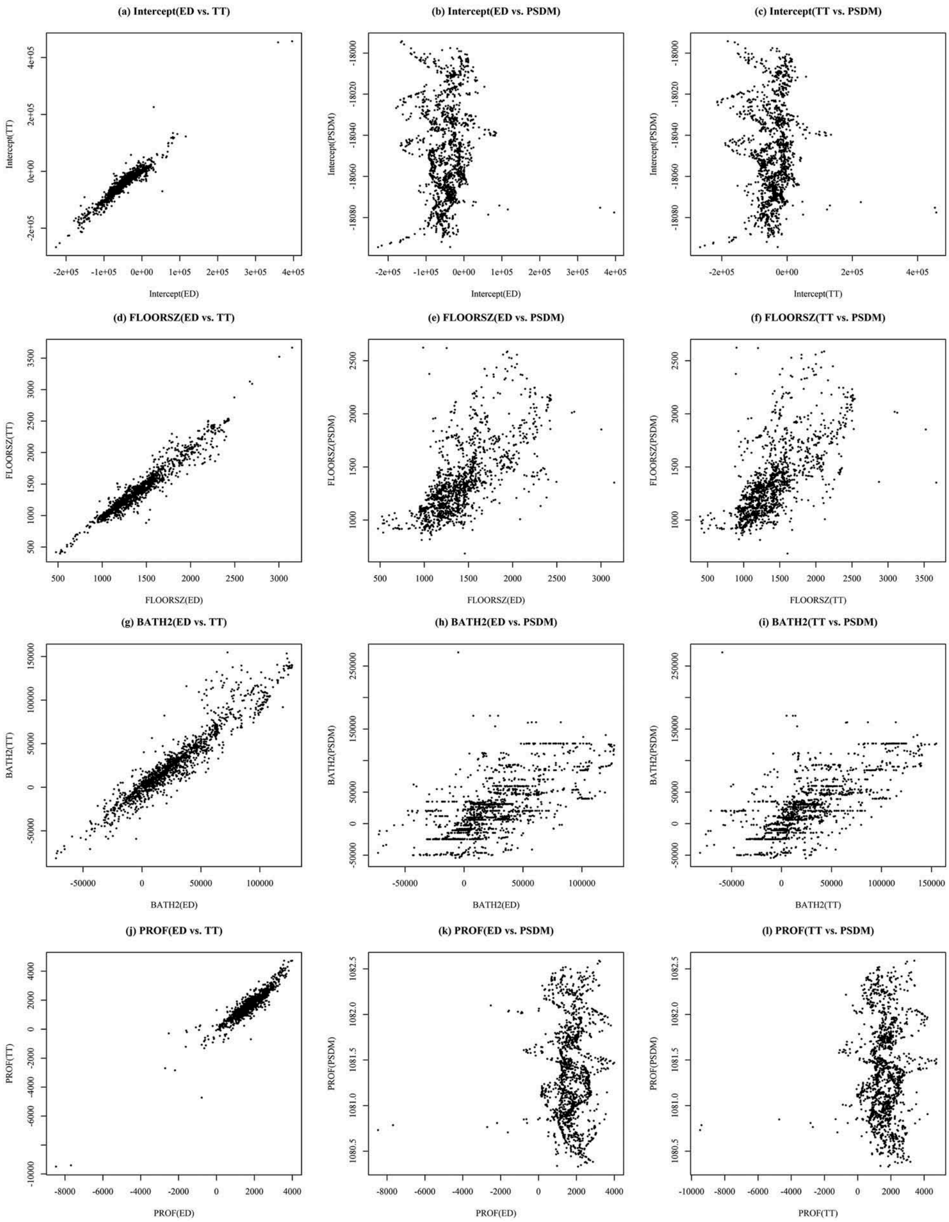


Figure 4. Parameter comparisons for the ED GWR, TT GWR, and PSDM GWR calibrations. The scatterplot label 'FLOORSZ(ED)' means the estimated parameters for the *FLOORSZ* variable from the ED GWR model. The rest of the scatterplot labels can be interpreted accordingly.

global estimation of these two parameters. This is logical for *PROF*, as it is a large scale economical index. Estimates of *FLOORSZ* and *BATH2* for the PSDM GWR model are positively correlated with the corresponding estimates from the ED GWR and TT GWR models, which indicate homogeneous patterns for these two independent variables.

Overall, the PSDM GWR model gives very different parameter estimates from the GWR models with only one distance metric, used uniformly. Differences are fundamentally caused by the distinctive weighting schemes that result from using both ED and TT metrics and employing flexible bandwidths, all within the same GWR model. Parameter estimates from PSDM GWR should display distinctive spatial scales of variation, for each parameter surface in turn. Results also suggest that a particular non-ED metric mixed GWR model would be a pragmatic choice for this data, that is, let the *Intercept* and *PROF* be fixed globally, but let *FLOORSZ* and *BATH2* vary locally (with their weights via the TT metric). In this respect, the PSDM GWR model provides a useful model selection tool for finding a simpler GWR fit.

4. Heuristics to reduce computational cost

The simulation study results of Lu *et al*. (2015) and this paper's empirical results demonstrate that a PSDM GWR model can provide a worthy improvement over alternative GWR models. However, the back-fitting algorithm for calibrating a PSDM GWR model is highly computer intensive. As stated in Section 2.2, there are five factors that might affect these costs: (i) choice of CVR; (ii) the threshold value τ; (iii) the maximum number of iterations $\mathbb{N}$; (iv) initial guess of $\widehat{\boldsymbol{\beta}}^{(0)}$; and (v) bandwidths specified or to be specified. The former three factors immediately determine the convergence speed, but it is quite straightforward to reset them accordingly. Addressing the latter two factors, however, is not so straightforward, and in this section we look at initial guesses of $\widehat{\boldsymbol{\beta}}^{(0)}$ and bandwidth selection to determine useful heuristics for reducing computational costs. In other words, we propose strategies for initializing the parameters and optimizing bandwidth selection in order to reduce computational burden.

In the case study, bandwidths used for the PSDM GWR model are always optimized via the AICc approach within every iteration step. Figure 5 shows all of the bandwidths used in the back-fitting process. Observe that the bandwidth values converge fast and won't change any more, even if all of them will be re-optimized in the next iteration step. This entails that we don't have to select the bandwidth in every iteration step, and instead we could stop optimizing when it converges. We can introduce another threshold value δ to define whether the bandwidth for a specific parameter has converged or not, and its value for this parameter will be kept the same in the following iterations, when its change from the last one is less than δ. Note here that the bandwidth selection for different parameters may stop at quite diverse steps, depending on the properties of the spatial process.

On the other hand, the converging speed is crucial for the executing efficiency of the back-fitting algorithm. Ideally, the objective is that the CVR value should decrease underτ within the minimum number of iterations. It is largely controlled by the choice of CVR and τ up to the accuracy requirements. Moreover, the initial values of $\widehat{\boldsymbol{\beta}}^{(0)}$ are also crucial. With the studied PSDM GWR model above, we tried three different strategies to initialize the parameters $\widehat{\boldsymbol{\beta}}^{(0)}$ for the back-fitting procedure, of which the details are as follows:

- Strategy 1: Use the parameters from a standard GWR model, that is, ED GWR calibration.

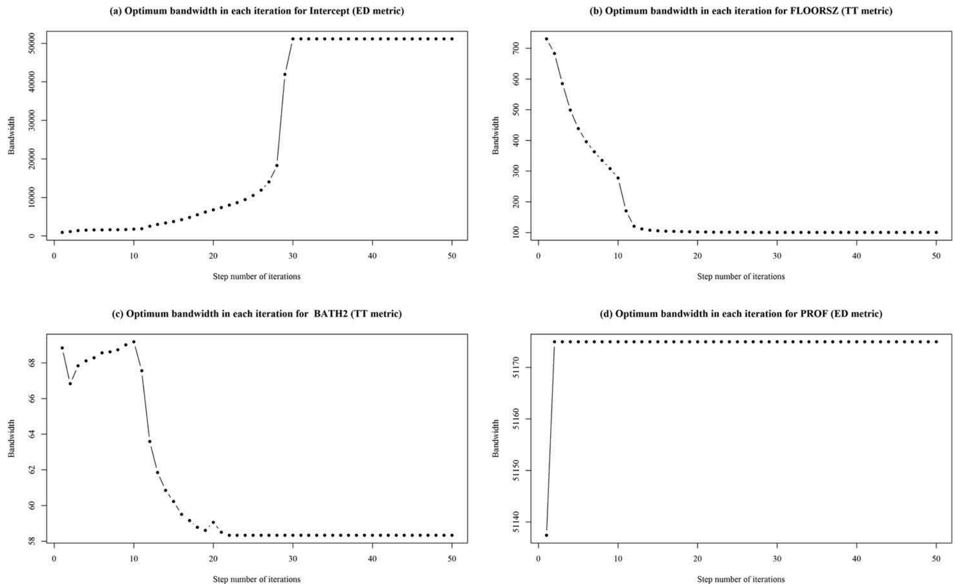


Figure 5. Bandwidths for all the iterations in the back-fitting algorithm[2].

- Strategy 2: Use the parameters from a separate back-fitting process with a τ value defined loosely (say 10^{-2}, while 10^{-6} for the PSDM GWR calibration) and bandwidths provided casually for each independent variable (say the average value from the corresponding metric matrix).
- Strategy 3: Similarly, use the parameters from a separate back-fitting process with τ as 10^{-2}, but use the parameter-specific bandwidths from optimizing GWR models between the dependent variable (i.e. *PURCHASE*) and each independent variable (i.e. *Intercept, FLOORSZ, BATH2,* and *PROF*, respectively) with the corresponding distance metric (i.e. ED, TT, TT and ED metrics, respectively).

The PSDM GWR model was calibrated following strategies 1, 2, and 3 for initializing $\widehat{\boldsymbol{\beta}}^{(0)}$, and the CVR values within the three back-fitting processes are presented in Figure 6. Observe that strategy 3 provides the fastest convergence, while strategy 1 appears the worst choice. However, strategy 3 actually provides the heaviest computational cost for initialization, while strategy 1 is the most straightforward. On balance (from a limited number of experiments), we recommend strategy 3 to initialize $\widehat{\boldsymbol{\beta}}^{(0)}$, and we made it the default strategy in the PSDM GWR calibration routine.

5. Discussion and concluding remarks

In this article, we proposed a new GWR model with parameter-specific distance metrics (PSDM GWR), which also allows parameter-specific bandwidths to be specified, as in the flexible bandwidth (FB GWR) model of Yang (2014). Thus, FB GWR is a special case of PSDM GWR, when only EDs are specified. Similarly, the global regression, basic GWR,

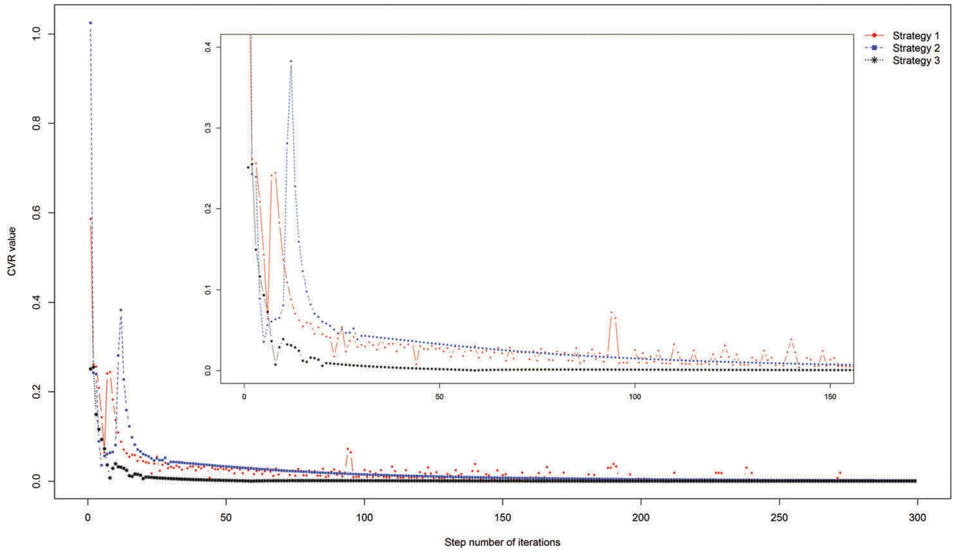


Figure 6. CVR values from three PSDM GWR calibrations with different initialization strategies.

mixed GWR, and GWR with only one distance metric are also specific cases of PSDM GWR.

Via a case study with London house price data, results indicated that a PSDM GWR model can clearly improve GWR model performance in terms of GoF and prediction accuracy over a GWR model specified with EDs (the basic model) and a GWR model specified with TT metrics. Here three independent (hedonic) variables were specified to predict house price (*PURCHASE)*, where large bandwidths under the ED metric of the PSDM GWR model were used to estimate the *Intercept* and *PROF* (percentage of professionals in the properties census enumeration district) parameters, while small bandwidths under the TT metric of the PSDM GWR model were used to estimate *BATH2* (indicator of more than two bathrooms) and *FLOORSZ* (house floor area) parameters.

Results from the case study suggest that PSDM GWR can not only detect variations in regression relationships across different spatial scales, as in FB GWR and mixed GWR, but also help determine the appropriate distance metric for such spatial scales. For this study, the distance metrics for PSDM GWR were user-specified in view of empirical knowledge. However in many instances, a natural distance metric for a given independent variable may not be forthcoming, and choosing metrics may be difficult when the number of independent variables is large. In such cases, a brute-force search with a range of candidate metrics could be tried. This solution however, could be extremely computation intensive, and its effectiveness would still depend on the variety of metric candidates chosen. Only two candidates were specified (Euclidean and TT) in this study, but metrics such as network and Minkowski distance could also have been considered. This difficult model specification issue is open for review and forms an essential part of our future work for the PSDM GWR approach. Although as a guide to its current implementation, the default metric should always be Euclidean, which can be replaced by an alternative when there is good empirical knowledge to do so.

As with mixed GWR and FB GWR, a back-fitting algorithm is needed for PSDM GWR calibration, and leads to a heavy computational cost. Heuristics for bandwidth selection and strategies on initializing the parameters were presented in order to improve the back-fitting algorithm's efficiency. Notably, three strategies for parameter initialization were tested, but from a quite limited number of experiments. In this respect, more strategies or more rigorous evaluations on their performances will be studied in the future. Future studies will also expand the model comparison to include all models that require the back-fitting algorithm: PSDM GWR (with only EDs – i.e. FB GWR) versus PSDM GWR (with only non-EDs but of the same type) versus PSDM GWR (with a mixture of distance metrics – as specified in this study).

Further work could also: (i) select distance metrics for the PSDM GWR approach via the use of Minkowski distance metrics (Lu *et al*. 2016); (ii) investigate in more detail, the distance metric for the intercept term of the PSDM GWR model; (iii) refine the computational efficiency for the PSDM GWR approach, via more heuristics and/or high performance computing techniques. Further work could also look at the application of PSDM and related GWR models to environmental data measured on river and stream networks where flow direction is also important (see the literature review given in Lu *et al*. 2014a). All routines and functions used in this paper will be integrated into the **GWmodel R package** (Lu *et al*. 2014b, Gollini *et al*. 2015), which provides a framework for handling spatially varying structures, via a wide range of geographically weighted models.

Acknowledgments

Research presented in this paper was jointly supported by projects from the National Natural Science Foundation of China [NSFC: 41401455], [NSFC: U1533102]; an open research fund of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (No. 14I01); an open research fund by State Key Laboratory of Resources and Environmental Information System (No. 1610); a UK Biotechnology and Biological Sciences Research Council grant (BBSRC BB/J004308/1). We thank all the reviewers for their valuable comments and suggestions, which are very important for improving this article.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Research presented in this paper was jointly supported by projects from the National Natural Science Foundation of China [NSFC: 41401455 and U1533102]; an open research fund of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing [Grant Number: 14I01]; an open research fund by State Key Laboratory of Resources and Environmental Information System [Grant Number: 1610]; a UK Biotechnology and Biological Sciences Research Council Grant Number: BBSRC BB/J004308/1.

Notes

1. There is no obvious choice of metric to estimate the *Intercept*, and as such we use the default ED metric.
2. As the bandwidth values converge fast and won't change anymore, only values from the first 50 iterations are drawn in this figure.

ORCID

Binbin Lu http://orcid.org/0000-0001-7847-7560
Chris Brunsdon http://orcid.org/0000-0003-4254-1780

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. *In: 2nd international symposium on information theory*, 2–8 September, Tsahkadsor, Armenian SSR, 267–281.

Assunção, R.M., 2003. Space varying coefficient models for small area data. *Environmetrics*, 14 (5), 453–473. doi:10.1002/env.599

Bowman, A.W., 1984. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71 (2), 353–360. doi:10.1093/biomet/71.2.353

Brunsdon, C., Fotheringham, A.S., and Charlton, M., 1999. Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science*, 39 (3), 497–524. doi:10.1111/0022-4146.00146

Brunsdon, C., Fotheringham, A.S., and Charlton, M.E., 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, 28 (4), 281–298. doi:10.1111/j.1538-4632.1996.tb00936.x

Casetti, E., 1972. Generating models by the expansion method: applications to geographical research. *Geographical Analysis*, 4 (1), 81–91. doi:10.1111/j.1538-4632.1972.tb00458.x

Cho, S.-H., Lambert, D.M., and Chen, Z., 2010. Geographically weighted regression bandwidth selection and spatial autocorrelation: an empirical example using Chinese agriculture data. *Applied Economics Letters*, 17 (8), 767–772. doi:10.1080/13504850802314452

Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74 (368), 829–836. doi:10.1080/01621459.1979.10481038

Farber, S. and Páez, A., 2007. A systematic investigation of cross-validation in GWR model estimation: empirical analysis and Monte Carlo simulations. *Journal of Geographical Systems*, 9 (4), 371–396. doi:10.1007/s10109-007-0051-3

Fotheringham, A.S. and Brunsdon, C., 1999. Local forms of spatial analysis. *Geographical Analysis*, 31 (4), 340–358. doi:10.1111/j.1538-4632.1999.tb00989.x

Fotheringham, A.S., Brunsdon, C., and Charlton, M., 2002. *Geographically weighted regression: the analysis of spatially varying relationships*. Chichester: Wiley.

Fotheringham, A.S., Crespo, R., and Yao, J., 2015. Geographical and Temporal Weighted Regression (GTWR). *Geographical Analysis*, 47 (4), 431–452. doi:10.1111/gean.2015.47.issue-4

Gelfand, A.E., *et al*., 2003. Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98 (462), 387–396. doi:10.1198/016214503000170

Gollini, I., *et al*., 2015. GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. *Journal of Statistical Software*, 63 (17), 1–50. doi:10.18637/jss.v063.i17

Goodchild, M.F., 2004. The validity and usefulness of laws in geographic information science and geography. *Annals of the Association of American Geographers*, 94 (2), 300–303. doi:10.1111/j.1467-8306.2004.09402008.x

Gorr, W.L. and Olligschlaeger, A.M., 1994. Weighted spatial adaptive filtering: Monte Carlo studies and application to illicit drug market modeling. *Geographical Analysis*, 26 (1), 67–87. doi:10.1111/j.1538-4632.1994.tb00311.x

Griffith, D.A., 2008. Spatial-filtering-based contributions to a critique of Geographically Weighted Regression (GWR). *Environment and Planning A*, 40 (11), 2751–2769. doi:10.1068/a38218

Harris, R., Dong, G., and Zhang, W., 2013. Using contextualized geographically weighted regression to model the spatial heterogeneity of land prices in Beijing, China. *Transactions in GIS*, 17 (6), 901–919. doi:10.1111/tgis.2013.17.issue-6

Hastie, T. and Tibshirani, R., 1986. Generalized additive models. *Statistical Science*, 1 (3), 297–310. doi:10.1214/ss/1177013604

Hoaglin, D.C. and Welsch, R.E., 1978. The hat matrix in regression and ANOVA. *The American Statistician*, 32 (1), 17–22.

Huang, B., Wu, B., and Barry, M., 2010. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, 24 (3), 383–401. doi:10.1080/13658810802672469

Hurvich, C.M., Simonoff, J.S., and Tsai, C.-L., 1998. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 60 (2), 271–293. doi:10.1111/1467-9868.00125

Lu, B., *et al*., 2016. The Minkowski approach for choosing the distance metric in geographically weighted regression. *International Journal of Geographical Information Science*, 30 (2), 351–368. doi:10.1080/13658816.2015.1087001

Lu, B., Charlton, M., and Fotheringham, A.S., 2011. Geographically weighted regression using a non-euclidean distance metric with a study on London house price data. *Procedia Environmental Sciences*, 7 (0), 92–97. doi:10.1016/j.proenv.2011.07.017

Lu, B., *et al*., 2014a. Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data. *International Journal of Geographical Information Science*, 28 (4), 660–681. doi:10.1080/13658816.2013.865739

Lu, B., *et al*., 2014b. The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-Spatial Information Science*, 17 (2), 85–101. doi:10.1080/10095020.2014.917453

Lu, B., *et al*., 2015. Calibrating a geographically weighted regression model with parameter-specific distance metrics. *Procedia Environmental Sciences*, 26, 109–114. doi:10.1016/j.proenv.2015.05.011

Mei, C.-L., He, S.-Y., and Fang, K.-T., 2004. A note on the mixed geographically weighted regression model. *Journal of Regional Science*, 44 (1), 143–157. doi:10.1111/j.1085-9489.2004.00331.x

Mei, C.-L., Xu, M., and Wang, N., 2016. A bootstrap test for constant coefficients in geographically weighted regression models. *International Journal of Geographical Information Science*, 30 (8), 1622–1643.

Nakaya, T., *et al*., 2005. Geographically weighted poisson regression for disease association mapping. *Statistics in Medicine*, 24 (17), 2695–2717. doi:10.1002/(ISSN)1097-0258

Páez, A., 2004. Anisotropic variance functions in geographically weighted regression models. *Geographical Analysis*, 36 (4), 299–314. doi:10.1111/gean.2004.36.issue-4

Tobler, W.R., 1970. A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46 (2), 234–240. doi:10.2307/143141

Yang, W., 2014. *An extension of geographically weighted regression with flexible bandwidths*. Doctoral. St Andrews.

Yang, W., Fotheringham, A.S., and Harris, P., 2011. Model selection in GWR: the development of a flexible bandwidth GWR. *In*: *The 11th International Conference on GeoComputation*, 20–22 July London.

Yang, W., Fotheringham, A.S., and Harris, P., 2012. An extension of geographically weighted regression with flexible bandwidths. *Gisruk*, 2012, Lancaster. 1–7.