



## The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models

Binbin Lu, Paul Harris, Martin Charlton & Chris Brunsdon

To cite this article: Binbin Lu, Paul Harris, Martin Charlton & Chris Brunsdon (2014) 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)

To link to this article: <https://doi.org/10.1080/10095020.2014.917453>



© 2014 Wuhan University



View supplementary material [↗](#)



Published online: 28 May 2014.



Submit your article to this journal [↗](#)



Article views: 5764



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 61 View citing articles [↗](#)

## The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models

Binbin LU<sup>a\*</sup>, Paul HARRIS<sup>b</sup>, Martin CHARLTON<sup>c</sup> and Chris BRUNSDON<sup>c</sup>

<sup>a</sup>*School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China;* <sup>b</sup>*Rothamsted Research, North Wyke, Okehampton, Devon, UK;* <sup>c</sup>*National Centre for Geocomputation, National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland*

(Received 16 November 2013; final version received 12 April 2014)

In this study, we present a collection of local models, termed geographically weighted (GW) models, which can be found within the **GWmodel** R package. A GW model suits situations when spatial data are poorly described by the global form, and for some regions the localized fit provides a better description. The approach uses a moving window weighting technique, where a collection of local models are estimated at target locations. Commonly, model parameters or outputs are mapped so that the nature of spatial heterogeneity can be explored and assessed. In particular, we present case studies using: (i) GW summary statistics and a GW principal components analysis; (ii) advanced GW regression fits and diagnostics; (iii) associated Monte Carlo significance tests for non-stationarity; (iv) a GW discriminant analysis; and (v) enhanced kernel bandwidth selection procedures. General Election data-sets from the Republic of Ireland and US are used for demonstration. This study is designed to complement a companion **GWmodel** study, which focuses on basic and robust GW models.

**Keywords:** principal components analysis; semi-parametric GW regression; discriminant analysis; Monte Carlo tests; election data

### 1. Introduction

In this study, we present a collection of local (non-stationary) statistical models, termed geographically weighted (GW) models (1). A GW model suits situations when spatial data are poorly described by the global (stationary) model form, and for some regions a localized fit provides a better description. This type of approaches uses a moving window weighting technique, where a collection of local models are found at target locations. Commonly, outputs or parameters of a GW model are mapped to provide a useful exploratory tool that can direct a more traditional or sophisticated statistical analysis. For example in a regression context, GW regression (2–6) can be used to explore relationships in the data. If relationships are deemed stationary across space, then a basic (non-spatial) regression or a regression that accounts for some spatial autocorrelation effect (e.g. (7)) is sufficient. Conversely, if relationships are deemed non-stationary, the GW regression can be replaced with a sophisticated, spatially varying coefficient model for improved inference, such as those proposed by Gelfand et al. (8) or Assunção (9).

Other notable GW models include: GW summary statistics (1, 10, 11); GW distribution analysis (12); GW principal components analysis (GW PCA) (1, 13); GW generalized linear models (1, 14); GW discriminant analysis (GWDA) (15); GW-Geostatistical hybrids (16–20); GW methods for outlier detection (21, 22); and GW

methods for network re-design (23). The GW modeling framework continues to evolve (24) and GW models have been usefully applied to data from a wide range of disciplines in the natural and social sciences.

Many of the listed GW models are included in an R package **GWmodel** (<http://www.r-project.org>). Notably, **GWmodel** provides functions to conduct: (i) a GW PCA; (ii) advanced GW regression fits and diagnostics; (iii) associated Monte Carlo significance tests for non-stationarity; (iv) a GW DA; and (v) enhanced bandwidth selection procedures; where all such functions are utilized in this study. In this respect, our study complements a companion **GWmodel** study (25), which focused on basic and robust GW models. Our study is structured as follows. Section 2 sets the scene, describing the specification of the weighting matrix in a GW model and the case study data-sets (General Election data for the Republic of Ireland and the US). Section 3 describes the use of GW summary statistics and a GW PCA, together with associated Monte Carlo tests. Section 4 describes the fitting of a mixed (semi-parametric) GW regression. Section 5 investigates further topics in GW regression; including: (a) multiple hypothesis tests, (b) collinearity diagnostics, and (c) the fitting of heteroskedastic models. Section 6 describes a GW DA. Section 7 describes enhanced bandwidth selection procedures for any GW model. Section 8 concludes this work. All of the analyses in this study can be reproduced with an associated R script

\*Corresponding author. Email: [binbin.lu@nuim.ie](mailto:binbin.lu@nuim.ie)

(see the online Section of Supplementary Material, which is also available at the GWR website (<http://gwr.nuim.ie/>)). Ideally, the R script should be run in conjunction with the reading of this article.

## 2. Context, the weighting matrix, and the study data-sets

### 2.1. Context

GW methods are primarily used to investigate spatial heterogeneity, where the form of the heterogeneity reflects the objective of the under-lying statistic or model. For example, a GW standard deviation (from GW summary statistics) investigates spatial change in data variability; a GW regression investigates spatial change in response and predictor data relationships; a GW variogram (16) investigates spatial change in spatial dependence. In all cases, a moving window weighting technique is used, where local models are calibrated at (sampled or un-sampled) locations (i.e. the window's center). For an individual model at some calibration location, all neighboring observations are weighted according to the properties of a distance-decay kernel function, and the model is locally fitted to this weighted data. Thus, the geographical weighting solely applies to the data in all GW methods, where each local model is fitted to its own GW data (sub-) set. The size of the window over which this localized model might apply is controlled by the kernel function's bandwidth. Small bandwidths lead to more rapid spatial variation in the results, while large bandwidths yield results increasingly close to the global model solution. The GW modeling paradigm encompasses many methods; each locally adapted from a global form.

### 2.2. Building the weighting matrix

Key to GW modeling is the weighting matrix, which sets the spatial dependency in the data. Here,  $W(u_i, v_i)$  is a  $n \times n$  diagonal matrix (where  $n$  is the sample size) denoting the geographical weighting attached to each observation point, for any model calibration point  $i$  at location  $(u_i, v_i)$ . Thus, a different weighting matrix is found at each calibration point. This matrix is determined according to the following three key elements: (i) the type of distance metric; (ii) the type of kernel weighting function; and (iii) the kernel weighting function's bandwidth. For the distance metric, **GWmodel** permits the Minkowski family of distance metrics, where the power of the Minkowski distance  $p$ , needs to be specified. For example, if  $p=2$  (the default), then the usual Euclidean distance metric results; or if  $p=1$ , then the Manhattan (or Taxi cab) distance metric results. As an example, Lu et al. (26) fit GW regressions using Euclidean and non-Euclidean distance metrics.

Four of the five kernel weighting functions available in **GWmodel** are defined in Table 1. Each function

Table 1. Four kernel weighting functions.

Box-car	$w_{ij} = 1$ , if $d_{ij} \leq r$ ; $w_{ij} = 0$ , otherwise
Bi-square	$w_{ij} = (1 - (d_{ij}/r)^2)^2$ , if $d_{ij} \leq r$ ; $w_{ij} = 0$ , otherwise
Gaussian	$w_{ij} = \exp(-d_{ij}^2/2b^2)$
Exponential	$w_{ij} = \exp(-d_{ij}/b)$

includes the bandwidth parameter ( $r$  or  $b$ ), which controls the rate of decay. All functions are defined in terms of weighting the sample data, where  $j$  is the index of the observation point and  $d_{ij}$  the distance between the points indexed by  $i$  and  $j$ . For the box-car and bi-square functions, the bandwidth  $r$  can be specified beforehand (i.e. a fixed distance) or specified as the distance between the point  $i$  and its  $N$ th nearest neighbor, where  $N$  is specified beforehand (i.e. an adaptive distance). The bi-square function gives fractional decaying weights according to the proximity of the data to each point  $i$ , up until a fixed distance or a distance according to a specified  $N$ th nearest neighbor. The local search strategy for this and the box-car function is simply  $N$  neighbors within a fixed radius  $r$  or  $N$  nearest neighbors for an adaptive approach. Both functions can suffer from discontinuity, although the bi-square function can be defined with a bandwidth that uses all of the data to minimize such problems.

The Gaussian and exponential functions are continuous and use all the data. Their weights decay according to a Gaussian or exponential curve. According to the bandwidth set, data that are a long way from the point  $i$  receive virtually zero weight. The key difference between these functions is their behavior at the origin. Usually these continuous functions are defined with a fixed bandwidth  $b$ , but can be constructed to behave in an adaptive manner. The bi-square function is useful as it can provide an intermediate weighting between the box-car and the Gaussian functions. To get similar weights from the bi-square and Gaussian functions, the bandwidths  $r$  and  $b$  can be approximately related by  $r \cong (3\sqrt{2}/2)b$ . For all functions, if  $r$  or  $b$  is set suitably large enough, then all of the data can receive a weight of one and the corresponding global model or statistic is found.

Bandwidths can be: (a) user-specified, when there exists some strong prior belief to do so; (b) optimally (or automatically) specified using cross-validation and related approaches, provided there exists an objective function (i.e. the method can be used as a predictor); or (c) user-specified, but guided by (b) where an automated approach is not viewed as a panacea for bandwidth selection (27). In **GWmodel**, automated bandwidths can be found for GW regression, GW regression with a locally compensated ridge term (to address local collinearity problems), generalized GW regression and GW PCA.

### 2.3. Study data-sets

**GWmodel** comes with five example data-sets, these are: (1) *Georgia*, (2) *LondonHP*, (3) *DubVoter*, (4) *EWHP*,

and (5) *USelect*. For this article's presentation of GW models, we use as case studies, the *DubVoter* and *USelect* data-sets.

### 2.3.1. Dublin 2004 voter turnout data

The *DubVoter* data is the main study data-set and is used throughout Sections 3 to 5 and Section 7. This data is composed of nine percentage variables,<sup>1</sup> measuring: (A) voter turnout in the Irish 2004 Dáil elections and (B) eight characteristics of social structure (census data); for  $n = 322$  Electoral Divisions (EDs) of Greater Dublin. Kavanagh et al. (28) modeled this data using GW regression; with voter turnout (*GenEl2004*), the dependent variable (i.e. the percentage of the population in each ED who voted in the election). The eight independent variables measure the percentage of the population in each ED, with respect to:

- one year migrants (i.e. moved to a different address one year ago) (*DiffAdd*);
- local authority renters (*LARent*);
- social class one (high social class) (*SCI*);
- unemployed (*Unempl*);
- without any formal educational (*LowEduc*);
- age group 18–24 (*Age18\_24*);
- age group 25–44 (*Age25\_44*); and
- age group 45–64 (*Age45\_64*).

The independent variables reflect measures of migration, public housing, high social class, unemployment, educational attainment, and three broad adult age groups. Other GW model studies using versions of this data include that of Harris et al. (13) and Gollini et al. (25).

### 2.3.2. US 2004 election data

The *USelect* data is only used in Section 6, for demonstrating a GW DA. It consists of the results of the 2004 US presidential election at the county level, together with a collection of socio-economic (census) variables (29). A variant of this data has been used for the visualization of GW DA outputs in Foley and Demšar (30). In terms of the election results, Bush or Kerry was always the winner within a county; while in some counties, the supporting ratio for a candidate ranged from 45 to 55%, which for our purposes is viewed as an 'unclear-winner' or a 'borderline' result. Thus, for our version of this data-set, we produce a categorical dependent variable with three classes: (i) Bush winner, (ii) Kerry winner, and (iii) Borderline. If we proceed with just two classes: (a) Bush winner and (b) Kerry winner; then an issue arises in that a GW logistic regression may provide a simpler approach to the local modeling of this data, than that found with a GW DA (31) (although observe that both methods can be applied to categorical dependent data with more than two classes (32)). For the *USelect*

data, the five independent variables are taken the same as that used in Foley and Demšar (30), as follows:

- percentage unemployed (*unemployed*)
- percentage of adults over 25 with 4 or more years of college education (*pctcoled*)
- percentage of persons over the age of 65 (*PEROVER65*)
- percentage urban (*pcturban*)
- percentage white (*WHITE*).

## 3. Exploration with GW summary statistics and GW PCA

This first section on GW modeling presents case studies on the use of GW summary statistics and a GW PCA. For demonstration, we investigate the *DubVoter* data.

### 3.1. GW summary statistics

Although simple to calculate and map, GW summary statistics can act as a vital pre-cursor to an application of a subsequent GW model. For example, GW standard deviations will highlight areas of high variability for a given variable; areas where a subsequent application of a GW PCA or a GW regression may warrant close scrutiny. For attributes  $x$  and  $y$  at any location  $i$  where  $w_{ij}$  accords to a kernel function of Section 2, definitions for a GW mean, GW standard deviation and GW correlation coefficient are, respectively

$$\mu(x_i) = \sum_{j=1}^n w_{ij}x_j / \sum_{j=1}^n w_{ij}, \quad (1)$$

$$s(x_i) = \sqrt{\sum_{j=1}^n w_{ij}(x_j - \mu(x_i))^2 / \sum_{j=1}^n w_{ij}}, \quad (2)$$

and

$$\rho(x_i, y_i) = c(x_i, y_i) / (s(x_i) s(y_i)) \quad (3)$$

with the GW covariance

$$c(x_i, y_i) = \sum_{j=1}^n w_{ij}\{(x_j - \mu(x_i))(y_j - \mu(y_i))\} / \sum_{j=1}^n w_{ij}. \quad (4)$$

### 3.2. GW PCA

In a PCA, a set of  $m$  correlated variables are transformed in to a new set of  $m$  uncorrelated variables called components. Components are linear combinations of the

original variables that can allow for a better understanding of sources of variation and trends in the data. Its use as a dimension reduction technique is viable if the first few components account for most of the variation in the original data. In a GW PCA, a series of localized PCAs are computed, where the local component outputs (variances, loadings and scores) are mapped, permitting a local identification of any change in structure of the multivariate data. This local exploration can pinpoint locations where results from a PCA are inappropriate. This in turn, allows for better-informed model decisions for any analysis that may follow, such as a clustering or regression analysis when orthogonal input data are required. GW PCA can assess: (i) how (effective) data dimensionality varies spatially and (ii) how the original variables influence each spatially varying component.

More formally, if an observation location  $j$  has coordinates  $(u_j, v_j)$ , then a GW PCA involves regarding a vector of observed variables  $x_j$  as having a certain dependence on its location  $j$ , where  $\mu(u_j, v_j)$  and  $\Sigma(u_j, v_j)$  are the local mean vector and the local variance-covariance matrix, respectively. The local variance-covariance matrix is

$$\Sigma(u_j, v_j) = \mathbf{X}^T \mathbf{W}(u_j, v_j) \mathbf{X} \quad (5)$$

where  $\mathbf{X}$  is the  $n \times m$  data matrix; and  $\mathbf{W}(u_j, v_j)$  is a diagonal matrix of geographic weights, generated by a kernel function of Section 2. The kernel's bandwidth can be user-specified or found optimally via cross-validation (13). To find the local principal components at location  $(u_j, v_j)$ , the decomposition of the local variance-covariance matrix provides the local eigenvalues and local eigenvectors (or loading vectors) with

$$\mathbf{L}(u_j, v_j) \mathbf{V}(u_j, v_j) \mathbf{L}(u_j, v_j)^T = \Sigma(u_j, v_j) \quad (6)$$

where  $\mathbf{L}(u_j, v_j)$  is a matrix of local eigenvectors and  $\mathbf{V}(u_j, v_j)$  is a diagonal matrix of local eigenvalues. A matrix of local component scores  $\mathbf{T}(u_j, v_j)$  can be found using

$$\mathbf{T}(u_j, v_j) = \mathbf{X} \mathbf{L}(u_j, v_j). \quad (7)$$

If we divide each local eigenvalue by  $\text{tr}(\mathbf{V}(u_j, v_j))$ , then we find a localized proportion of the total variance (PTV) in the original data accounted for by each component. Thus, at each location  $j$  for a GW PCA with  $m$  variables, there are  $m$  components,  $m$  eigenvalues,  $m$  sets of component loadings (with each set  $m \times m$ ), and  $m$  sets of component scores (with each set  $n \times m$ ). We can obtain eigenvalues and their associated eigenvectors at un-observed locations, but as no data exists for these locations, we cannot obtain component scores.

### 3.3. Monte Carlo tests for non-stationarity

For GW summary statistics and GW PCA, Monte Carlo tests are possible that test for non-stationarity (1, 13).

Tests confirm whether or not the GW summary statistic or aspects of the GW PCA are *significantly* different to that found by chance or artifacts of random variation in the data. Here, the sample data are successively randomized and the GW model is applied after each randomization. A basis of a significance test is then possible by comparing the true result with results from a large number of randomized distributions. The randomization hypothesis is that any pattern seen in the data occurs by chance and therefore any permutation of the data is equally likely.

As an example for GW correlation, the test proceeds as follows: (i) calculate the true GW correlation at all locations; (ii) randomly choose a permutation of the data where the coordinates are kept in the same pairs, as are the chosen attribute pairs; (iii) calculate a simulated GW correlation at all locations using the randomized data of (ii); (iv) repeat steps (ii) and (iii), say 99 times; (v) at each location  $i$ , rank the one true GW correlation with the 99 simulated GW correlations; (vi) at each location  $i$ , if the true GW correlation lies in the top or bottom 2.5% tail of the ranked distribution then the true GW correlation can be said to be significantly different (at the 95% level) to such a GW correlation found by chance. The results from this Monte Carlo test are mapped.

For a GW PCA, a similar procedure is followed where the test evaluates whether the local eigenvalues vary significantly across space. Here, the paired coordinates are successively randomized amongst the variable data-set and after each randomization, GW PCA is applied (with an optimally re-estimated bandwidth) and the standard deviation (SD) of a given local eigenvalue is calculated. The true SD of the same local eigenvalue is then included in a ranked distribution of SDs. Its position in this ranked distribution relates to whether there is significant (spatial) variation in the chosen local eigenvalue. The results from this Monte Carlo test are presented via a graph.

### 3.4. Examples: GW correlations

For a demonstration of an analysis using a GW summary statistic, we calculate GW correlations to investigate the local relationships between: (a) voter turnout (*GenEl4004*) and *LARent* and (b) *LARent* and *Unempl*. In the former case, the correlations provide a preliminary assessment of relationship non-stationarity between the dependent and an independent variable of a GW regression of Sections 4 and 5. In the latter case, the correlations provide an assessment of local collinearity between two independent variables of such a GW regression (33). In both cases, we specify a bi-square kernel. Furthermore, as the spatial arrangement of the EDs in Greater Dublin is not a tessellation of equally sized zones, it makes sense to specify an adaptive bandwidth, that we user-specify at  $N=48$ . This entails that the bi-square kernel will change in radius, but will always include the closest 48 EDs for each local correlation. Bandwidths for



GW correlations cannot be found optimally using cross-validation (see also (23)). We also conduct the corresponding Monte Carlo tests for the two GW correlation specifications. Here, we use the **GWmodel** functions *gwss* to find GW summary statistics and *monte-carlo.gwss* to conduct the Monte Carlo tests. The output of our GW correlation analysis is visualized in Figure 1.

From Figure 1(a), the relationship between turnout and *LARent* appears non-stationary, where this relationship is strongest in areas of central and southwest Dublin. Here, turnout tends to be low, while local authority renting tends to be high. The associated Monte Carlo test suggests many instances of unusual relationships, such as those found in the north, that are unusually weak. From Figure 1(b), consistently strong positive correlations between *LARent* and *Unempl* are found in three distinct areas of Greater Dublin; areas where local collinearity effects in the GW regression of Sections 4 and 5 are likely to be a cause for concern (see also (25)).

### 3.5. Examples: PCA to GW PCA

For applications of PCA and GW PCA, we investigate these eight variables: *DiffAdd*, *LARent*, *SC1*, *Unempl*, *LowEduc*, *Age18\_24*, *Age25\_44* and *Age45\_64*. We standardize the data and specify the PCA with the covariance matrix. The same (globally) standardized data is also used in the GW PCA calibration, which is similarly specified with (local) covariance matrices. The effect of this standardization is to make each variable have equal importance in the subsequent analysis (at least for the

PCA case).<sup>2</sup> The PCA results (PTV data and loadings) are found using *scale* and *princomp* R functions, and are presented in Table 2.

From the PTV data, the first two components collectively account for 61.6% of the variation in the data. From the loadings, components one and two mainly represent older (*Age45\_64*) and affluent (*SC1*) residents, respectively. However, these results may not reliably represent local social structure, and an application of GW PCA may be useful. Here, a bandwidth for GW PCA is found using cross-validation, where it is necessary to decide a priori on the number of components,  $k$  to retain, provided  $m \neq k$ . Thus, we choose to find an optimal adaptive bandwidth using a bi-square kernel with the *bw.gwpca* **GWmodel** function, with  $k=3$ . A bandwidth of  $N=131$  is returned and will be used to calibrate the GW PCA fit. Observe that we now specify all  $k=8$  components, but we will focus our investigation on only the first two components. This specification ensures that the PTV data is estimated correctly. The GW PCA fit is conducted using the *gwpca* **GWmodel** function.

The GW PCA outputs are visualized and interpreted, focusing on: (1) how data dimensionality varies spatially and (2) how the original variables influence the components. For the former, the spatial distribution of local PTV for the first two components can be mapped. For the latter, we look at the change in size and sign of the eight local loadings together, for a given component, at each of the 322 EDs. In this respect, we map multivariate glyphs that have spokes around a central hub in which the length of the spoke corresponds to the size of

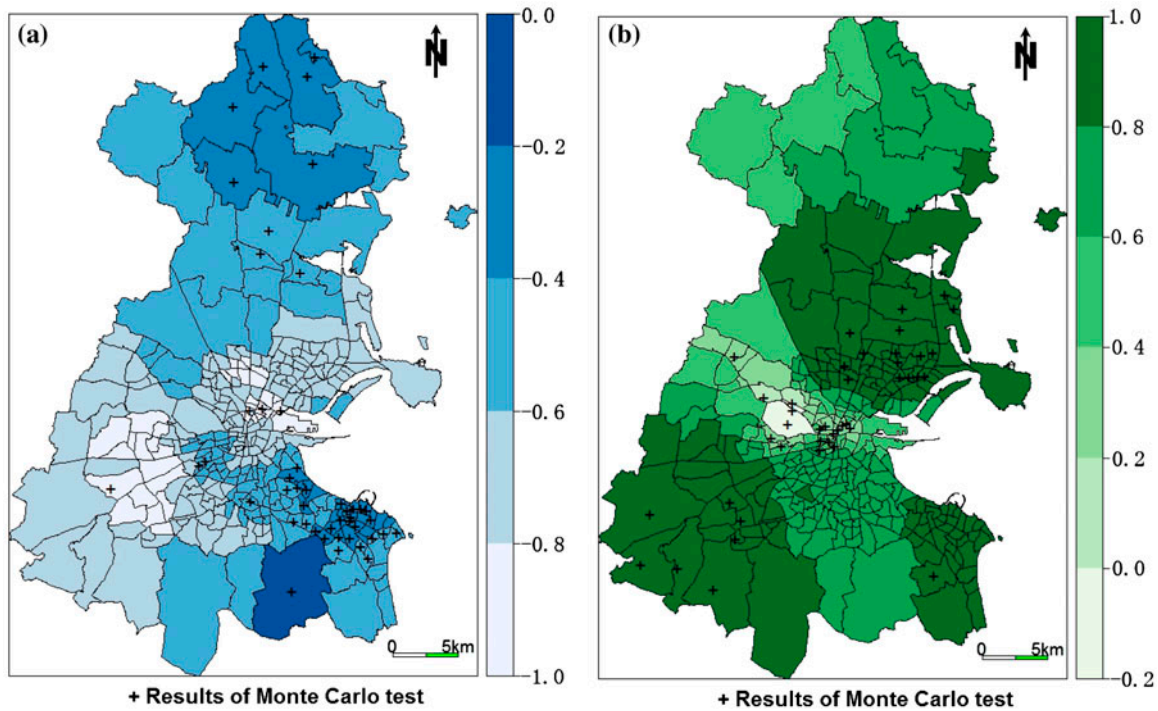


Figure 1. GW correlations and associated Monte Carlo tests for: (a) *GenEl2004* and *LARent*; and (b) *LARent* and *Unempl*. Global correlations are  $-0.68$  and  $0.67$ , respectively.

Table 2. Summary of global PCA.

	PC-1	PC-2	PC-3	PC-4	PC-5	PC-6	PC-7	PC-8
PTV	36.08	25.59	11.92	10.53	6.89	3.68	3.11	2.20
Loadings								
<i>DiffAdd</i>	-0.389	-0.444	-0.004	-0.149	0.123	0.293	0.445	0.575
<i>LARent</i>	-0.441	0.226	0.144	0.172	0.612	0.149	-0.539	0.132
<i>SCI</i>	0.130	-0.576	-0.030	-0.135	0.590	-0.343	0.076	-0.401
<i>Unempl</i>	-0.361	0.462	0.022	0.189	0.197	-0.085	0.670	-0.355
<i>LowEduc</i>	-0.131	0.308	-0.362	-0.861	0.079	-0.062	-0.065	-0.011
<i>AGE18_24</i>	-0.237	-0.080	0.845	-0.359	-0.224	-0.051	-0.045	-0.200
<i>AGE25_44</i>	-0.436	-0.302	-0.317	0.053	-0.291	0.448	-0.177	-0.546
<i>AGE45_64</i>	0.493	0.118	0.179	-0.144	0.289	0.748	0.142	-0.164

the local loading, and its color corresponds to the sign (in this case, blue signifies positive and red signifies negative). The glyphs are scaled relative to the spoke with the largest absolute loading. The variable corresponding to each local loading is always in the same place on the glyph, as follows: *DiffAdd* is at  $0^\circ$  (north); *LARent* is  $45^\circ$  (northeast); *SCI* is  $90^\circ$  (east); *Unempl* is  $135^\circ$  (southeast), *LowEduc* is  $180^\circ$  (south), *Age18\_24* is  $225^\circ$  (southwest), *Age25\_44* is  $270^\circ$  (west) and *Age45\_64* is  $315^\circ$  (northeast).

Figure 2(a) presents the local PTV map, where there is clear spatial variation in the PTV data. Higher percentages are located in the south and in central Dublin, whilst lower percentages are located in the north. The PTV data is also generally higher in the local case, than in the global case (at 61.6%). Figure 2(b) presents a multivariate glyph map for the loadings on the first component, where a spatial preponderance of glyphs of one color or another, or larger spokes on the same variables provide a general indication of the structures being represented at each of the 322 EDs. This map is not intended to be scrutinized in detail, but clearly indicates strong local trends in social structure across Greater Dublin. To provide support for our chosen GW PCA specification, the associated Monte Carlo test evaluates whether the local eigenvalues for the first component vary significantly across space. The results are given in Figure 2(c), where the  $p$ -value for the true SD of the eigenvalues is calculated at 0.02. Thus, an application of GW PCA is considered worthy as the null hypothesis of local eigenvalue stationarity is firmly rejected at the 95% level, for the dominant first component.

#### 4. Mixed model building for GW regression

The first and most commonly applied GW model is GW regression (3). This model enables an exploration of spatially varying data relationships, via the visualization and interpretation of sets of local regression coefficients and associated estimates. For **GWmodel**, a range of GW regression models are included: (i) basic; (ii) robust; (iii) generalized; (iv) mixed; and (v) locally compensated ridge. The use of basic, robust and locally compensated ridge GW regression functions is presented in the

companion paper of Gollini et al. (25). For this paper, we present the use of basic and mixed GW regression functions; together with the further GW regression topics of Section 5.

##### 4.1. Basic and mixed GW regression

The basic form of the GW regression model is

$$y_i = \beta_{i0} + \sum_{k=1}^m \beta_{ik} x_{ik} + \varepsilon_i \quad (8)$$

where  $y_i$  is the dependent variable at location  $i$ ;  $x_{ik}$  is the value of the  $k$ th independent variable at location  $i$ ;  $m$  is the number of independent variables;  $\beta_{i0}$  is the intercept parameter at location  $i$ ;  $\beta_{ik}$  is the local regression coefficient for the  $k$ th independent variable at location  $i$ ; and  $\varepsilon_i$  is the random error at location  $i$ . A key assumption in this basic (and related forms of) GW regression is that the local coefficients vary at the same scale and rate across space (depending on the particular kernel weighting function that is specified). However, some coefficients (and relationships) may be expected to have different degrees of variation over the study region. In particular, some coefficients (and relationships) are viewed as constant (or stationary) in nature, whilst others are not. For these situations, a mixed GW regression can be specified (1, 2). This semi-parametric model treats some coefficients as global (and stationary), whilst the rest are treated as local (and non-stationary), but with the same rate of spatial variation. This model's general form can be written as

$$y_i = \sum_{j=1, k_a} a_j x_{ij}(a) + \sum_{l=1, k_b} b_l(u_i, v_i) x_{il}(b) + \varepsilon_i \quad (9)$$

where  $\{a_1, \dots, a_{k_a}\}$  are the  $k_a$  global coefficients;  $\{b_1(u_i, v_i), \dots, b_{k_b}(u_i, v_i)\}$  are the  $k_b$  local coefficients;  $\{x_{i1}(a), \dots, x_{ik_a}(a)\}$  are the independent variables associated with global coefficients; and  $\{x_{i1}(b), \dots, x_{ik_b}(b)\}$  are the independent variables associated with local coefficients. In a vector-matrix notation, Equation (9) can be rewritten as

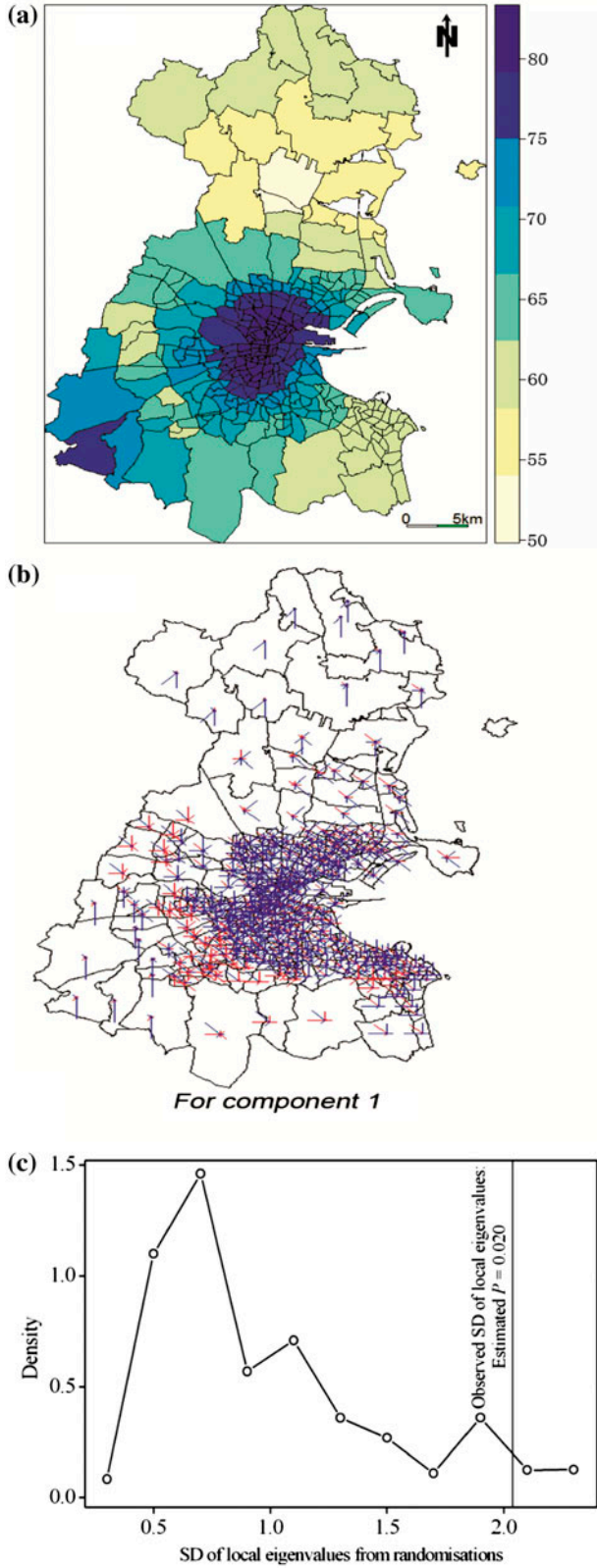


Figure 2. GW PCA results: (a) PTV data for the first two components; (b) multivariate glyphs of the loadings for all eight components; and (c) Monte Carlo test for first component only.

$$y = X_a a + X_b b + \varepsilon \quad (10)$$

where  $y$  is the vector of dependent variables;  $X_a$  is the matrix of globally fixed variables;  $a$  is the vector of  $k_a$  global coefficients;  $X_b$  is the matrix of locally varying variables; and  $b$  is the matrix of local coefficients. To calibrate this model in **GWmodel**, we follow that of Brunsdon et al. (2), where a back-fitting procedure is adopted (34). If we define the hat matrix for the global regression part of the model, as  $S_a$ ; and that for the GW regression part, as  $S_b$ ; then Equation (9) can be rewritten as

$$\hat{y} = \hat{y}_a + \hat{y}_b \quad (11)$$

where the two components,  $\hat{y}_a$  and  $\hat{y}_b$  can be expressed as

$$\hat{y}_a = S_a y \quad (12)$$

$$\hat{y}_b = S_b y \quad (13)$$

and thus the calibration procedure can be briefly described in the following six steps:

- Step 1. Supply an initial value for  $\hat{y}_a$ , say  $\hat{y}_a^{(0)}$ , practically by regressing  $X_a$  on  $y$  using ordinary least squares (OLS)
- Step 2. Set  $i = 1$
- Step 3. Set  $\hat{y}_{(bi)} = S_b [y - \hat{y}_{(ai)}]$
- Step 4. Set  $\hat{y}_{(ai)} = S_a [y - \hat{y}_{(bi)}]$
- Step 5. Set  $i = i + 1$
- Step 6. Return to Step 3 unless  $\hat{y}^{(i)} = \hat{y}_a^{(i)} + \hat{y}_b^{(i)}$  converges to  $\hat{y}^{(i-1)}$

An alternative mixed GW regression fitting procedure is given in Fotheringham et al. (1), that uses a method from Speckman (35). This alternative is not as computationally intensive as that presented, and is under consideration for future mixed GW regressions in **GWmodel**.

#### 4.2. Monte Carlo tests for regression coefficient non-stationarity

For a mixed GW regression, difficulties arise when deciding whether a relationship should be fixed globally or allowed to vary locally. Here, Fotheringham et al. (36) adopt a stepwise procedure, where all possible combinations of global and locally varying relationships are tested, and an optimal mixed model is chosen according to a minimized AIC value. This approach is comprehensive, but computationally expensive, and is utilized in the GW regression 4.0 executable software (37). Alternatively, a Monte Carlo approach can be used to test



for significant (spatial) variation in each regression coefficient (or relationship) from the basic GW regression model, where the null hypothesis is that the relationship between dependent and independent variable is constant (1, 3, 4). The procedure is analogous to that presented for the local eigenvalues of a GW PCA in Section 3.3, where for the basic GW regression the true variability in each local regression coefficient is compared to that found from a series of randomized data-sets. If the true variance of the coefficient does not lie in the top 5% tail of the ranked results, then the null hypothesis cannot be rejected at the 95% level; and the corresponding relationship should be globally fixed when specifying the mixed GW regression model. Observe, that if all relationships are viewed as non-stationary, then the basic GW regression should be preferred. Conversely, if all relationships are viewed as stationary, then the standard global regression should be preferred. Advances on the mixed GW regression model, where the relationships can be allowed to vary at different rates across space can be found in the study of Yang et al. (38).

#### 4.3. Example: mixed GW regression model specification

We now demonstrate modeling building for mixed GW regression using the *DubVoter* data. First we calibrate a basic GW regression. We then conduct the Monte Carlo test on this model's outputs, to gauge for significant variation (or non-stationarity) in each coefficient, including the intercept term. Finally, we fit a mixed GW regression

according to the results of the Monte Carlo test. Our regressions investigate the local/global relationships between the response: *GenEl2004* and these eight predictors: *DiffAdd*, *LARent*, *SC1*, *Unempl*, *LowEduc*, *Age18\_24*, *Age25\_44* and *Age45\_64*. For both GW regressions, we specify a bi-square kernel with an adaptive bandwidth.

The optimal bandwidth for the basic GW regression is found at  $N=109$  in accordance to an automatic AICc approach via the **GWmodel** function *bw.gwr*. This bandwidth is then used to calibrate the basic GW regression via the **GWmodel** function *gwr.basic*. We then conduct the Monte Carlo test (via the **GWmodel** function *monte-carlo.gwr*), where the results are presented in Table 3.

The results suggest (say, at the 95% level) that the *Intercept* term together with the *DiffAdd*, *LARent*, *LowEduc*, *Age25\_44*, and *Age45\_64* variables, should all be fixed as global in the mixed model. Accordingly, the mixed model is calibrated using the **GWmodel** function *gwr.mixed* with the same adaptive bandwidth as that found for the basic model. Observe that the geographically varying coefficients in the mixed model are less variable than the corresponding coefficients from the basic model, although the same bandwidth is used. The estimated coefficients from the functions *gwr.basic* and *gwr.mixed* can be summarized in a manner that imitates the layout of the GW regression 3.0 executable software (39), as shown in Tables 4–6.

As an example visualization, Figure 3(a–b) present the coefficient surfaces corresponding to *Unempl* found from the basic and mixed GW regressions, respectively.

Table 3. Monte Carlo test for the basic GW regression.

Variable	<i>Intercept</i>	<i>DiffAdd</i>	<i>LARent</i>	<i>SC1</i>	<i>Unempl</i>	<i>LowEduc</i>	<i>Age18_24</i>	<i>Age25_44</i>	<i>Age45_64</i>
<i>p</i> -value	0.35	0.17	0.28	0.02	0.00	0.19	0.04	0.29	0.19

Table 4. Summary of GWR coefficient estimates.

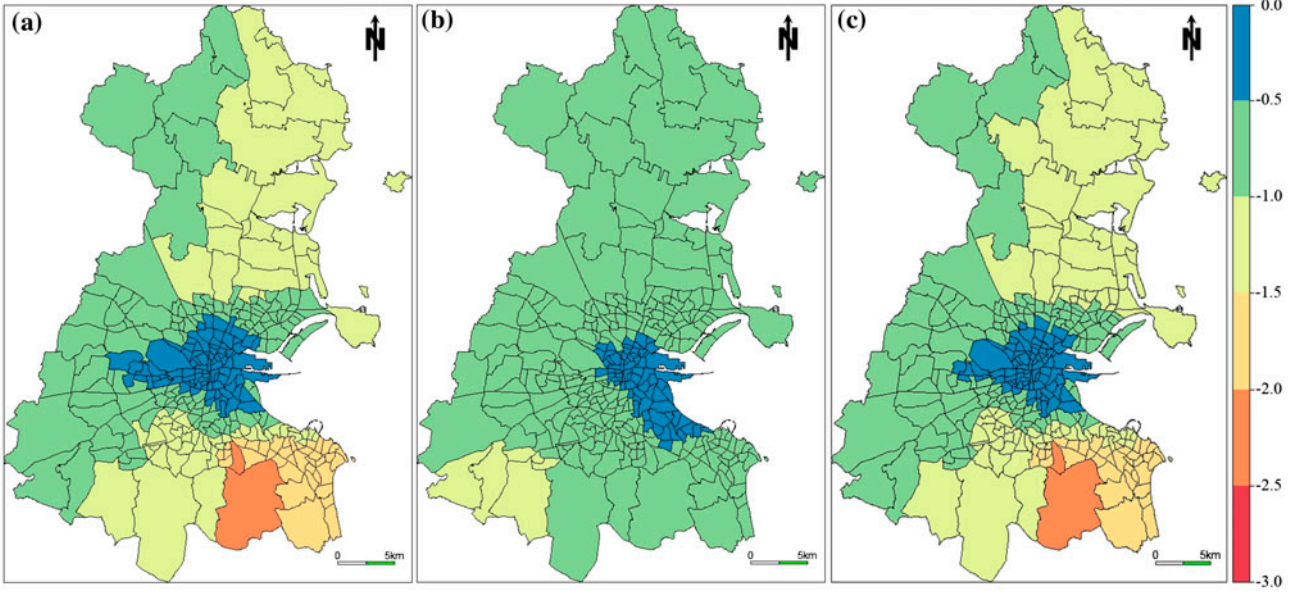
Variable	Min.	1st Qu.	Median	3rd Qu.	Max.
<i>Intercept</i>	53.2300000	73.3200000	81.6600000	95.0700000	116.8000
<i>DiffAdd</i>	−0.7281000	−0.3338000	−0.1584000	0.1586000	0.5465
<i>LARent</i>	−0.1949000	−0.1206000	−0.0844400	−0.0369200	0.0940
<i>SC1</i>	−0.1578000	0.0352800	0.3088000	0.4201000	0.8796
<i>Unempl</i>	−2.3180000	−1.1440000	−0.7649000	−0.4753000	−0.0925
<i>LowEduc</i>	−7.6750000	−0.7369000	0.5332000	1.8100000	3.4140
<i>Age18_24</i>	0.3970000	−0.2529000	−0.1457000	0.0007642	0.3669
<i>Age25_44</i>	−1.0950000	−0.7209000	−0.4536000	−0.3048000	0.2184
<i>Age45_64</i>	−0.9236000	−0.4098000	−0.1102000	0.0467900	0.4931

Table 5. Summary of mixed GWR coefficient estimates.

Estimated global variables	<i>Intercept</i>	<i>DiffAdd</i>	<i>LARent</i>	<i>LowEduc</i>	<i>Age25_44</i>	<i>Age45_64</i>
Estimated global coefficients	86.31399	−0.15299	−0.11481	0.12894	−0.53151	−0.2579

Table 6. Estimated GWR variables.

Variable	Min.	1st Qu.	Median	3rd Qu.	Max.
<i>SCI</i>	0.01948	0.10360	0.19820	0.42970	0.7112
<i>Unempl</i>	-1.03400	-0.77250	-0.65630	-0.52440	-0.0668
<i>Age18_24</i>	-0.40890	-0.20600	-0.12650	-0.06580	0.1139

Figure 3. (a) Basic; (b) mixed; and (c) heteroskedastic GW regression coefficients estimates for *Unempl*.

The spatial variation in this coefficient is clearer greater when using the basic GW regression. Differences in the coefficient surfaces primarily occur in the northwest and southwest areas of Dublin.

## 5. Further GW regression topics

In this section, we present the use of **GWmodel** GW regression functions to conduct: (i) multiple hypothesis tests, (ii) collinearity diagnostics, and (iii) heteroskedastic fits.

### 5.1. Multiple hypothesis tests with GW regression

For GW regression, pseudo *t*-values can be used to test, in a purely informal sense, for evidence of local coefficient estimates that are *significantly* different from zero (e.g. 22). For each coefficient estimate,  $\beta_k(u_i, v_i)$  at location *i*, the pseudo *t*-value can be calculated using:

$$T_{k,i} = \frac{\hat{\beta}_k(u_i, v_i)}{SE(\hat{\beta}_k(u_i, v_i))} \quad (14)$$

where  $SE(\hat{\beta}_k(u_i, v_i))$  is the standard error of  $\hat{\beta}_k(u_i, v_i)$ . For details on the standard error calculations, see (1). However, as GW regression yields a separate model at each location *i*, (where each model is calibrated with the same

observations but with different weighting schemes), we need to conduct a large number of simultaneous *t*-tests. This operation is likely to result in high order multiple inference problems, where the likelihood of increased type I errors needs to be controlled. In this respect, various standard approaches are available that adjust each test's *p*-value; these include: (a) Benjamini-Hochberg (40), (b) Benjamini-Yekutieli (41), and (c) Bonferroni (42) approaches. All three approaches can be used with GW regression, but a fourth, the Fotheringham-Byrne (43) approach is specifically designed for this purpose. This Bonferroni-style adjustment can be briefly described as follows. Let the probability of rejecting one or more true null hypothesis (i.e. the family-wise error rate, FWER) be denoted by  $\xi_m$  (with *m* the number of tests). Then the FWER for testing hypotheses about GW regression coefficients can be controlled at  $\xi_m$  or less, by selecting

$$\alpha = \xi_m / \left(1 + p_e - \frac{p_e}{np}\right) \quad (15)$$

where  $\alpha$  is the probability of a type I error in the *i*th test;  $p_e$  is the effective number of parameters in the GW regression;  $np$  is the number of parameters in each individual local regression (i.e. the same as that found in the global regression); and *n* is the sample size.

To demonstrate this topic, the results from the basic GW regression of Section 4 are investigated. Here all four adjustment approaches are provided in the **GWmodel** function *gwr.t.adjust*, which simply needs to be specified with the results from the GW regression run. Accordingly, we can map the competing test results, where as an example, Figure 4(a) displays the surface of the original (un-adjusted)  $p$ -values for the *Unempl*

coefficient; and Figures 4(b–c) and 5(a–b) present the surfaces for the corresponding adjusted  $p$ -values (all significant results are colored red). Observe that the standard (Benjamini-Hochberg, Benjamini-Yekutieli, Bonferroni) approaches adjust the  $p$ -values to zero or one (i.e. in-significant and significant), whilst the Fotheringham-Byrne approach provides outputs from zero to one. From the five maps, it can be observed that: (1) the

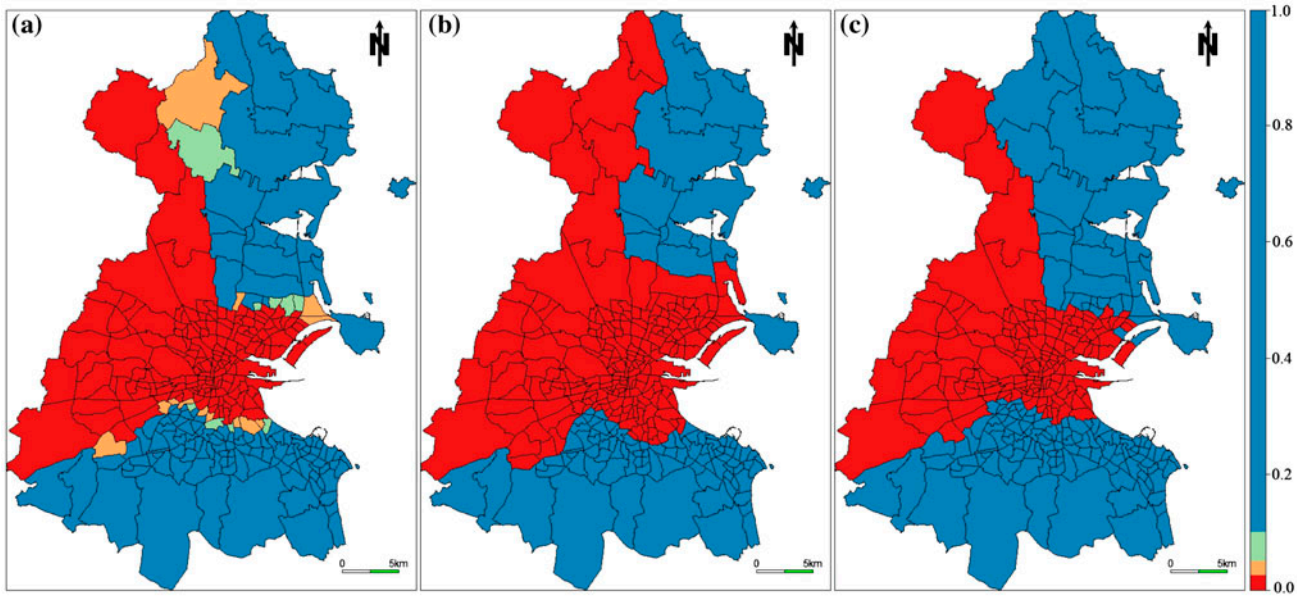


Figure 4. The  $p$ -values associated with *Unempl* from the basic GW regression of Section 4: (a) un-adjusted; (b) adjusted by Benjamini-Hochberg; and (c) adjusted by Benjamini-Yekutieli.

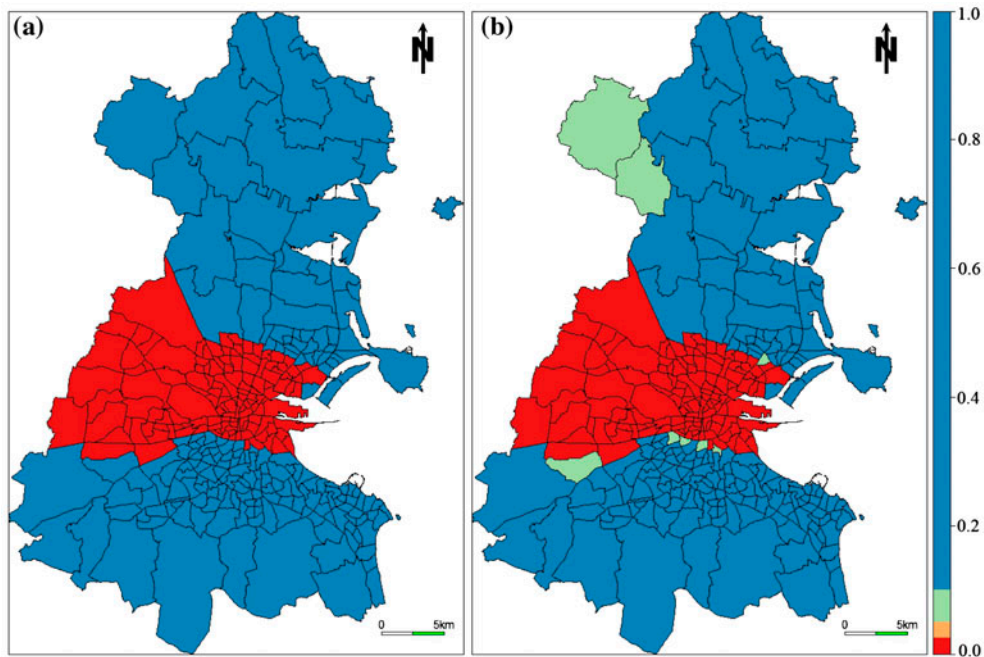


Figure 5. The  $p$ -values associated with *Unempl* from the basic GW regression of Section 4: (a) adjusted by Bonferroni; and (b) adjusted by Fotheringham-Byrne.



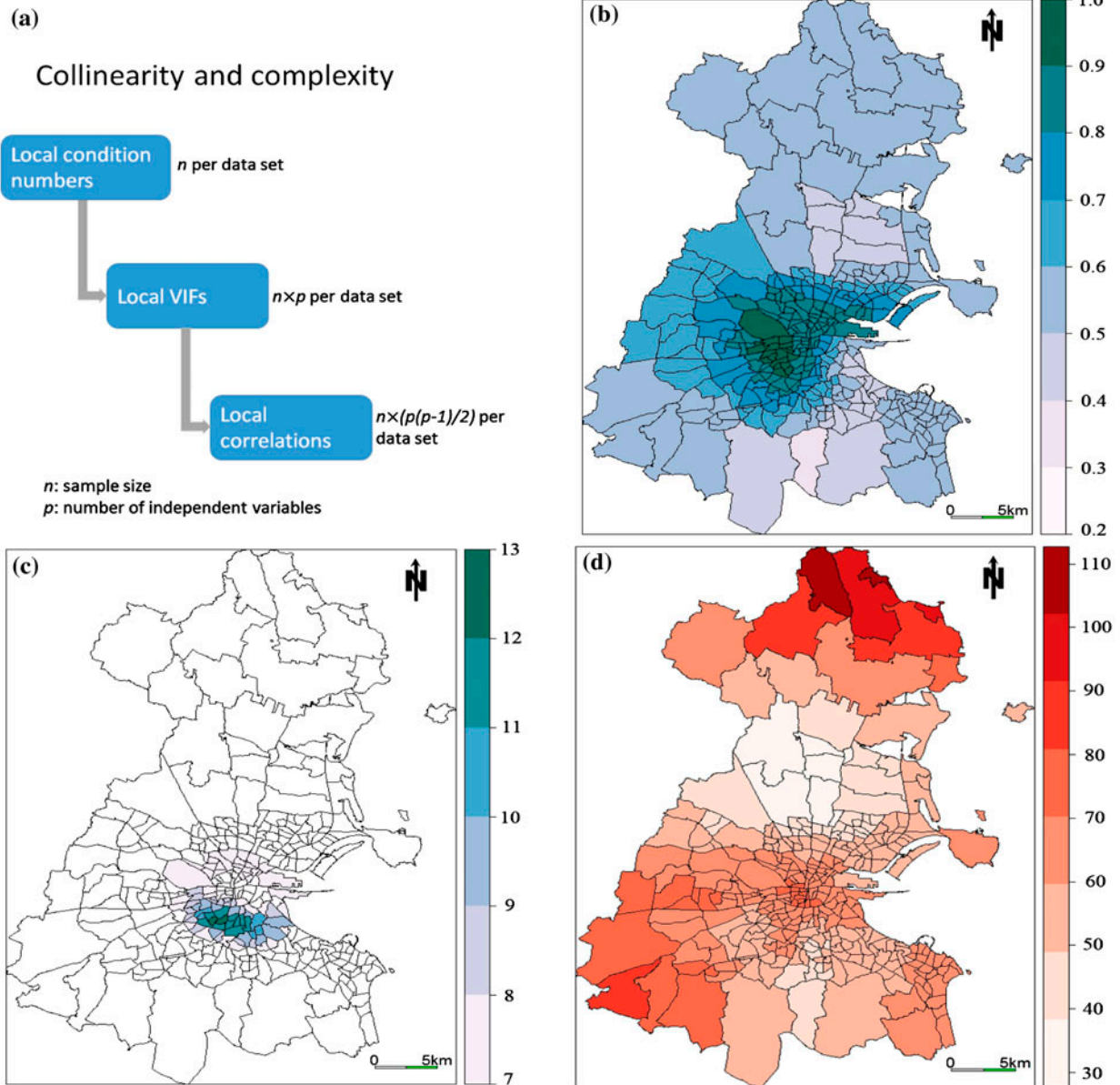


Figure 6. (a) Levels of complexity for different localized collinearity diagnostics; (b) GW correlations; (c) local VIFs; and (d) local condition numbers.

un-adjusted  $p$ -values are similar to that found from Benjamini-Hochberg and Benjamini-Yekutieli adjustments; and (2) the Bonferroni and Fotheringham-Byrne adjustments provide similar results.

### 5.2. Local collinearity diagnostics for a basic GW regression

The problem of collinearity amongst the predictor variables of a regression model has long been acknowledged and can lead to a loss of precision and power in the coefficient estimates (44). This issue is heightened in GW regression since: (A) its effects can be more pronounced with the smaller samples that are used to calibrate each local regression; and (B) if the data is

spatially heterogeneous in terms of its correlation structure, some localities may exhibit collinearity when others do not. In both cases, (local) collinearity may cause serious problems in GW regression, when none are found in the corresponding global regression. Diagnostics to investigate local collinearity in a GW regression model, include finding: (i) GW correlations amongst pairs of independent variables; (ii) local variance inflation factors (VIFs) for each independent variable; (iii) local variance decomposition proportions (VDPs); and (iv) local (design matrix) condition numbers (CN). Accordingly, the following *rules of thumb* can be taken to indicate likely local collinearity problems in the GW regression model: (a) absolute GW correlation values greater than 0.8 for a given independent variable pair; (b) VIFs greater than 10



for a given independent variable; (c) VDPs greater than 0.5; and (d) CN values greater than 30. Observe that all diagnostics are found at the same spatial scale as each local regression of the GW regression model and can thus be mapped. These diagnostics are considered an integral part of an analytical toolkit that should always be employed in any GW regression analysis. Figure 6(a) depicts the levels of complexity associated with such investigations (excluding the use of VDPs). Details on the use and merit of these diagnostics, possible model solutions, and critical discussions on this issue with GW regression, can be found in (6, 25, 33, 45–49). For this study, our objective is to simply demonstrate some of the collinearity diagnostics used. The application of a (possible) model solution with **GWmodel**, for example, via a locally compensated ridge GW regression, is demonstrated in Gollini et al. (25) using the **GWmodel** function *gwr.lcr*.

For a given GW regression specification, GW correlations, local VIFs, local VDPs and the local CNs can be found using the **GWmodel** function *gwr.collin.diagno*. The same local CNs can also be found using the function *gwr.lcr*. Example maps presenting GW correlations, local VIFs and local CNs are given in Figure 6(b–d), reflecting diagnostics for same the basic GW regression of Section 4. Scales of each map are chosen to highlight our critical thresholds. Clearly, significant collinearity is present in our study GW regression model, where Diff-Add appears to be a major cause with respect to its relationship to Age25\_44 in central areas of Dublin. As the local CNs are large everywhere, the simple removal of one variable from the analysis may go some way in alleviating this problem; before proceeding to a more locally focused analysis with some locally compensated model. This of course brings into question the validity of the results of Kavanagh et al. (28), where basic GW regression was applied to this data.

### 5.3. Heteroskedastic GW regression

Basic GW regression assumes that the error term is normally distributed with zero mean and constant (stationary) variance over the study region ( $\varepsilon_i \sim N(0, \sigma^2)$ ). An extension of GW regression is possible, which allows a non-stationary error variance ( $\varepsilon_i \sim N(0, \sigma^2(u_i, v_i))$ ). This (possibly more realistic) model was first proposed in Fotheringham et al. (1) and has been further extended to a predictive form in Harris et al. (18). Details of such heteroskedastic GW regressions can be found in (1, 18), where an iterative modeling technique is used, requiring the model to converge to some pre-specified level of tolerance. An alternative (parametric) heteroskedastic GW regression can be found the work of Paez et al. (50–55).

For **GWmodel**, the function *gwr.hetero* allows the non-parametric version to be specified. Currently, the model is only given in a very rudimentary form, where the kernel function used to control the coefficient estimates is also used to control the local error variances;

which are themselves approximated by the squared residuals ( $e^2$ ). Outputs from *gwr.hetero* are the regression coefficients only, which can be compared to the corresponding coefficients found using a basic GW regression (*gwr.basic*). As an example, Figure 3(c) displays the coefficient surface for *Unempl* from the heteroskedastic model. Clearly, there is little difference from the coefficient surface of the basic model (Figure 3(a)). Thus, modeling with a stationary error variance appears reasonable. The use of a non-stationary error variance can be useful, however, for improved measures of uncertainty (18) and for outlier detection (22).

## 6. GW discriminant analysis

Discriminant analysis (DA) allows the modeling and prediction of a categorical dependent variable explained by a set of independent variables. As with GW regression, the relationships between the dependent and independent variables may vary across space. In such a case, a GW discriminant analysis (GW DA) (15) provides a useful investigative tool, where the discrimination rule is localized. DA (and in turn, GW DA) provides an alternative to logistic regression (and in turn, GW logistic regression); a useful comparison of which can be found in the simulation study of Pohar et al. (32), where guidelines to choosing one method in preference to the other are presented.

### 6.1. GW discriminant analysis

The theoretical context for DA is briefly described. Suppose a population, of which each object belongs to  $k$  possible categories; and a training set  $X$ , where each row vector  $\mathbf{x}_i$  indicates an observation belonging to category  $l$  ( $l \in \{1, \dots, k\}$ ); then for an observation vector  $\mathbf{x}$ , the discrimination rule is to assign  $\mathbf{x}$  to the  $l$ th category with the maximum probability that  $\mathbf{x}$  belongs to this category, say

$$P_l(\mathbf{x}) = \max_{j \in \{1, \dots, k\}} \{P_j(\mathbf{x}) = p_j f_j(\mathbf{x})\}. \quad (16)$$

Here,  $p_j f_j(\mathbf{x})$  is proportional to the posterior probability of an observation arising from population  $j$  once the value of  $\mathbf{x}$  is known. Now a linear DA (LDA) assumes that the distribution for each category is multivariate normal with a discrimination rule of

Table 7. DA and GW DA confusion matrices.

	Borderline	Bush	Kerry	Total
(DA)				
Borderline	1	1	4	6
Bush	543	2099	166	2808
Kerry	92	49	156	297
Total	636	2149	326	3111
(GW DA)				
Borderline	29	15	12	56
Bush	522	2103	145	2770
Kerry	85	31	169	285
Total	636	2149	326	3111

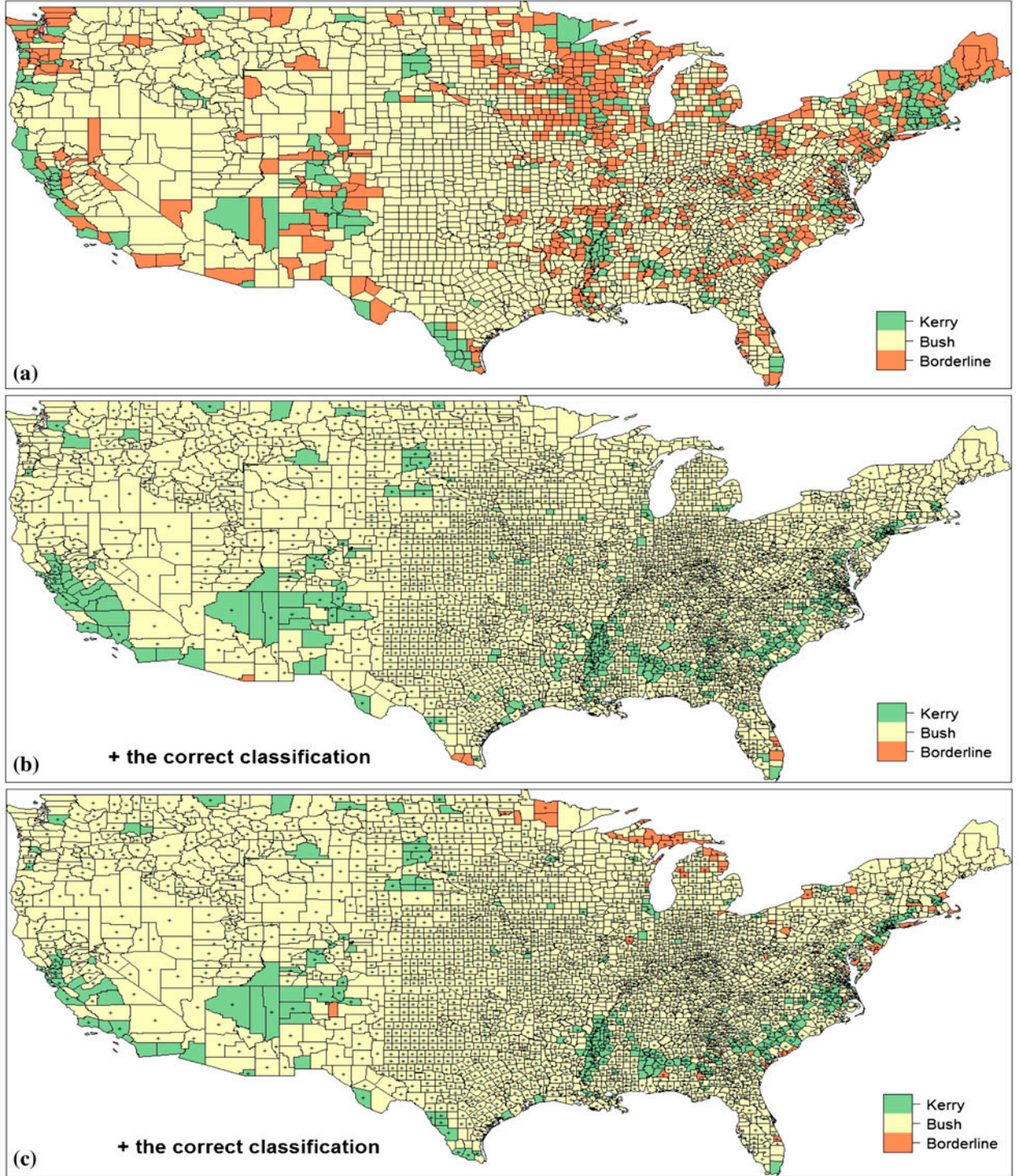


Figure 7. (a) Results of 2004 US presidential election. Classification results using: (b) DA; and (c) GW DA.

$$P_l(\mathbf{x}) = \max_{j \in \{1, \dots, k\}} \left\{ P_j(\mathbf{x}) = p_j \frac{1}{(2\pi|\Sigma|)^{q/2}} e^{-\frac{1}{2}(\mathbf{x} - \mu_j)^T \Sigma^{-1}(\mathbf{x} - \mu_j)} \right\} \quad (17)$$

where  $\Sigma$  is the covariance matrix,  $q$  is the number of independent variables in  $\mathbf{x}$ , and  $\mu_j$  is the mean for

population  $j$ . This function can be simplified by taking logs and changing signs

$$LP_l(\mathbf{x}) = \min_{j \in \{1, \dots, k\}} \left\{ LP_j(\mathbf{x}) \right. \\ \left. = \frac{1}{2}(\mathbf{x} - \mu_j)^T \Sigma^{-1}(\mathbf{x} - \mu_j) + \frac{q}{2} \log(|\Sigma|) - \log(p_j) \right\}. \quad (18)$$



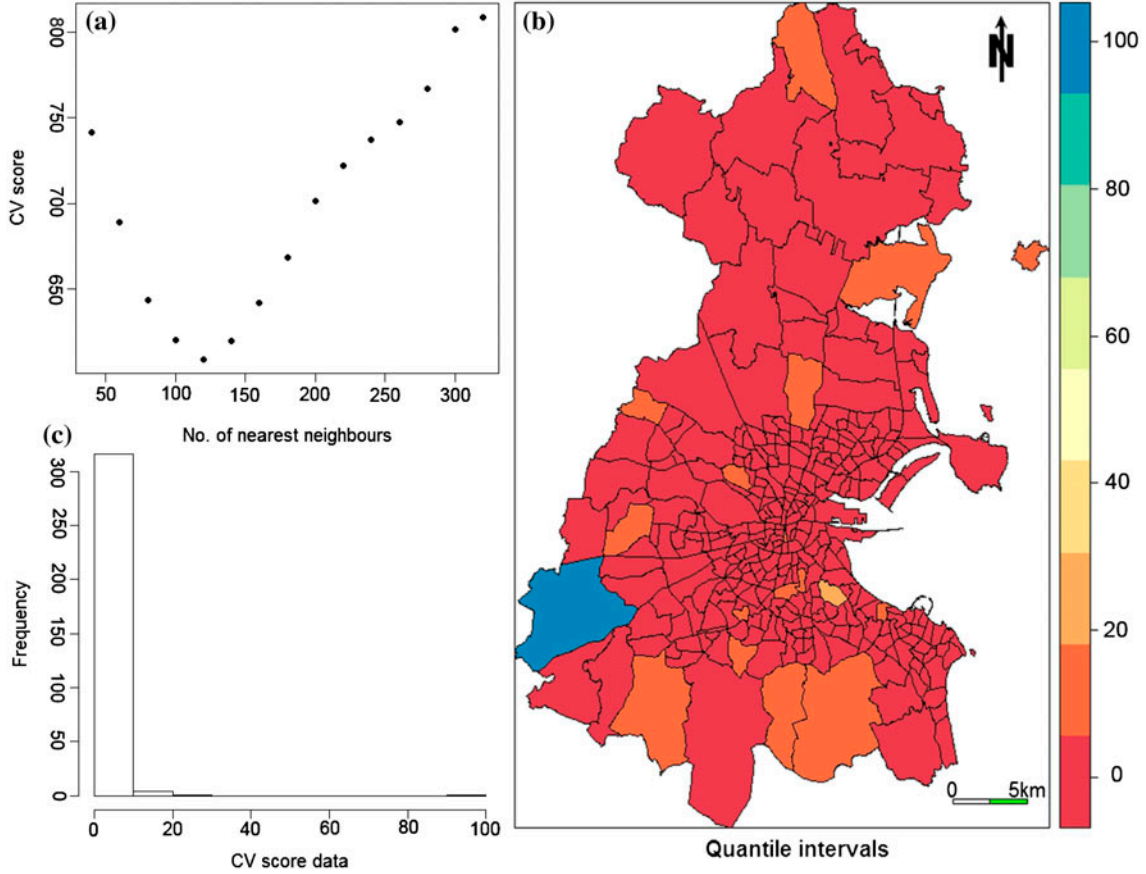


Figure 8. GW PCA calibration: (a) adaptive CV bandwidth function; (b) map; and (c) histogram of the CV score data for the optimal bandwidth of  $N=131$ .

LDA assumes that  $\Sigma$  is identical for each category, whilst an alternative, quadratic DA (QDA) assumes that  $\Sigma$  is different for each category  $j$ , where its discrimination rule replaces  $\Sigma$  in Equation (18) with  $\Sigma_j$ .

GW DA is a direct local adaption of DA, with the chosen discrimination rule varying across space. Here the stationary mean and covariance estimates of DA are replaced with respective GW mean estimates (Equation (1)) and GW covariance estimates (Equation (4)) in the discrimination rules for both LDA and QDA. Thus, a local LDA or QDA can be found at any location  $(u, v)$  using GW DA. Bandwidth selection follows a cross-validation approach, where an optimum bandwidth is identified by minimizing this score:

$$CV_{GWDA}(b) = \sum_i P_{\neq i}(b) \quad (19)$$

where  $P_{\neq i}$  is the proportion of incorrect assignments when the observation  $i$  is removed from the sample data.

### 6.2. Example: GW discriminant analysis

To demonstrate a GW DA, we use the *USelect* data described in Section 2. Here, we calibrate a GW DA using the function *gwda* in **GWmodel**, together with a

standard DA (LDA) using the function *lda* from **MASS** (51). The GW DA is conducted with an adaptive bandwidth (bi-square kernel) with its optimum found using the **GWmodel** function *bw.gwda*. The resultant confusion matrices are presented in Table 7. Here, the DA classification accuracy is 72.5%, whilst GWDA provides a slightly improved classification accuracy of 74.0%. An interesting feature is that the global model predicts only one county in the ‘Borderline’ category. Results of the actual presidential election are mapped at the county level in Figure 7(a), where Bush was a clear winner in most counties, while the election was more competitive in areas like Wisconsin and Maine. The classification results using DA and GW DA are mapped in Figure 7(b–c), where the spatial pattern in the GW DA classifications appears marginally closer to the true results than that found with the DA.

### 7. Enhanced kernel bandwidth selection

In **GWmodel**, a number of functions are provided to aid bandwidth selection. These include: *bw.ggwr*, *bw.gwda*, *bw.gwpcr*, *bw.gwr*, and *bw.gwr.lcr* for automatic bandwidth selection when calibrating a generalized GW regression, a GW DA, a GW PCA, (basic) GW regression and GW regression with a local compensated ridge

term, respectively. However, it is not always recommended to simply plug the resultant (optimal) bandwidth into the given GW model, without first checking the behavior of the full bandwidth function. Here we demonstrate how to: (i) investigate for multiple minima in this function and (ii) assess if (outlying) observations adversely affect the behavior of this function. In this respect, **GWmodel** provides the following functions for constructing a cross-validation (CV) bandwidth function (i.e. bandwidth vs. the CV score): `ggwr.cv`, `gwpca.cv`, `gwr.cv` and `gwr.lcr.cv` (with `gwda.cv` still to be coded). Similarly, **GWmodel** provides the following functions for finding the CV score data at each observation location, for a given bandwidth: `ggwr.cv.contrib`, `gwpca.cv.contrib`, `gwr.cv.contrib`, and `gwr.lcr.cv.contrib` (with `gwda.cv.contrib` still to be coded).

Observe that the CV score data is summed to provide the CV score for a given bandwidth. As an example of using these functions, we further investigate the GW PCA conducted in Section 3. Here we use the `bw.gwpca`, `gwpca.cv` and `gwpca.cv.contrib` functions. The CV bandwidth function, and the histogram and map of the CV score data for a bandwidth of  $N=131$ , are presented in Figure 8(a–c). Clearly, the CV bandwidth function is well-behaved, reaching a clear minimum at  $N=131$ ; and thus provides re-assurance in this bandwidth's use. At this specific bandwidth, the CV score data is heavily positively skewed, with one extreme value at 98.4 that corresponds to an ED in the southwest of Greater Dublin. The census data at this ED may warrant additional scrutiny.

## 8. Concluding remarks

This study, together with its companion study (25), demonstrates the application of a wide range of techniques for investigating spatial heterogeneity, using functions provided by the **GWmodel** R package. Topics include that of (i) GW summary statistics, (ii) GW principal component analysis, (iii) GW regression, and (iv) GW Discriminant Analysis. The GW modeling paradigm provides a simple, yet powerful analytical toolkit for exploring change in a statistical model's parameters and outputs across space; a paradigm that continues to evolve (e.g. (16, 21, 23, 26, 49, 52–55)). Functions for these more recent advances in GW modeling will be incorporated into **GWmodel** in due course.

## Acknowledgments

Research presented in this paper was funded by a Strategic Research Cluster grant (07/SRC/I1168) by Science Foundation Ireland under the National Development Plan. The authors gratefully acknowledge this support.

## Supplementary material

The Supplementary material for this paper is available at <http://dx.doi.org/10.1080/10095020.2014.917453>.

## Notes

1. Observe that none of these variables constitute a closed system and as such, do not need to be treated as compositional data.
2. The use of un-standardized data, or the use of locally standardized data with GW PCA is a subject of current research.

## Notes on contributors

Binbin Lu is a lecturer at School of Remote Sensing and Information Engineering, Wuhan University. His research interests are spatial statistics, spatial analysis, GW models, open-source GIS coding in R and python.

Paul Harris is a senior research scientist at Rothamsted Research, North Wyke, UK. His research interests are geostatistics, GW models, and robust spatially varying parameter models.

Martin Charlton is a senior research associate in National Centre for Geocomputation, National University of Ireland Maynooth. His research areas include spatial statistics, GIScience, geodemographics, and GW models.

Chris Brunsdon is the director of the National Centre for Geocomputation, National University of Ireland Maynooth. Chris Brunsdon and Martin Charlton together with Stewart Fotheringham are the originators of GW Regression. His research interests are geocomputation, spatial statistics, exploratory spatial data analysis, and visualization.

## References

- (1) Fotheringham, A.S.; Brunsdon, C.; Charlton, M. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*; Wiley: Chichester, 2002.
- (2) Brunsdon, C.; Fotheringham, A.S.; Charlton, M. Some Notes on Parametric Significance Tests for Geographically Weighted Regression. *J. Regional Sci.* **1999**, *39*, 497–524.
- (3) Brunsdon, C.; Fotheringham, A.S.; Charlton, M. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geogr. Anal.* **1996**, *28*, 281–298.
- (4) Brunsdon, C.; Fotheringham, S.; Charlton, M. Geographically Weighted Regression. *J. Royal Statist. Soc. Series D* **1998**, *47*, 431–443.
- (5) Leung, Y.; Mei, C.L.; Zhang, W.X. Statistical Tests for Spatial Nonstationarity Based on the Geographically Weighted Regression Model. *Environ. Plan. A* **2000**, *32*, 9–32.
- (6) Wheeler, D.C. Diagnostic Tools and a Remedial Method for Collinearity in Geographically Weighted Regression. *Environ. Plan. A* **2007**, *39*, 2464–2481.
- (7) Waller, L.A.; Gotway, C.A. *Applied Spatial Statistics for Public Health Data*; Wiley: New York, 2004.
- (8) Gelfand, A.E.; Kim, H.-J.; Sirmans, C.F.; Banerjee, S. Spatial Modeling with Spatially Varying Coefficient Processes. *J. Am. Stat. Assoc.* **2003**, *98*, 387–396.
- (9) Assunção, R.M. Space Varying Coefficient Models for Small Area Data. *Environmetrics* **2003**, *14*, 453–473.
- (10) Brunsdon, C.; Fotheringham, A.; Charlton, M. Geographically Weighted Local Statistics Applied to Binary Data. *Lecture Notes in Computer Science* **2002**, *2478*, 38–50.
- (11) Brunsdon, C.; Fotheringham, A.S.; Charlton, M. Geographically Weighted Summary statistics – A Framework for Localised Exploratory Data Analysis. *Comput. Environ. Urban Syst.* **2002**, *26*, 501–524.



- (12) Dykes, J.; Brunson, C. Geographically Weighted Visualization: Interactive Graphics for Scale-Varying Exploratory Analysis. *IEEE Trans. Visual. Comput. Graph.* **2007**, *13*, 1161–1168.
- (13) Harris, P.; Brunson, C.; Charlton, M. Geographically Weighted Principal Components Analysis. *Int. J. Geogr. Inform. Sci.* **2011**, *25*, 1717–1736.
- (14) Nakaya, T.; Fotheringham, A.S.; Brunson, C.; Charlton, M. Geographically Weighted Poisson Regression for Disease Association Mapping. *Stat. Med.* **2005**, *24*, 2695–2717.
- (15) Brunson, C.; Fotheringham, S.; Charlton, M. Geographically Weighted Discriminant Analysis. *Geogr. Anal.* **2007**, *39*, 376–396.
- (16) Harris, P.; Charlton, M.; Fotheringham, A.S. Moving Window Kriging with Geographically Weighted Variograms. *Stochast. Environ. Res. Risk Assess.* **2010**, *24*, 1193–1209.
- (17) Harris, P.; Fotheringham, A.S.; Crespo, R.; Charlton, M. The Use of Geographically Weighted Regression for Spatial Prediction: An Evaluation of Models Using Simulated Data Sets. *Math. Geosci.* **2010**, *42*, 657–680.
- (18) Harris, P.; Brunson, C.; Fotheringham, A.S. Links, Comparisons and Extensions of the Geographically Weighted Regression Model When Used as a Spatial Predictor. *Stochast. Environ. Res. Risk Assess.* **2011**, *25*, 123–138.
- (19) Harris, P.; Juggins, S. Estimating Freshwater Acidification Critical Load Exceedance Data for Great Britain Using Space-Varying Relationship Models. *Math. Geosci.* **2011**, *43*, 265–292.
- (20) Machuca-Mory, D.; Deutsch, C. Non-Stationary Geostatistical Modeling Based on Distance Weighted Statistics and Distributions. *Math. Geosci.* **2012**, *45*, 31–48.
- (21) Harris, P.; Brunson, C.; Charlton, M.; Juggins, S.; Clarke, A. Multivariate Spatial Outlier Detection Using Robust Geographically Weighted Methods. *Math. Geosci.* **2014**, *46*, (1), 1–31.
- (22) Harris, P.; Fotheringham, A.S.; Juggins, S. Robust Geographically Weighted Regression: A Technique for Quantifying Spatial Relationships between Freshwater Acidification Critical Loads and Catchment Attributes. *Ann. Assoc. Am. Geogr.* **2010**, *100*, 286–306.
- (23) Harris, P.; Clarke, A.; Juggins, S.; Brunson, C.; Charlton, M. Geographically Weighted Methods and Their Use in Network Re-designs for Environmental Monitoring. *Stoch. Environ. Res. Risk Assess.* [Online] **2014**, 1–19. doi:10.1007/s00477-014-0851-1.
- (24) Harris, P.; Charlton, M.; Brunson, C. Geographically Weighted (GW) Models: Advances in Modelling Spatial Heterogeneity. *GeoENV 2012*, Valencia, Spain, 2012.
- (25) Gollini, I.; Lu, B.; Charlton, M.; Brunson, C.; Harris, P. *GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models*. arXiv: 1306.0413 [stat.AP], 2013.
- (26) Lu, B.; Charlton, M.; Harris, P.; Fotheringham, A.S. Geographically Weighted Regression with a Non-euclidean Distance Metric: A Case Study Using Hedonic House Price Data. *Int. J. Geogr. Inform. Sci.* **2014**, *28*, 660–681.
- (27) Silverman, B. *Density Estimation for Statistics and Data Analysis*; Chapman & Hall: London, 1986.
- (28) Kavanagh, A.; Sinnott, R.; Fotheringham, A.S.; Charlton, M. In *A Geographically Weighted Regression Analysis of General Election Turnout in the Republic of Ireland*, Political Studies Association of Ireland Conference, Cork; University College Cork: Cork, **2006**, 1–13.
- (29) Robinson, A.C. In *Geovisualization of the 2004 Presidential Election*. In National Institutes of Health, P.S.U., Eds.; Penn State, 2013. <http://www.personal.psu.edu/users/a/c/acr181/election.html>
- (30) Foley, P.; Demšar, U. Using Geovisual Analytics to Compare the Performance of Geographically Weighted Discriminant Analysis versus Its Global Counterpart, Linear Discriminant Analysis. *Int. J. Geogr. Inform. Sci.* **2012**, *27*, 633–661.
- (31) Maddala, G.S. *Limited-dependent and Qualitative Variables in Econometrics*; Cambridge University Press: Cambridge, UK, 1983.
- (32) Pohar, M.; Blas, M.; Turk, S. Comparison of Logistic Regression and Linear Discriminant Analysis: A Simulation Study. *Metodoloski* **2002**, *1*, 143–161.
- (33) Wheeler, D.; Tiefelsdorf, M. Multicollinearity and Correlation among Local Regression Coefficients in Geographically Weighted Regression. *J. Geogr. Syst.* **2005**, *7*, 161–187.
- (34) Hastie, T.; Tibshirani, R. Generalized Additive Models. *Stat. Sci.* **1986**, *1*, 297–310.
- (35) Speckman, P. Kernel Smoothing in Partial Linear Models. *J. Royal Stat. Soc. Series B* **1988**, *50*, 413–436.
- (36) Fotheringham, A.S.; Charlton, M.; Brunson, C. Nakaya, T. In *Model Selection Issues in Geographically Weighted Regression, Proceedings of the 8th International Conference on GeoComputation, University of Michigan, USA, 2005*; Xie, Y. and Brown, D.G., Eds.; University of Michigan: Ann Arbor, MI, 2005.
- (37) Nakaya, T.; Charlton, M.; Fotheringham, A.S.; Brunson, C. *How to Use SGWRWIN (GWR4.0)*; National Centre for Geocomputation, National University of Ireland Maynooth: Maynooth, 2009.
- (38) Yang, W.; Fotheringham, A.S.; Harris, P. An Extension of Geographically Weighted Regression with Flexible Bandwidths. *Proceedings GISRUK 20th Annual Conference*, Lancaster, 2012.
- (39) Charlton, M.; Fotheringham, A.S.; Brunson, C. *GWR 3: Software for Geographically Weighted Regression*; National Centre for Geocomputation, National University of Ireland Maynooth: Maynooth, 2003.
- (40) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. Royal Stat. Soc. Series B* **1995**, *57*, 289–300.
- (41) Benjamini, Y.; Yekutieli, D. The Control of the False Discovery Rate in Multiple Testing under Dependency. *Ann. Stat.* **2001**, *29*, 1165–1188.
- (42) Shaffer, J.P. Multiple Hypothesis Testing. *Ann. Rev. Psychol.* **1995**, *46*, 561–584.
- (43) Byrne, G.; Charlton, M.; Fotheringham, A.S. Multiple Dependent Hypothesis Tests in Geographically Weighted Regression. In *10th International Conference on Geocomputation*; Lees, B.; Laffan, S., Eds.; Sydney, 2009; pp 1–5.
- (44) Belsley, D.; Kuh, E.; Welsch, R. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*; Wiley: New York, 1980.
- (45) Griffith, D.A. Spatial-Filtering-Based Contributions to a Critique of Geographically Weighted Regression (GWR). *Environ. Plan. A* **2008**, *40*, 2751–2769.
- (46) Wheeler, D.; Waller, L. Comparing Spatially Varying Coefficient Models: A Case Study Examining Violent Crime Rates and Their Relationships to Alcohol Outlets and Illegal Drug Arrests. *J. Geogr. Syst.* **2009**, *11*, (1), 1–22.
- (47) Wheeler, D.C. Simultaneous Coefficient Penalization and Model Selection in Geographically Weighted Regression: The Geographically Weighted Lasso. *Environ. Plan. A* **2009**, *41*, 722–742.
- (48) Páez, A.; Farber, S.; Wheeler, D. A Simulation-Based Study of Geographically Weighted Regression as a Method for Investigating Spatially Varying Relationships. *Environ. Plan. A* **2011**, *43*, 2992–3010.

- (49) Brunsdon, C.; Charlton, M.; Harris, P. Living with Collinearity in Local Regression Models. *Spatial Accuracy Conference Brazil*, 2012.
- (50) Páez, A.; Uchida, T.; Miyamoto, K. A General Framework for Estimation and Inference of Geographically Weighted Regression Models: 1. Location-Specific Kernel Bandwidths and a Test for Locational Heterogeneity. *Environ. Plan. A* **2002**, *34*, 733–754.
- (51) Ripley, B. *MASS: Support Functions and Datasets for Venables and Ripley's MASS*, 7.3-29; CRAN, 2013.
- (52) Huang, B.; Wu, B.; Barry, M. Geographically and Temporally Weighted Regression for Modeling Spatio-Temporal Variation in House Prices. *Int. J. Geogr. Inform. Sci.* **2010**, *24*, 383–401.
- (53) Crespo, R. Statistical Extensions of GWR: Spatial Interpolation and a Spatiotemporal Approach. PhD thesis. National University of Ireland, Maynooth: Maynooth, 2009.
- (54) Chen, V.Y.-J.; Deng, W.-S.; Yang, T.-C.; Matthews, S.A. Geographically Weighted Quantile Regression (GWQR): an Application to U.S. Mortality Data. *Geogr. Anal.* **2012**, *44*, 134–150.
- (55) Wang, Q.; Zhao, P.; Ren, H.; Kakubari, Y. Spatiotemporal Dynamics of Forest Net Primary Production in China over the past Two Decades. *Global Planet. Change* **2008**, *61*, 267–274.