

# The Minkowski approach for choosing the distance metric in geographically weighted regression

Binbin Lu<sup>a</sup>, Martin Charlton<sup>b</sup>, Chris Brunsdon <sup>b</sup> and Paul Harris<sup>c</sup>

<sup>a</sup>School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China; <sup>b</sup>National Centre for Geocomputation, National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland;

<sup>c</sup>Sustainable Soils and Grassland Systems, Rothamsted Research, North Wyke, Okehampton, Devon, UK

## ABSTRACT

In this study, the geographically weighted regression (GWR) model is adapted to benefit from a broad range of distance metrics, where it is demonstrated that a well-chosen distance metric can improve model performance. How to choose or define such a distance metric is key, and in this respect, a 'Minkowski approach' is proposed that enables the selection of an optimum distance metric for a given GWR model. This approach is evaluated within a simulation experiment consisting of three scenarios. The results are twofold: (1) a well-chosen distance metric can significantly improve the predictive accuracy of a GWR model; and (2) the approach allows a good approximation of the underlying 'optimal distance metric', which is considered useful when the 'true' distance metric is unknown.

## ARTICLE HISTORY

Received 16 March 2015

Accepted 23 August 2015

## KEYWORDS

Non-stationarity; GW model; Minkowski distance; simulation experiment

## 1. Introduction

In many early developments of spatial regression modelling, models were applied at a global level, where data relationships were assumed constant across the study region (Fotheringham and Brunsdon 1999). However, the existence of relationship non-stationarity, which commonly appears as uncontrolled spatial variability, challenges such global forms. To account for such heterogenic spatial processes, a number of localized regression techniques have been proposed, most notably the expansion method (Casetti 1972), multilevel models (Jones 1991), geographically weighted regression (GWR) (Brunsdon *et al.* 1996), panel data models (Elhorst 2003) and space-varying parameter models using Bayesian constructs (Assunção 2003, Gelfand *et al.* 2003). For this study, we concentrate only on the GWR model, where it is adapted to benefit from a broad range of distance metrics.

For GWR, spatially varying relationships are explored between the dependent and the independent variables, in turn. The data are geographically weighted using some distance-decay kernel weighting function, allowing nearer observations to have more influence in estimating a local set of regression coefficients than observations farther away. The resultant regression coefficients and associated (pseudo) *t*-values are then

mapped to determine evidence of relationship non-stationarity (e.g. [Harris et al. 2010b](#)). Such maps can also be complemented by associated tests for relationship non-stationarity (e.g. [Brunsdon et al. 1998](#)). In most applications of GWR, Euclidean distance (ED) is used to determine the geographical weights, that is, a straight-line distance metric.

The first use of non-Euclidean distance (non-ED) metrics in GWR modelling can be found in [Lu et al. \(2011, 2014a\)](#) using a house price data set for London, UK. Here, the use of network distance (ND) and travel time (TT) metrics provided greater insight into data relationships than that found with the usual ED metric. The dominance of the River Thames, with its limited number of crossing points, was found to be crucial in assessing the choice of distance metric. This initial work only allowed three distinct choices of distance metric (ED, ND and TT), whereas for this study a 'Minkowski distance approach' is followed that enables an optimal metric selection from a continuous range of metrics for a given GWR model.

The scope of possible distance metrics in many spatial analyses is often far larger than the single ED option. Such metrics include, for example, a stream/river distance (e.g. [Legleiter and Kyriakidis 2008](#), [Money et al. 2009](#)) or some transportation ND (e.g. [Kent et al. 2006](#)), where spatial continuity only makes sense along the stream/river/etc., or road/rail/canal/etc., respectively. A spatial process may also vary with direction, and in this respect, anisotropic processes can be represented using non-ED metrics (e.g. [Boisvert and Deutsch 2011](#)). In particular, [Páez \(2004\)](#) found GWR models calibrated with anisotropic kernels to outperform their isotropic (standard) counterparts. An optimal distance metric for a given spatial process is not always clear; metrics can vary due to the diversity of the sample data and the complexity of the geography. Conditions may be such that the distance metric itself cannot be properly measured or calculated. For example, TTs are commonly based on judged approximations. Thus, defining an appropriate distance metric when calibrating a spatial model presents a certain challenge.

In this respect, the Minkowski distance metric can be used, as by varying its exponent parameter  $p$ , together with the coordinate rotation angle  $\theta$ , a range of (commonly used) non-ED metrics can be accounted for. For example, [Shahabi et al. \(2002\)](#) and [Shahid et al. \(2009\)](#) use Minkowski distance to approximate road distances and TTs. [Love et al. \(1988\)](#), cited by [Miller and Wentz \(2003\)](#), indicate that the value of  $p$  typically ranges from 1 to 2 for representing the true travel distance at urban and regional scales. For a spatial model that is adapted to use Minkowski distances, optimal values of  $p$  and  $\theta$  are commonly found that minimize some goodness-of-fit (GoF) criteria. For GWR, such a dual optimization needs to be used in conjunction with optimally finding the kernel bandwidth parameter  $b$  (which controls the spatial scale of relationship non-stationarity), and as such, this presents one of many challenges that this study seeks to address.

We present our research over four sections. Firstly, we describe how the GWR model is adapted to use Minkowski distances. Secondly, we assess our adapted GWR model within a simulation experiment consisting of three scenarios. Thirdly, we investigate ways to improve the computational efficiency of this adaptation. Finally, we summarize the study. All modelling functions used in this article can be found in the GW model (after version 1-2.5) R package ([Lu et al. 2014b](#), [Gollini et al. 2015](#)), which is an integrated framework for handling spatially varying model parameters, moments, statistics and

data structures, via geographically weighted models. R is an open-source statistical programming environment (R Development Core Team 2013).

## 2. Methodology

### 2.1. The Minkowski distance

The Minkowski distance, also known as the  $p$ -norm distance, is a general form of a distance metric in Euclidean space. It is expressed as

$$d = \left( \sum_{i=1}^m |u_i - v_i|^p \right)^{1/p} \quad (1)$$

where  $(u_1, u_2, \dots, u_m)$  and  $(v_1, v_2, \dots, v_m)$  are two vectors in  $m$ -dimension Euclidean space, and  $p$  is a positive real number. It is a generalization of all commonly used metrics on a Euclidean space. When  $p$  is 1, 2 and *infinity*, the distance is known as the Manhattan distance, the ED and the Chebyshev distance, respectively. A different value of  $p$  represents a distinctive measure of distance in a Euclidean space. Take the 2-D Euclidean space, for example, where distance iso-surfaces for different values of  $p$  can be produced, as shown in Figure 1 for nine values of  $p$ . Surfaces in Figure 1 show that adjacent values of  $p$  depict similar patterns. Larger metrics are found when values of  $p$  are less than 2, whilst smaller metrics are found when values of  $p$  larger than 2.

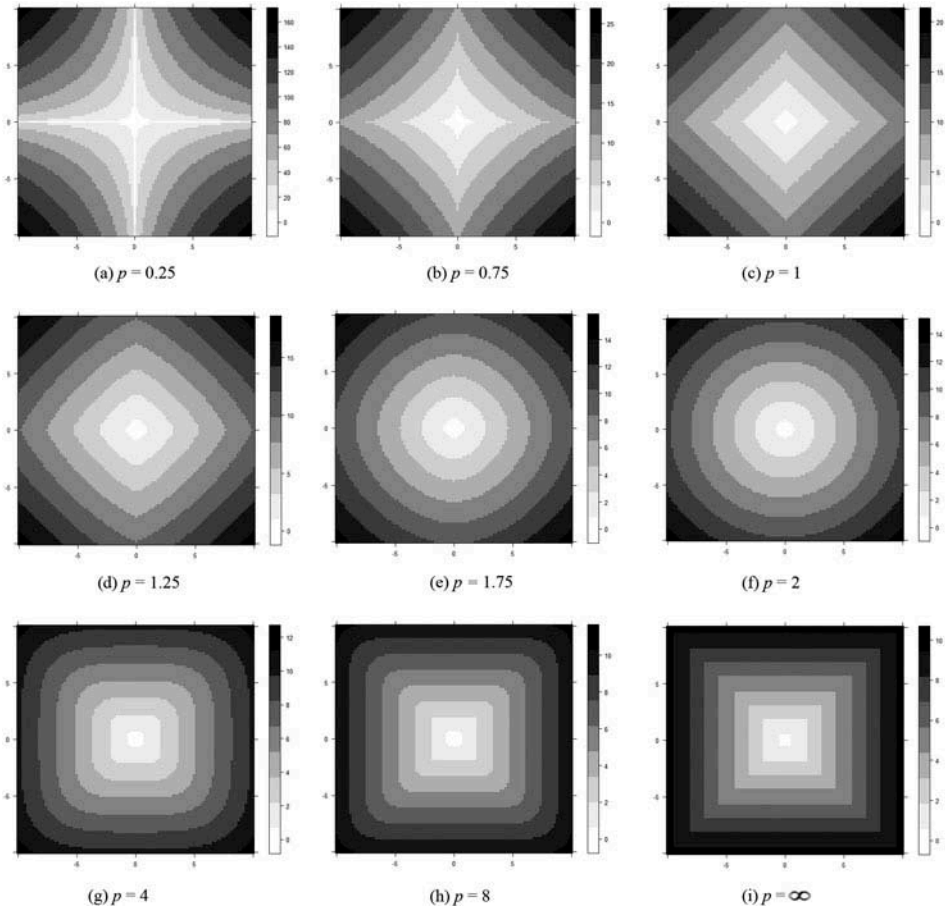
From the geometric properties of the surfaces in Figure 1, the rotation of the coordinate system will also lead to changes in the distance measurement, except the circular shapes in Figure 1f when  $p = 2$ . Thus, the rotation angle  $\theta$  also needs to be taken into consideration for distance metric selection. In the 2-D Euclidean space, the expression of Minkowski distance can be defined as

$$d_{p,\theta} = \sqrt{(u_1 - u_2)^2 + (v_1 - v_2)^2} (|\sin(\theta + \alpha)|^p + |\cos(\theta + \alpha)|^p)^{\frac{1}{p}} \quad (2)$$

where  $(u_i, v_i)_{i=1,2}$  are the 2-D Euclidean coordinates, and the angle  $\alpha = \tan^{-1} \left( \frac{u_1 - u_2}{v_1 - v_2} \right)$ . It is easy to prove that  $d_{p,\theta}$  is a function of  $\theta$  with the primitive period  $\frac{\pi}{2}$  radians or  $90^\circ$ .

### 2.2. GWR with Minkowski distances

To model spatially varying relationships, GWR allows the coefficients of a regression model to vary continuously over space. Thus, unlike a basic (ordinary least squares (OLS)) regression model, the coefficients in a GWR model can be estimated at any location (e.g. on a dense grid so that a coefficient surface can be visualized). GWR makes a pointwise calibration concerning a 'bump of influence' around each (local calibration) regression point, where nearer observations (or datapoints) have more influence in estimating the local set of coefficients than observations farther away (Fotheringham *et al.* 1998). This procedure is depicted in Figure 2, where in this study observations are geographically weighted using a (discontinuous or truncated) Gaussian kernel function. Other types of kernel functions include the exponential, (continuous) Gaussian, bi-square, tri-cube and box-car (see Gollini *et al.* 2015). GWR measures the inherent relationships around each

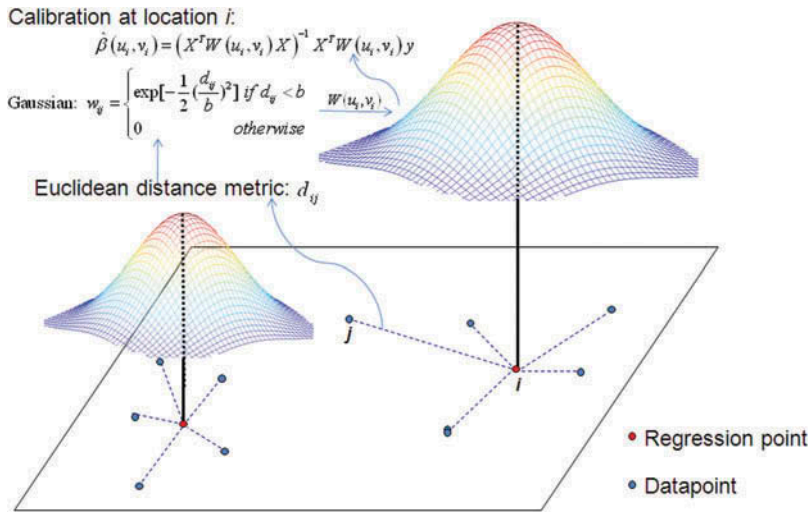


**Figure 1.** Distance iso-surface plots with nine different values of  $p$ .

regression point by estimating sets of regression coefficients by a weighted least-squares approach (of which the matrix expression is shown in [Figure 2](#)). The distance between the datapoint  $j$  and the regression point  $i$  is  $d_{ij}$  that is traditionally an ED metric with planar coordinates. [Lu et al. \(2011, 2014a\)](#) extend GWR to use two common non-ED metrics (network and TT), where for this study, we extend further, with the use of Minkowski distance metrics.

The kernel bandwidth  $b$  (also shown in [Figure 2](#)) is the key controlling parameter for any weighting scheme specified and strongly influences all GWR outputs. It can be in a fixed or adaptive distance form (see [Gollini et al. 2015](#)) and can be optimally found by minimizing some GoF diagnostic, for example, via a cross-validation or an Akaike information criterion (AIC) approach. AIC ([Akaike 1973](#)) is derived from the Kullback–Liebler information distance (Kullback and Leibler 1951) between two statistical distributions ([Brunsdon et al. 2000](#)), and its minimization provides a trade-off between GoF and degrees of freedom (i.e. the goal is model parsimony). In most GWR studies, a corrected AIC (AICc) ([Hurvich et al. 1998](#)) is used for selecting  $b$ .

We also adopt an AICc approach to find an *optimal* Minkowski distance metric (defined in terms of  $p$  and  $\theta$ ), together with an *optimal* bandwidth  $b$ , for a given GWR specification.



**Figure 2.** Interpretation of a basic GWR model and its calibration.

Based on the resultant variation in AICc, an optimal triplet  $(p, \theta, b)$  is found that corresponds to the smallest AICc. Finding  $(p, \theta, b)$  can follow a numerical search routine or an exhaustive search. Both can be computationally intensive, where using a numerical routine requires the derivatives of the model with respect to  $p$ ,  $\theta$  and  $b$ . Our Minkowski distance approach to calibrate a GWR model can be described in three stages:

- (1) Using AICc, find optimal values of  $b$  for a range of Minkowski distance metrics defined by  $p$  and  $\theta$  ( $p$  can be any non-negative real number and  $\theta$  can lie between 0 and  $\frac{\pi}{2}$  in radians). Thus, a bandwidth-AICc function is effectively found for every distinct distance metric that is specified.
- (2) Tabulate the minimized AICc values from step 1 in terms of  $p$  and  $\theta$ , where the smallest AICc of the table corresponds to the optimal triplet  $(p, \theta, b)$ . Observe that this tabulation could also be constructed in terms of  $b$  rather than AICc.
- (3) Interrogate the tabulated outputs for the best-fitting GWR model.

### 3. Simulation experiment

It is often appropriate to use simulated data when evaluating a new statistical method or an adaptation of an existing method, as, for example, [Wang et al. \(2008\)](#), [Harris et al. \(2010a\)](#) and [Lu et al. \(2012\)](#) demonstrate for GWR-related studies. In comparison with an empirical study, a simulation study can be done in a controllable way by following a certain number of rules; rules that enable a more objective model evaluation. Furthermore, GWR fits with different distance metrics will commonly result in different and unrelatable optimum bandwidths, and as such, it is inherently difficult to evaluate them with any objectivity using only real data. Therefore, we evaluate our proposed extension of GWR via a simulation experiment consisting of three scenarios. Our study aim is to demonstrate that the Minkowski approach can provide an approximation of

the underlying ‘optimal distance metric’ for a given GWR model, and that this approach is useful when the ‘true’ distance metric is unknown.

### 3.1. Simulation design

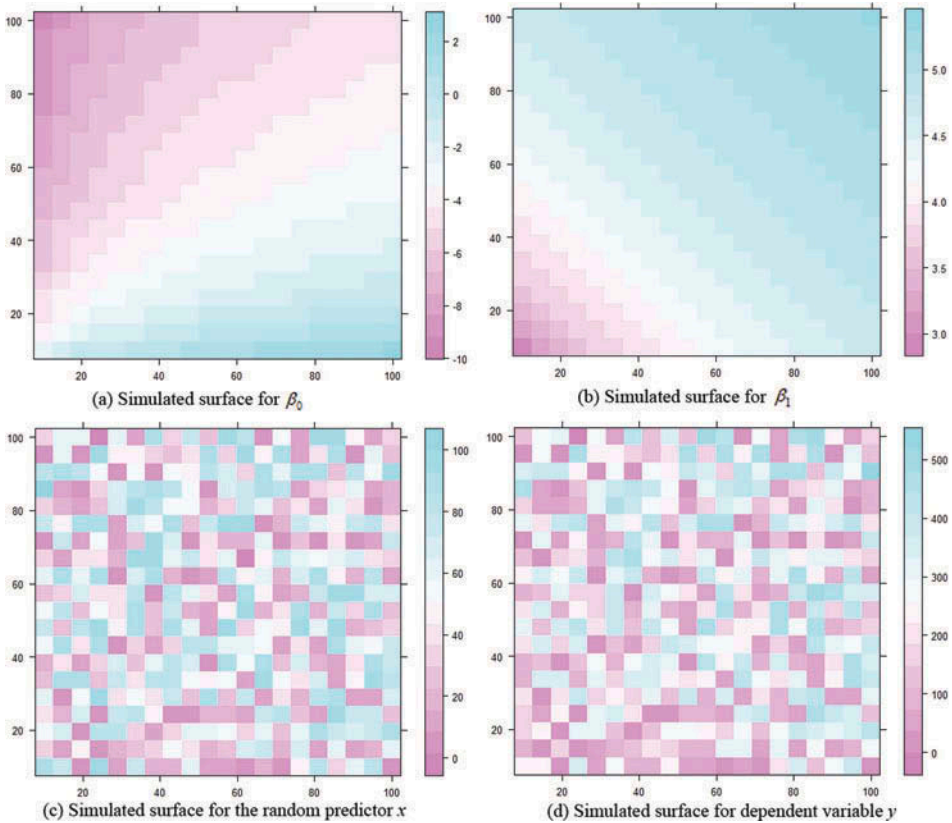
#### 3.1.1. Scenario 1: simulation of relationship spatial heterogeneity on a regular grid

For this first scenario, we use the same deterministic simulation design as that used in [Lu et al. \(2012\)](#). A data set of size  $20 \times 20$  is first generated on a square grid. For each cell, a predictor variable  $x$  is generated as a uniformly random numeric vector ranging from 1 to 100. Two non-stationary regression coefficients  $\beta_0$  and  $\beta_1$  are then generated on the same grid using Equations (3) and (4), respectively. The dependent variable  $y$  is then generated naturally from Equation (5). The resultant coefficients and data are visualized in [Figure 3](#).

$$\beta_0(u_i, v_i) = 2 \log(u_i) - 3 \log(v_i) \quad (3)$$

$$\beta_1(u_i, v_i) = \log(u_i + v_i) \quad (4)$$

$$y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_i \quad (5)$$



**Figure 3.** Scenario 1: visualization of the simulated coefficients and the predictor and dependent variables.

### 3.1.2. Scenario 2: manipulation of scenario 1 to reflect anisotropy and disjointedness

In scenario 1, the simulation is simply made on a  $20 \times 20$  square grid with spatially heterogenic coefficients. In this second scenario, anisotropic manipulations and a sense of disjointedness are incorporated to the same simulation, where it would be expected that GWR using a non-ED metric would be the ideal choice and perform relatively well. Here, we divide the square grid of scenario 1 into four equal parts, that is,  $10 \times 10$  square subgrids. As shown in Figure 4a, these four parts are labelled in different colours and symbols. For each part, we then rotated their coordinates in different angles and rescaled them as follows:

- Part 1 (red crosses): coordinates rotated by  $\pi/6$  (in radians) and multiplied by 2.
- Part 2 (green dots): coordinates rotated by  $\pi/3$  (in radians) and multiplied by 3.
- Part 3 (yellow squares): coordinates rotated by  $\pi/4$  (in radians) and multiplied by 4.
- Part 4 (blue panes): coordinates rotated by  $3\pi/4$  (in radians) and multiplied by 5.

In comparison with scenario 1, all the coefficients and variables are still exactly the same, but now the spatial layout of this simulated data has changed into the one shown in Figure 4b.

### 3.1.3. Scenario 3: simulation of random relationships on a regular grid

Scenario 3 acts as a control or benchmark simulation, where we create an entirely 'unknown' pattern in the regression coefficients on a regular grid. Thus with a random variable,  $x_i$ , and random coefficients,  $\beta_i$ , we generate the dependent variable,  $y_i$ , as follows:

$$y_i = \beta_1(u_i, v_i)x_i \quad (6)$$

where the surfaces for the predictor variable and the regression coefficients  $\beta_1$  are shown in Figure 5. No intercept is generated in this case.

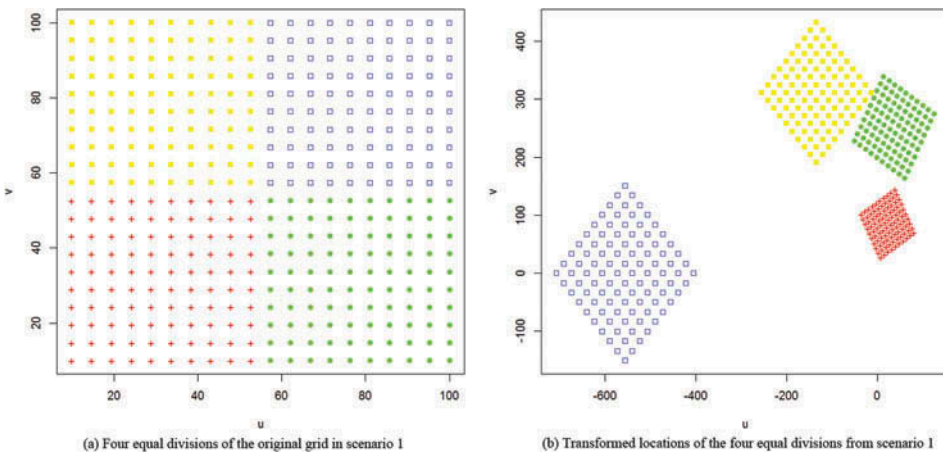
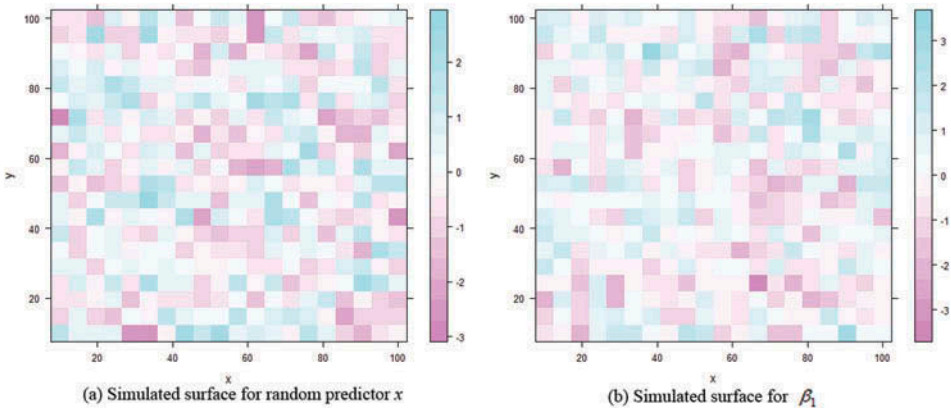


Figure 4. Scenario 2: anisotropic and disjoint transformations applied to the scenario 1 simulation.



**Figure 5.** Scenario 3: visualization of the simulated predictor variable and the coefficients.

### 3.2. GWR with Minkowski and Euclidean distances

With the simulated data, we calibrate GWR models using Minkowski distances (MD-GWR), each with their own optimally found bandwidth and MD parameters. As it is impractical to try every possible combination of  $p$  and  $\theta$ , we limit ourselves to

- (1)  $p = \{0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75, 3, 3.25, 3.5, 3.75, 4, 4.25, 4.5, 4.75, 5, 5.25, 5.5, 5.75, 6, 6.25, 6.5, 6.75, 7, 7.25, 7.5, 7.75, 8, \infty\}$
- (2)  $\theta$  from 0 to  $\frac{\pi}{2}$  sampled in nine equal intervals of  $\frac{\pi}{18}$  (in radians).

Results are given in terms of improvement in AICc with respect to using a particular MD-GWR calibration over that found with a corresponding basic GWR calibration using the usual ED metric (ED-GWR). From [Figure 6](#), entries coloured green indicate where a significant improvement is found using an MD metric (i.e. the AICc difference is less than  $-3$ ); entries coloured black are for AICc differences of between  $-3$  and  $+3$  which are taken to indicate an insignificant difference in model fit; entries coloured red indicate where the ED metric should be preferred (i.e. the AICc difference is greater than  $+3$ ). The use of three units as our AICc cut-off value is discussed in [Fotheringham \*et al.\* \(2002, pp. 69–70\)](#). The rows in [Figure 6](#) correspond to different rotated angles  $\theta$ , while the columns correspond to different values of  $p$ . These descriptions and rules are similarly applied to all related figures.

As shown in [Figure 6](#), the choice of distance metric clearly influences GWR model fit; different structures of relationship non-stationarity associate with distinct metrics. From [Figure 6a](#) for scenario 1, MD-GWR can be preferred when (1) the rotated angle  $\theta$  is 0, 1.22 and 1.4,<sup>1</sup> and the value of  $p$  lies between 0.75 and 2; and (2) the rotated angle is 0.35, 0.52 and 0.7, and the value of  $p$  is greater than 2. From [Figure 6b](#) for scenario 2 (the asymmetrically rotated and rescaled partitioned data from scenario 1), the preferred MD-GWR cases are when (1) the rotated angle  $\theta$  is 0, 0.17 and 1.4, and the value of  $p$  is larger than 2; (2) the rotated angle  $\theta$  is 0.7 and 0.87, and the value of  $p$  lies between 0.5 and 2. These scenarios show similar responses to varying values of  $p$ , in that the ‘best performing’ distance (BPD) metrics for both are reached at same value of  $p$  (i.e. *infinity*, the Chebyshev distance); and non-improvements occur at  $p = 0.25$ . In contrast, these scenarios present diverse responses to the rotated angles  $\theta$ . These similarities and



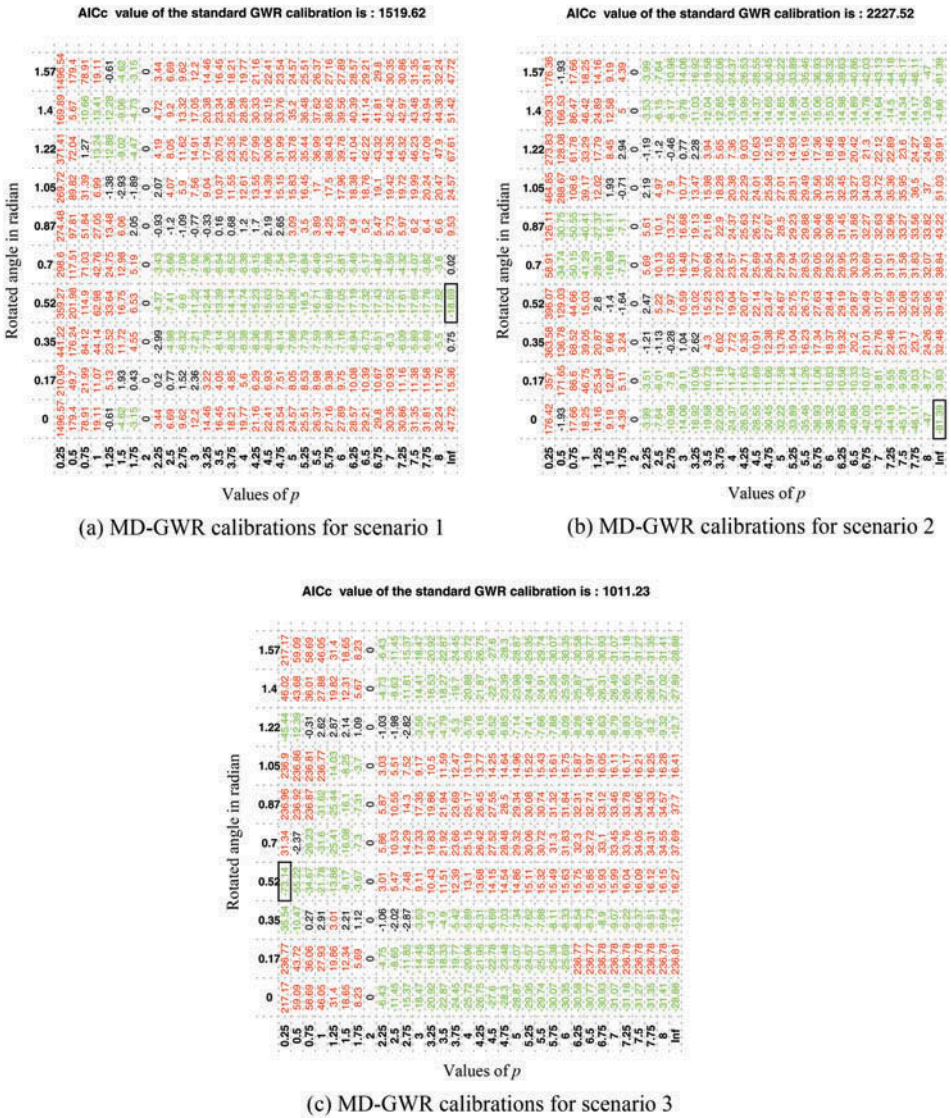


Figure 6. AICc differences between ED-GWR and MD-GWR for different sets of  $(p, \theta)$ .

differences directly relate to the constructions of the two simulations. From Figure 6c for scenario 3 (unknown and random variation in the coefficients), MD-GWR can be preferred for cases when (1) the rotated angle  $\theta$  is from 0.35 to 1.22, and the value of  $p$  is less than 2; and (2) the rotated angle is 0, 0.17, 0.35, 1.22 and 1.4, and the value of  $p$  is greater than 2. The BPD comes out to be 0.25 with the rotated angle  $\theta$  as 0.52.

### 3.3. GWR with the BPD metrics

#### 3.3.1. GoF comparisons

The GoF outputs for the ED-GWR calibrations and the GWR calibrations using the BPD metrics (BPD-GWR), together with the global OLS regression results, are given in Table 1.

**Table 1.** GoF diagnostics for OLS regression, ED-GWR and BPD-GWR in the three scenarios.

Scenario	OLS		ED-GWR			BPD-GWR			
	RSS	AICc	Bandwidth	RSS	AICc	$(p, \theta)$	Bandwidth	RSS	AICc
1	211261	3648.97	4.64	395.52	1519.62	$(Inf, 0.52)$	4.16	376.94	1500.94
2			14.56	1544.61	2227.52	$(Inf, 0)$	12.41	1343.97	2146.12
3	525	1247.98	2.68	79.23	1011.23	$(0.25, 0.52)$	22.33	53.26	938.10

As would be expected, all GWR fits provide an improvement over the corresponding OLS regression. For the GWR fits, the BPD-GWR calibrations all use non-ED metrics (i.e. all are MD-GWR fits) and show significant improvements in terms of AICc and the residual sum of squares (RSS) over their ED-GWR counterparts.

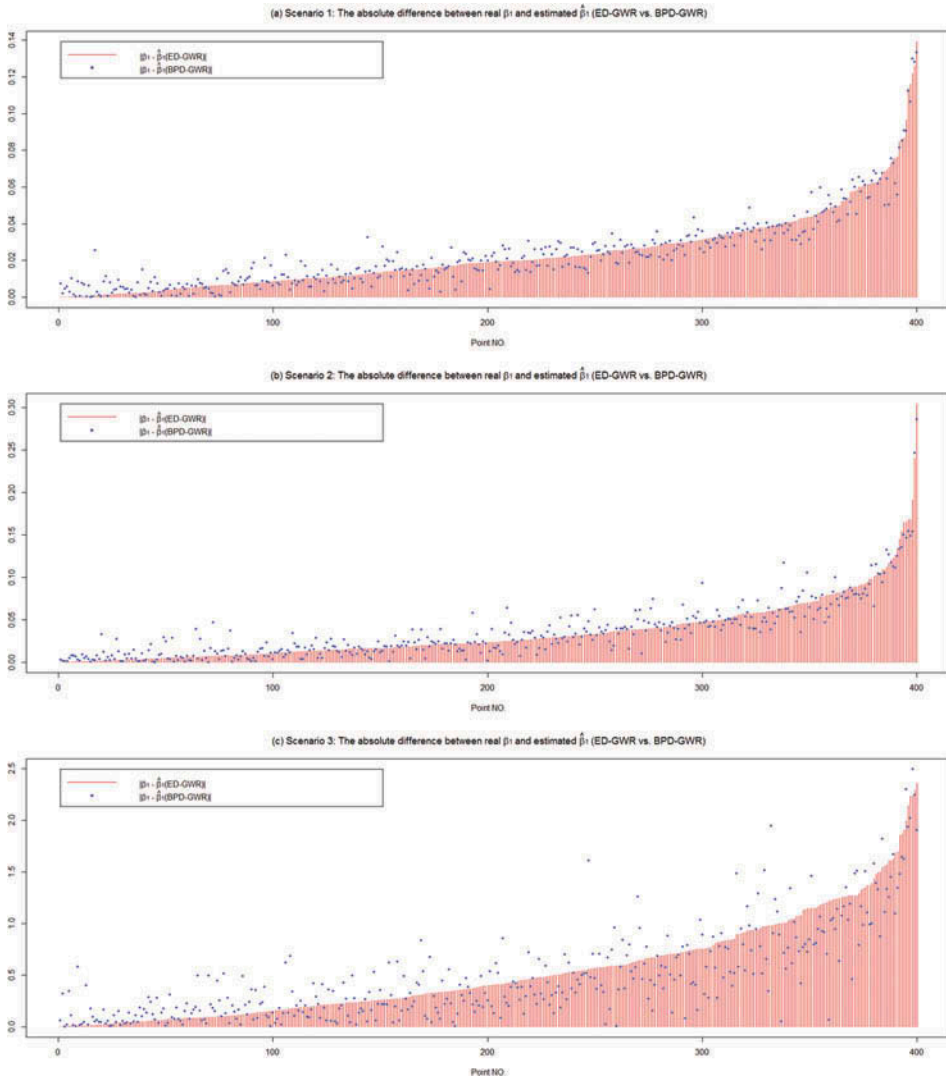
For scenario 1, the BPD-GWR model uses an MD at  $p = Inf$  and  $\theta = 0.52$  (i.e. Chebyshev distance with coordinates rotated by 0.52 in radians). This calibration shows a reduction in AICc of 18.71 from its ED-GWR counterpart, and the RSS is also reduced by 4.8%. It is hypothesized that the main reason for an MD-GWR fit to outperform its ED-GWR counterpart is due to the asymmetrical nature of the coefficient surface for  $\beta_0$  for this scenario (see Figure 3). For scenario 2, and as expected, a much greater improvement is seen by using MD-GWR. This MD-GWR model uses the same Chebyshev distance ( $p = Inf$ ) but without rotation ( $\theta = 0$ ), as that found in scenario 1. The MD-GWR calibration shows a reduction in AICc of 81.39 from its ED-GWR counterpart and the RSS is also reduced by 13%. Thus, an MD-GWR model is considered a superior GWR calibration for detecting anisotropic and disjoint data relationships than that found with the ED-GWR model.

For scenario 3, the simulation design represents an 'arbitrary' set of laws in how data relationships vary across space. For this scenario, the BPD-GWR model uses an MD with  $p = 0.25$  and  $\theta = 0.52$ . This MD-GWR model outperforms the corresponding ED-GWR model as the AICc value is reduced by 73.13 and RSS is reduced by 32.8%. This result suggests that the Minkowski approach can be useful when there is absolutely no knowledge, understanding or insight into the 'true' distance metric.

### 3.3.2. Comparisons of the estimated coefficients

As with any GWR study, it is important to interpret the localized regression coefficients, where for this study we need to ascertain the accuracy of our estimated coefficients with respect to the known (or real) ones that we have simulated. Here, Figure 7 displays the (ordered) absolute differences between the real values of  $\beta_1$  and the estimates  $\hat{\beta}_1$ , for the ED-GWR and BPD-GWR calibrations, at all 400 locations, across all three scenarios. In Table 2, we summarize the same results from Figure 7 via (a) the RSS between  $\beta_1$  and  $\hat{\beta}_1$  (call this, RSS-C) and (b) the number of locations where one GWR calibration provides a coefficient estimate that is closer to the real coefficient than the other GWR calibration (i.e. the number of winning locations, call this, NWL).

From Figure 7, both GWR calibrations appear to make highly accurate estimations of  $\beta_1$  for scenarios 1 and 2, with the BPD-GWR model performing moderately better in both scenarios (as confirmed in Table 2). For scenario 3, the BPD-GWR model makes impressively better coefficient estimates than the ED-GWR model, where the RSS-C is reduced by 15% and where the BPD-GWR model provides more accurate coefficient estimates at over 60% of the simulation locations. This suggests that an MD-GWR model can provide



**Figure 7.** Comparisons of the real  $\beta_1$  and estimated  $\hat{\beta}_1$  respectively from ED-GWR and BPD-GWR calibrations: (a) scenario 1, (b) scenario 2 and (c) scenario 3.

**Table 2.** Quantitative descriptions for estimated  $\hat{\beta}_1$  from ED-GWR and BPD-GWR calibrations in the three scenarios.

Scenario	$\hat{\beta}_1$ (ED-GWR)		$\hat{\beta}_1$ (BPD-GWR)	
	RSS-C	NWL	RSS-C	NWL
1	0.40	195	0.38	205
2	0.84	192	0.74	208
3	201.91	159	169.91	241

more accurate coefficient estimates when the ‘optimum’ distance metric is uncertain, and thus provide a truer picture of relationship non-stationarity.

#### 4. Reducing computational overheads

The simulation-based results show that an MD metric is useful in improving the GWR fit in terms of its overall GoF and also in terms of its coefficient accuracy. However, our approach uses brute force to search for the BPD metric for a given GWR fit. For the chosen values of  $p$  and rotation angles  $\theta$ , 321 GWR calibrations<sup>2</sup> were necessary for each simulation scenario, calibrations that each included an additional ‘nested’ optimization to select an optimum bandwidth,  $b$ . All such calculations are computationally demanding, especially when there exists a large number of sample observations and/or local regression calibration locations. However, we can reduce such costs by taking more care in choosing the intervals for  $p$  and  $\theta$ . For example, from [Figure 6](#), it is observed that

- (1) The change in AICc can be very small between adjacent values of  $p$ . For example, AICc changes little ( $<1$ ) across  $p$  values of 7, 7.25, 7.5, 7.75 and 8 for scenario 3.
- (2) As  $p$  increases away from 2 (i.e. ED), AICc is more sensitive to changes in the rotation angle  $\theta$  for all three scenarios.

##### 4.1. Heuristics to reduce computational cost

With these observations in mind, we can reduce computational cost using this ‘forward direction’ procedure:

- (1) Given an interval for  $p$ :  $[p_{\min}, p_{\max}]$ , a searching pace  $s$ , an AICc difference threshold of  $t = 3$  (as discussed in [Section 3.2](#)), and a fixed rotation angle  $\theta_0$ ; calculate the AICc values with the MDs respectively at  $(p_{\min}, \theta_0)$  and  $(p_{\max}, \theta_0)$ , namely  $AICc_{p_{\min}}$  and  $AICc_{p_{\max}}$ .
- (2) Set the iteration number  $i = 0$ ; include  $p_{\min}$  in the vector  $P_{\text{sel}}$  for holding the values of  $p$  to be used in the Minkowski approach and set the AICc value with the latterly selected  $p$ :  $AICc_{\text{last\_sel}} = AICc_{p_{\min}}$ .
- (3) If the absolute difference between  $AICc_{\text{last\_sel}}$  and  $AICc_{p_{\max}}$  is less than  $t$ , terminate; otherwise, include  $p_{\max}$  in  $P_{\text{sel}}$ , and start the iteration:
  - (i) Set  $i = i + 1$ ,  $p_i = p_{\min} + i \times s$
  - (ii) Calculate AICc with MD at  $(p_i, \theta_0)$ , name it as  $AICc_{p_i}$ .
  - (iii) If the absolute differences between  $AICc_{\text{last\_sel}}$  and  $AICc_{p_i}$ ,  $AICc_{\text{last\_sel}}$  and  $AICc_{p_{\max}}$  are not less than  $t$ , include  $p_i$  in  $P_{\text{sel}}$  and set  $AICc_{\text{last\_sel}} = AICc_{p_i}$ .
- (4) If the absolute difference between  $AICc_{\text{last\_sel}}$  and  $AICc_{p_{\max}}$  is less than  $t$  or  $p_{i+1}$  is larger than  $p_{\max}$ , terminate; otherwise, continue the iteration.

Similarly, computational cost can be reduced using this ‘backward direction’ procedure:

- (1) Repeat step 1 in the ‘forward direction’ procedure for the initial settings.

- (2) Set the iteration number  $i = 0$ ; include  $p_{\max}$  in the vector  $P_{\text{sel}}$  for holding the values of  $p$  to be used in the Minkowski approach and set the AICc value with the latterly selected  $p$ :  $\text{AICc}_{\text{last\_sel}} = \text{AICc}_{p_{\max}}$ .
- (3) If the absolute difference between  $\text{AICc}_{\text{last\_sel}}$  and  $\text{AICc}_{p_{\min}}$  is less than  $t$ , terminate; otherwise, include  $p_{\min}$  in  $P_{\text{sel}}$ , and start the iteration:
  - (i) Set  $i = i + 1$ ,  $p_i = p_{\min} + i \times s$ .
  - (ii) Calculate AICc with MD at  $(p_i, \theta_0)$ , name it as  $\text{AICc}_{p_i}$ .
  - (iii) If the absolute differences between  $\text{AICc}_{\text{last\_sel}}$  and  $\text{AICc}_{p_i}$ ,  $\text{AICc}_{\text{last\_sel}}$  and  $\text{AICc}_{p_{\min}}$  are not less than  $t$ , include  $p_i$  in  $P_{\text{sel}}$  and set  $\text{AICc}_{\text{last\_sel}} = \text{AICc}_{p_i}$ .
- (4) If the absolute difference between  $\text{AICc}_{\text{last\_sel}}$  and  $\text{AICc}_{p_{\min}}$  is less than  $t$  or  $p_{i+1}$  is less than  $p_{\min}$ , terminate; otherwise, continue the iteration.

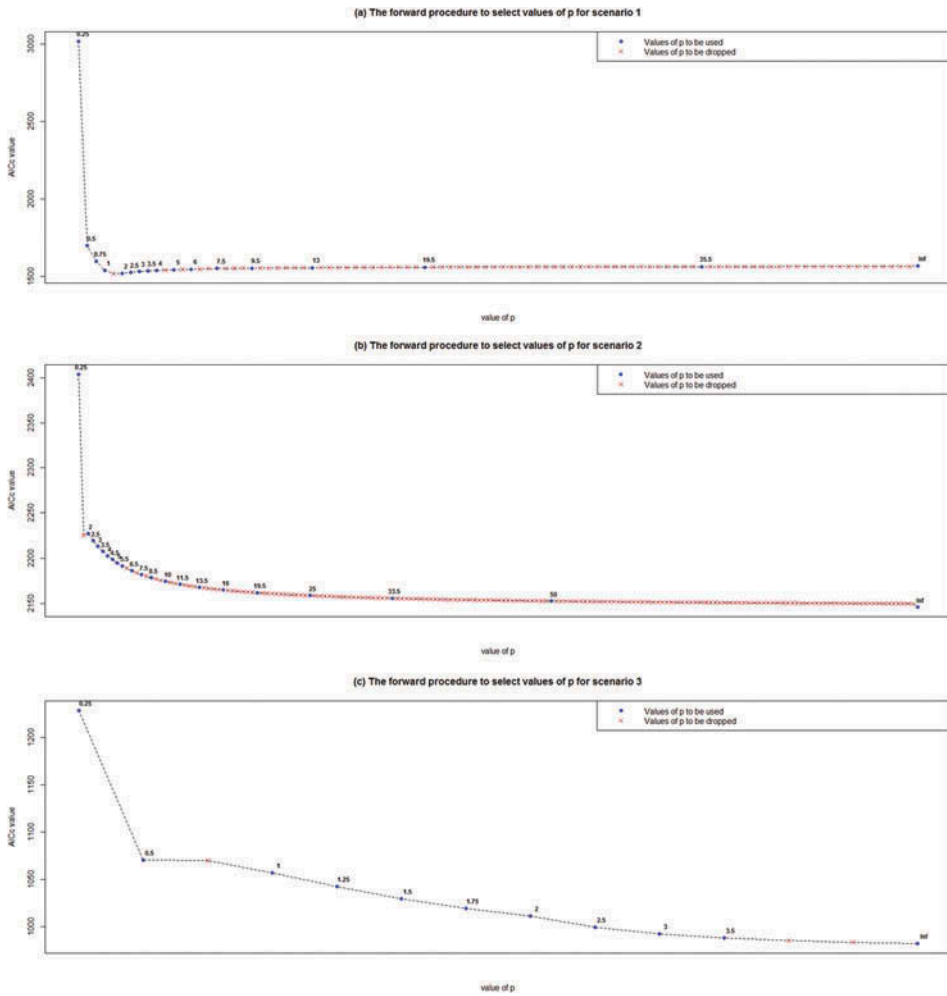
Generally, there is little computational difference between the described procedures, but only the ‘forward direction’ can be used when  $p_{\max}$  is *infinity*. Observe also that the greater the value of  $p_{\min}$ , the greater the value of pace  $s$  could be set. For the rotation angles  $\theta$ , further computational reductions are possible if we sample at varying densities on the interval  $[0, \frac{\pi}{2})$  for a specific  $p$ , that is, given the results in Figure 6, sparse sampling for the values of  $p$  close to 2; but intensive sampling for the values far from 2 is appropriate.

#### 4.2. Revisiting the simulation experiment

We can now rerun this study’s simulation experiment using the described procedures. Firstly, we ran the ‘forward direction’ procedure to choose the values of  $p$  for the three scenarios. For each scenario, we made the selections on two successive intervals:  $[0.25, 2]$  and  $(2, \infty)$ . We also used different paces  $s$  for each interval: 0.25 for the former interval and 0.5 for the latter one, but the same threshold  $t = 3$  and a fixed rotation angle  $\theta_0 = 0$ .

Figure 8 displays the search results for the three scenarios, where blue points denote the values of  $p$  used for the MD-GWR calibrations, whereas red points denote those dropped. Figure 8a and b are similar, as the data sets in scenarios 1 and 2 are closely related, while Figure 8c for scenario 3 presents a very different process. The results indicate that setting  $s$  is important; where in this case, we used  $s = 0.5$  on the interval  $(2, \infty)$  for all three scenarios. This value of  $s$  seems too small for scenarios 1 and 2, where 279 values of  $p$  larger than 2 are tried, but more than 90% are dropped. In contrast, it performs well for scenario 3, where values of  $p$  are chosen efficiently. With the selected values of  $p$ , we define a varying number of arithmetical sequences on  $[0, \frac{\pi}{2})$  for sampling  $\theta$  instead of 9 for all: 1 is defined for  $p = 2$  (i.e. only the rotation angle  $\theta = 0$  is used for ED); and increasing numbers are used when the value of  $p$  increases from 2. The full set of values for  $p$  and  $\theta$  for the MD-GWR calibrations is given in Table 3.

For each scenario, the total number of model calibrations has significantly decreased in comparison to the exhaustive version (i.e. 321) from before. Figure 9 displays the AICc values from the MD-GWR calibrations for all three scenarios, where points are coloured

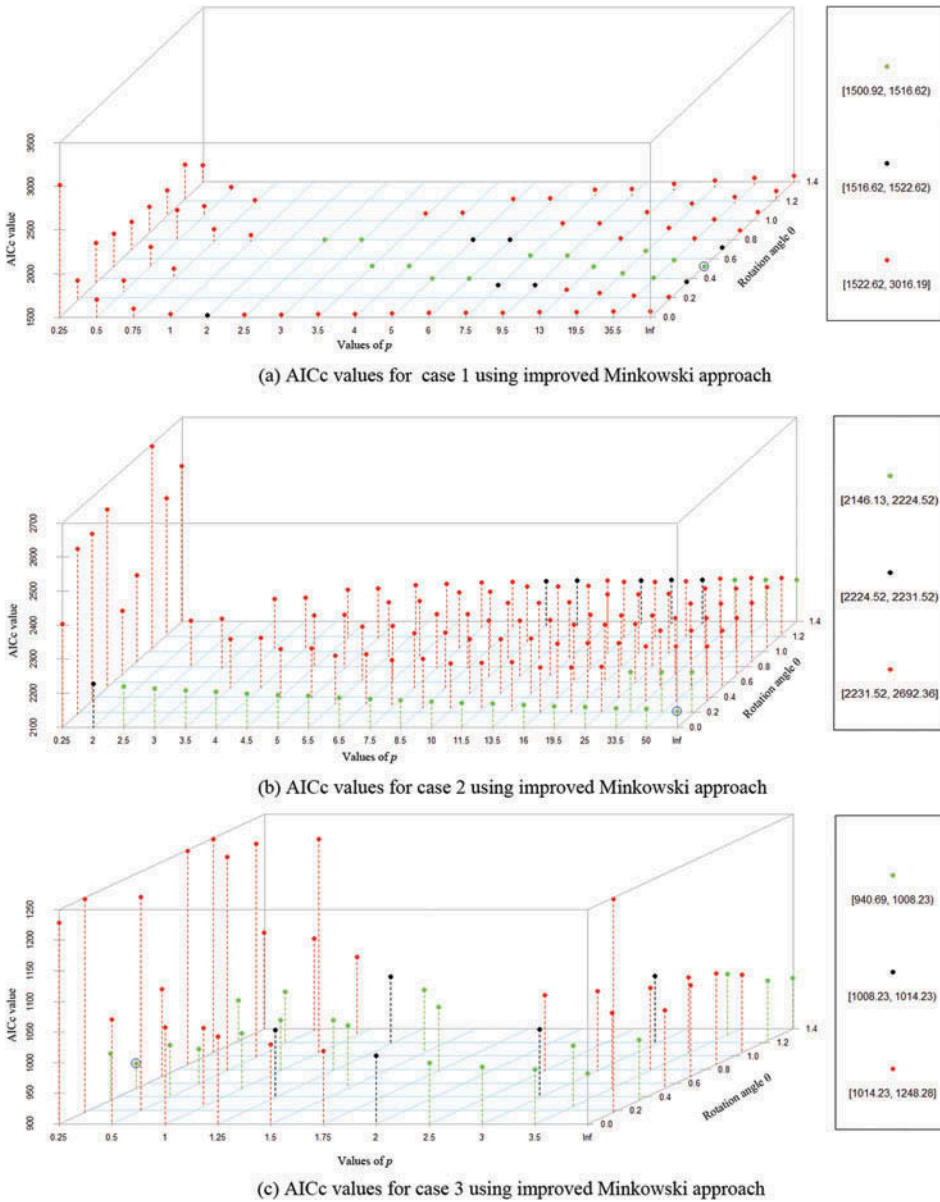


**Figure 8.** The results from the ‘forward direction’ procedure for (a) scenario 1, (b) scenario 2 and (c) scenario 3.

**Table 3.** Information of improved Minkowski approach for the three scenarios.

Scenario	$p$ -Values	Number of arithmetical sequences for sampling $\theta$	Total number of calibrations
1	{0.25, 0.5, 0.75, 1, 2, 2.5, 3, 3.5, 4, 5, 6, 7.5, 9.5, 13, 19.5, 35.5, $Inf$ }	{9, 6, 4, 2, 1, 2, 2, 3, 3, 4, 4, 5, 5, 6, 7, 8, 9}	80
2	{0.25, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6.5, 7.5, 8.5, 10, 11.5, 13.5, 16, 19.5, 25, 33.5, 50, $Inf$ }	{9, 1, 2, 2, 3, 3, 4, 4, 5, 5, 6, 6, 7, 7, 7, 8, 8, 8, 9, 9, 9}	122
3	{0.25, 0.5, 1, 1.25, 1.5, 1.75, 2, 2.5, 3, 3.5, $Inf$ }	{9, 8, 6, 4, 3, 2, 1, 2, 4, 6, 9}	54

red, black and green. The colouring scheme partitions are defined as the AICc value from the ED-GWR calibration  $\pm 3$  (i.e. analogous to that used in Section 3 and Figure 6). As expected, the BPD metrics correspond to those found before (for which the corresponding point is circled).



**Figure 9.** AICc values from the MD-GWR calibrations using the information in Table 3.

In summary, the results provide value in using the described set of heuristics for improving computational efficiency for MD metric selection in GWR. However, there are caveats, primarily (1) different choices of  $\theta_0$  and  $s$  within both searching procedures may cause diverse values of  $p$ , so how to preset these values needs further work; and (2) varying numbers of arithmetical sequences were used for sampling  $\theta$  within the MD-GWR calibrations, and here a more coherent set of rules needs to be established (in addition to the concerns raised on the disparity of the values of  $p$  moving away from 2).

## 5. Discussion and concluding remarks

In this study, we have demonstrated a Minkowski approach to approximate the underlying distance metric for calibrating a GWR model. Via a simulation experiment, we have shown that this approach can improve the GoF of a GWR model. Such a GWR model can also provide more accurate coefficient estimates, especially when the ideal distance metric is uncertain, thus potentially providing a truer picture of relationship non-stationarity. These positive results may be particularly useful when there is insufficient information for calculating a suitable distance, like TT or ND. The Minkowski approach is not viewed as a negation of the standard GWR technique with EDs, but more a useful option, especially in urban applications of GWR.

The Minkowski approach to GWR calibration can be computationally demanding, and as such, care needs to be taken in choosing suitable candidate intervals for the Minkowski distance parameters  $p$  and  $\theta$ . Here,  $p$  can be chosen to approximate known distance types. For example, Miller and Wentz (2003) show that TT in an urban setting could be represented by  $p$  values ranging from 1 to 2. Furthermore, as each individual GWR calibration is independent from each other, it would be straightforward to apply parallel computing techniques (e.g. Harris *et al.* 2010c) to reduce computational costs.

However, there exists a key drawback to the Minkowski approach, where it is often difficult to sketch out how a specific Minkowski distance function measures in our mind, except for some common cases, such as the Manhattan ( $p = 1$ ) or ED ( $p = 2$ ). The rotation of coordinate system (using  $\theta$ ) also adds to this perception difficulty. This drawback tends to make the Minkowski approach more suitable for prediction purposes with GWR (see Harris *et al.* 2010a) rather than exploratory/inferential purposes with GWR as the latter is more concerned with *understanding* relationship non-stationarity.

Finally, it is important to note that our extension of GWR is still liable to the same issues that basic GWR may suffer from, such as (1) model misspecification (e.g. Griffith 2008) and (2) local collinearity (Páez *et al.* 2011), which is somewhat dependent on issue (1). Further work could investigate these key issues in GWR when a non-ED metric is chosen.

## Notes

1. The period of Equation (5) is  $\frac{\pi}{2}$ , that is, very close to 1.57, so a rotation of 1.57 will be almost no different from a 0 rotation.
2. Coordinate rotations are invalid for ED, so the model is only calibrated once, when  $p$  is 2.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

Research presented in this paper was supported by projects from the National Natural Science Foundation of China [NSFC: 41401455], [NSFC: 41331175]; a Strategic Research Cluster [grant number 07/SRC/I1168] by the Science Foundation Ireland under the National Development Plan; and the North Wyke Farm Platform supported by the Biotechnology and Biological Sciences Research Council of the UK [BBSRC BB/J004308].



**ORCID**

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 Tsahkadzor. 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
- Boisvert, J.B. and Deutsch, C.V., 2011. Programs for kriging and sequential Gaussian simulation with locally varying anisotropy using non-Euclidean distances. *Computers & Geosciences*, 37 (4), 495–510. doi:10.1016/j.cageo.2010.03.021
- Brunsdon, C., Fotheringham, A.S., and Charlton, M., 2000. *Geographically weighted regression as a statistical model*. Newcastle-upon-Tyne: Spatial Analysis Research Group.
- 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
- Brunsdon, C., Fotheringham, S., and Charlton, M., 1998. Geographically weighted regression-modelling spatial non-stationarity. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47 (3), 431–443. doi:10.1111/rssd.1998.47.issue-3
- 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
- Elhorst, J., 2003. Specification and estimation of spatial panel data models. *International Regional Science Review*, 26 (3), 244–268. doi:10.1177/0160017603253791
- 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., Charlton, M.E., and Brunsdon, C., 1998. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*, 30 (11), 1905–1927. doi:10.1068/a301905
- 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.
- 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, P., et al., 2010a. The use of geographically weighted regression for spatial prediction: an evaluation of models using simulated data sets. *Mathematical Geosciences*, 42 (6), 657–680. doi:10.1007/s11004-010-9284-7
- Harris, P., Fotheringham, A.S., and Juggins, S., 2010b. Robust geographically weighted regression: a technique for quantifying spatial relationships between freshwater acidification critical loads and catchment attributes. *Annals of the Association of American Geographers*, 100 (2), 286–306. doi:10.1080/00045600903550378
- Harris, R., et al., 2010c. Grid-enabling geographically weighted regression: a case study of participation in higher education in England. *Transactions in GIS*, 14 (1), 43–61. doi:10.1111/tgis.2010.14.issue-1
- 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/rssb.1998.60.issue-2

- Jones, K., 1991. Specifying and estimating multi-level models for geographical research. *Transactions of the Institute of British Geographers*, 16 (2), 148–159. doi:[10.2307/622610](https://doi.org/10.2307/622610)
- Kent, J., Leitner, M., and Curtis, A., 2006. Evaluating the usefulness of functional distance measures when calibrating journey-to-crime distance decay functions. *Computers, Environment and Urban Systems*, 30 (2), 181–200. doi:[10.1016/j.compenvurbsys.2004.10.002](https://doi.org/10.1016/j.compenvurbsys.2004.10.002)
- Kullback, S. and Leibler, R.A., 1951. On information and sufficiency. *The Annals of Mathematical Statistics*, 22 (1), 79–86. doi:[10.1214/aoms/1177729694](https://doi.org/10.1214/aoms/1177729694)
- Legleiter, C.J. and Kyriakidis, P.C., 2008. Spatial prediction of river channel topography by kriging. *Earth Surface Processes and Landforms*, 33 (6), 841–867. doi:[10.1002/\(ISSN\)1096-9837](https://doi.org/10.1002/(ISSN)1096-9837)
- Love, R.F., Morris, J.G., and Wesolowsky, G.O., 1988. *Facility location: models and methods*. New York, NY: North-Holland.
- 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, 92–97. doi:[10.1016/j.proenv.2011.07.017](https://doi.org/10.1016/j.proenv.2011.07.017)
- Lu, B., Charlton, M., and Harris, P., 2012. Geographically weighted regression using a non-Euclidean distance metric with simulation data. In: *2012 first international conference on agro-geoinformatics*, 2–4 August Shanghai. IEEE, 1–4.
- 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](https://doi.org/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](https://doi.org/10.1080/10095020.2014.917453)
- Miller, H.J. and Wentz, E.A., 2003. Representation and spatial analysis in geographic information systems. *Annals of the Association of American Geographers*, 93 (3), 574–594. doi:[10.1111/1467-8306.9303004](https://doi.org/10.1111/1467-8306.9303004)
- Money, E.S., Carter, G.P., and Serre, M.L., 2009. Modern space/time geostatistics using river distances: data integration of turbidity and E. coli measurements to assess fecal contamination along the Raritan River in New Jersey. *Environmental Science & Technology*, 43 (10), 3736–3742. doi:[10.1021/es803236j](https://doi.org/10.1021/es803236j)
- 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](https://doi.org/10.1111/gean.2004.36.issue-4)
- Páez, A., Farber, S., and Wheeler, D., 2011. A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. *Environment and Planning A*, 43 (12), 2992–3010. doi:[10.1068/a44111](https://doi.org/10.1068/a44111)
- R Development Core Team, 2013. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Shahabi, C., Kolahdouzan, M.R., and Sharifzadeh, M., 2002. A road network embedding technique for k-nearest neighbor search in moving object databases. In: *Proceedings of the 10th ACM international symposium on advances in geographic information systems*, 8–9 November. McLean, VA: ACM, 94–100.
- Shahid, R., et al., 2009. Comparison of distance measures in spatial analytical modeling for health service planning. *BMC Health Services Research*, 9 (1), 200–214. doi:[10.1186/1472-6963-9-200](https://doi.org/10.1186/1472-6963-9-200)
- Wang, N., Mei, C.-L., and Yan, X.-D., 2008. Local linear estimation of spatially varying coefficient models: an improvement on the geographically weighted regression technique. *Environment and Planning A*, 40 (4), 986–1005. doi:[10.1068/a3941](https://doi.org/10.1068/a3941)