An Extension of Geographically Weighted Regression with Flexible Bandwidths

Wenbai Yang[1], A. Stewart Fotheringham[1], Paul Harris[2]

[1]School of Geography & Geosciences,
University of St Andrews, Scotland, UK
Tel.: +44(0)1334 462894 Fax: +44(0)1334 463949
Email: wy5@st-andrews.ac.uk
[2]National Centre for Geocomputation, National University of Ireland Maynooth, Ireland

KEYWORDS: GWR, Bandwidth, Backfitting, Simulation, Nonstationarity

1. Introduction

Geographically weighted regression (GWR) (Brunsdon et al. 1996; Fotheringham et al. 2002) is a useful technique for modelling local spatial relationships between variables. The essential idea of GWR is that observations near to a model calibration point have more influence in the estimation of regression coefficients than observations farther away do. The standard GWR model employs a single bandwidth to control the distance-decay in this influence. In practice however, such a uniform bandwidth may not be sufficient in reflecting complex spatial variations in relationships between dependent and independent variables. In an attempt to produce a more realistic model, this paper develops an extension to GWR, where flexible bandwidths are found providing coefficient surfaces that vary at different spatial scales. Experiments are carried out on simulated datasets to test the model.

2. Background

In GWR a series of local regressions are calibrated at target regression locations. Observations are weighted according to their proximity to the regression point so that data from near observations are weighted more than data from far observations. This geographical weighting is achieved through a kernel function with a given bandwidth, which determines the rate at which the weights decay around a regression point. The larger a bandwidth is, the more slowly the weights decay. When the bandwidth tends to infinity, the model will tend to a global regression where the relationships are stable over space.

Current approaches in GWR allow a bandwidth to be selected by some optimizing criteria such as cross-validation, Akaike Information Criterion (AIC) or AICc (small sample bias corrected AIC) (Fotheringham et al. 2002). All the relationships examined in the model are assumed to follow this uniform 'one-size-fits-all' bandwidth. However we may want to weight the observations differently, by each independent variable, using different rates of distance-decay to reflect a multivariate process that varies across different spatial scales. In such circumstances, a flexible bandwidth GWR (FBGWR) model can be specified.

3. Methodology

The form of FBGWR can be written as:

$$y_i = \beta_{bw1}(u_i, v_i)x_{i1} + \beta_{bw2}(u_i, v_i)x_{i2} + ... + \varepsilon_i \tag{1}$$

where y_i is the dependent variable measured at observation point i; x_{i1}, x_{i2}, … are the independent variables (including an intercept of ones) at point i; (u_i, v_i) stands for the location of point i; ε_i is the error term; $\beta_{bw1}(u_i, v_i)$, $\beta_{bw2}(u_i, v_i)$, … are coefficients describing the relationships between y_i and different independent variables around location (u_i, v_i). The regression coefficients vary with location *and* for each independent variable, vary at a different spatial scale. This property distinguishes FBGWR from standard GWR.

One approach to calibrate FBGWR is to use the backfitting method (Hastie et al. 2001) which is similarly used by Brunsdon et al. (1999) in calibratring semi-parametric GWR. The idea is to calibrate each term in turn, assuming that all the other terms are known. Partial residuals are regressed on each individual variable in an iterative manner, where each step will give new calibrations for each term, and eventually these should converge, provided some regularity conditions apply to all hat matrices. In this way, all the calibrations are solved simultaneously.

4. Experiment

To evaluate the performance of a FBGWR model, a well-advised practice is to design experiments on simulated datasets where properties of the data, including size, distribution, variation and heterogeneity can be controlled. Model evaluation using simulated data avoids problems due to any unwanted effect that is present in empirical data; effects that often compromise gaining a clear understanding of the model.

4.1 Simulation data

In this experiment, three datasets are simulated following the approach proposed by Farber and Páez (2007); Wang et al.(2008); and used in Harris et al. (2010), for investigating GWR models. Here 625 observation points are located on a 25*25 grid, and a data generating process is defined as

$$y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{1i} + \varepsilon_i \quad \text{for } i = 1, 2, \ldots, 625, \tag{2}$$

where y_i is the generated dependent variable; x_{1i} is a single independent variable randomly drawn from a uniform distribution over interval (0, 1); and ε_i is an error term independently drawn from a normal distribution with mean zero and deviation at a proportion of 33.3% to the variance of the mean process. The two coefficients, β_0 (for the intercept) and β_1 (for the single independent variable) are specified as functions of (u_i, v_i) in the following three cases.

Case 1: $\beta_0 = 3$, $\beta_1 = 3$

Case 2: $\beta_0(u, v) = 1 + (1/6)(u + v)$, $\beta_1(u, v) = 1 + u/3$

Case 3: $\beta_0(u, v) = 1 + 4 \sin[(1/12)\pi u]$, $\beta_1(u, v) = 1 + (1/324)[36 - (6 - u)^2][36 - (6 - v)^2]$

Each case represents a different heterogeneity level, with zero heterogeneity in case 1, low heterogeneity in case 2 and high heterogeneity in case 3. Thus three simulated data sets are built, each with different properties in data relationships. Coefficient surfaces for the latter two cases are depicted in Figure 1.

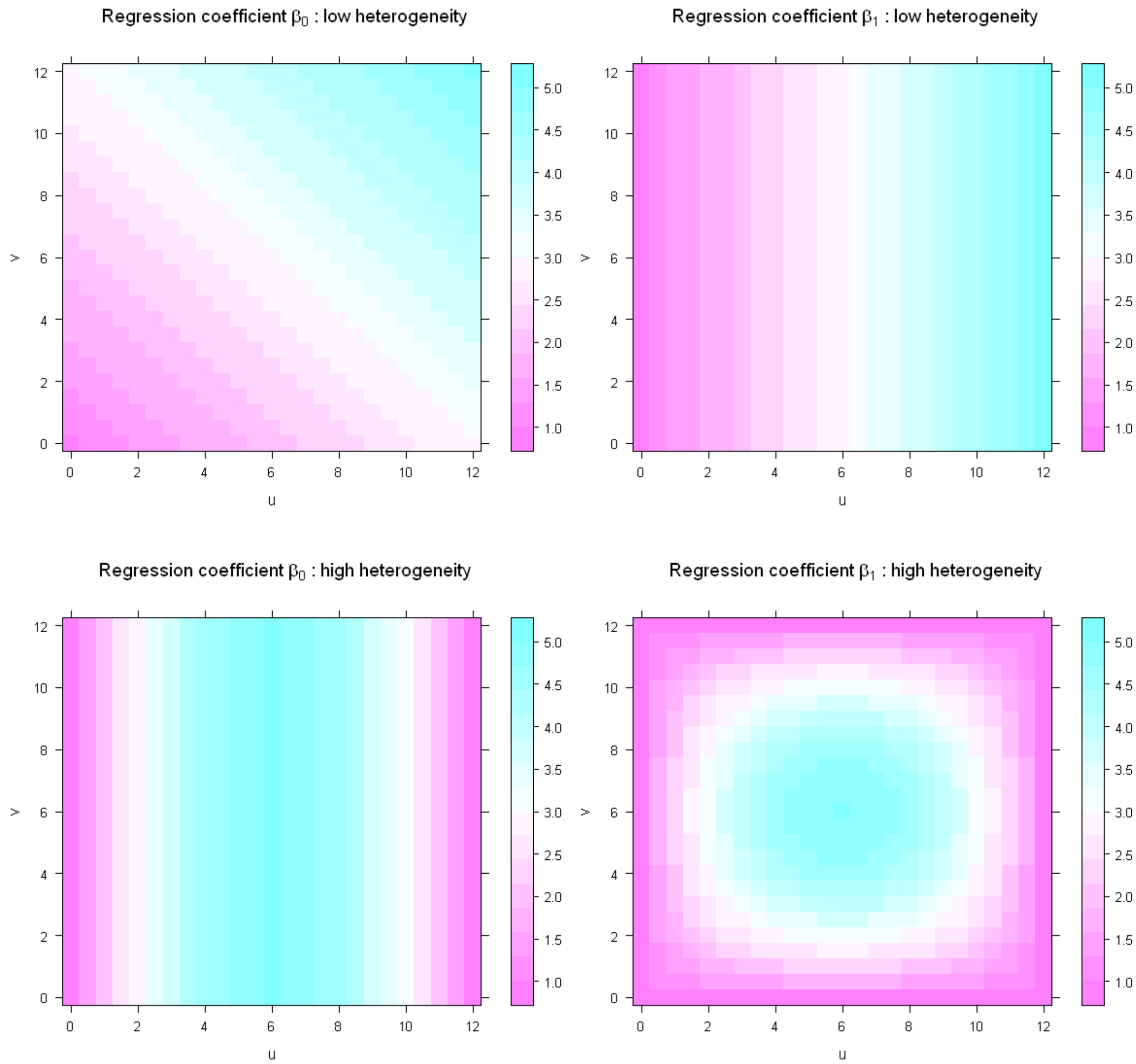

Figure 1. Simulated coefficient surfaces for case 2 (the upper) and case 3 (the lower).

4.2 Model calibration and results

As a first step, the performance of FBGWR is compared to standard GWR in prediction accuracy and ability to reproduce the coefficient surfaces. Both models are calibrated using an adaptive bi-square kernel function and AICc to select an optimal bandwidth.

Table 1 compares the results from the two models for each case. According to the residual sum of squares (RSS) for the predicted and actual y_i data, FBGWR performs better than GWR in cases 1 and 3. In each case, FBGWR has tuned clearly different bandwidths for the two coefficients, suggesting different scales of relationship nonstationarity. The bandwidth of 1 is considered to probably reflect the existence of a stationary coefficient.

Table 1. Results from GWR and FBGWR on simulated data sets

		Case/data 1 (zero heterogeneity)		Case/data 2 (low heterogeneity)		Case/data 3 (high heterogeneity)	
		GWR	*FBGWR*	*GWR*	*FBGWR*	*GWR*	*FBGWR*
Bandwith	Intercept(β_0)	0.36	0.12	0.20	0.17	0.20	0.05
	X1(β_1)		(appr.)1		0.28		0.19
RSS		153.69	147.07	596.56	599.59	841.54	826.53

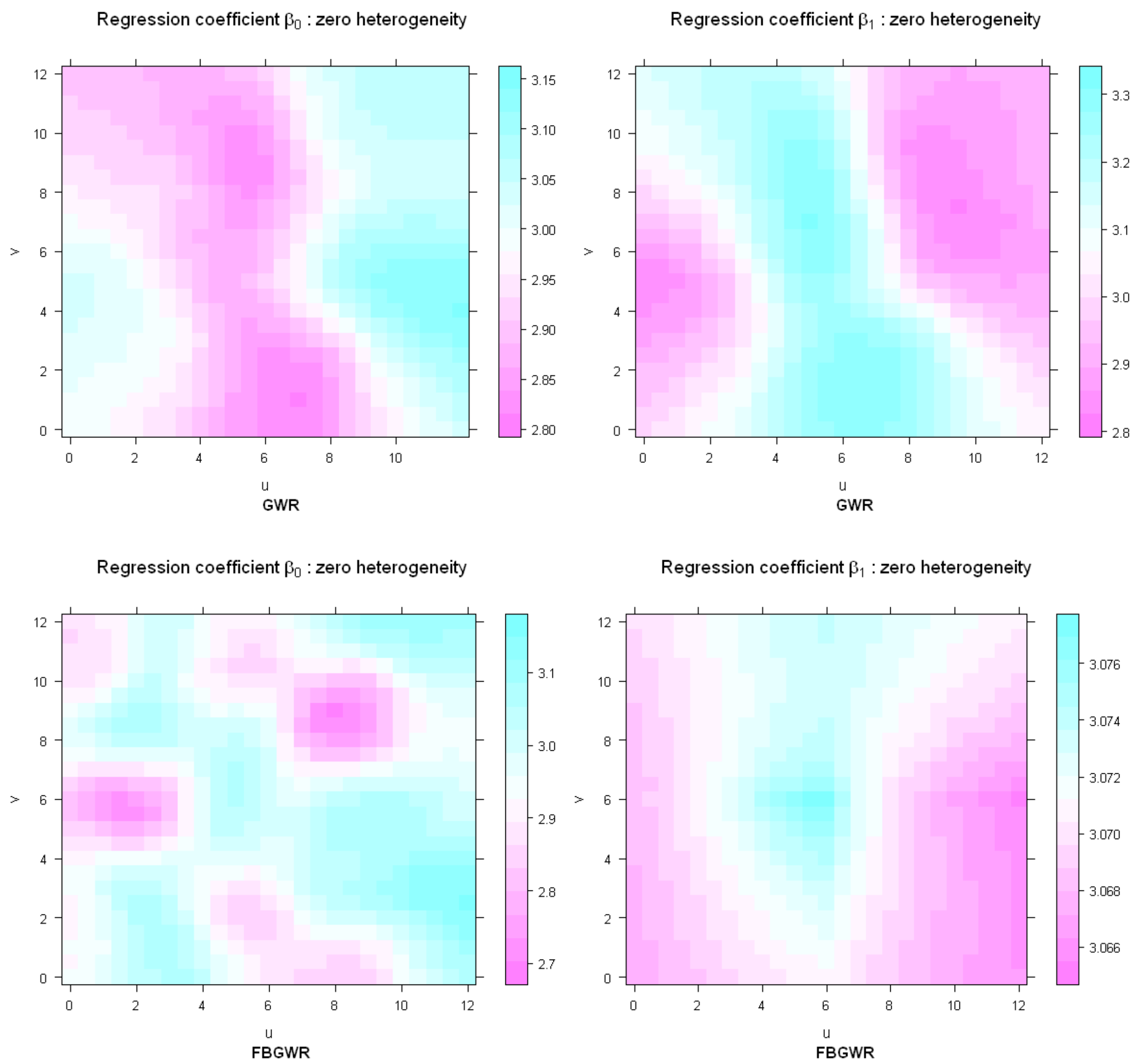

Figure 2. Estimated coefficient surfaces for case 1

The estimated coefficients are mapped for each case, in Figures 2 to 4, to be compared with real surfaces given in Figure 1. In case 1, the estimated coefficients from GWR show some vague patterning but tend to hover around 3 as expected. For FBGWR, the intercept β_0 tends to random, whilst β_1 is a little over-estimated. In case 2, FBGWR performs better than GWR in estimating both coefficients. In case 3, FBGWR also reproduces the two coefficient surfaces quite well, while standard GWR hardly represents the real patterns at all, especially β_1 the more complex surface.

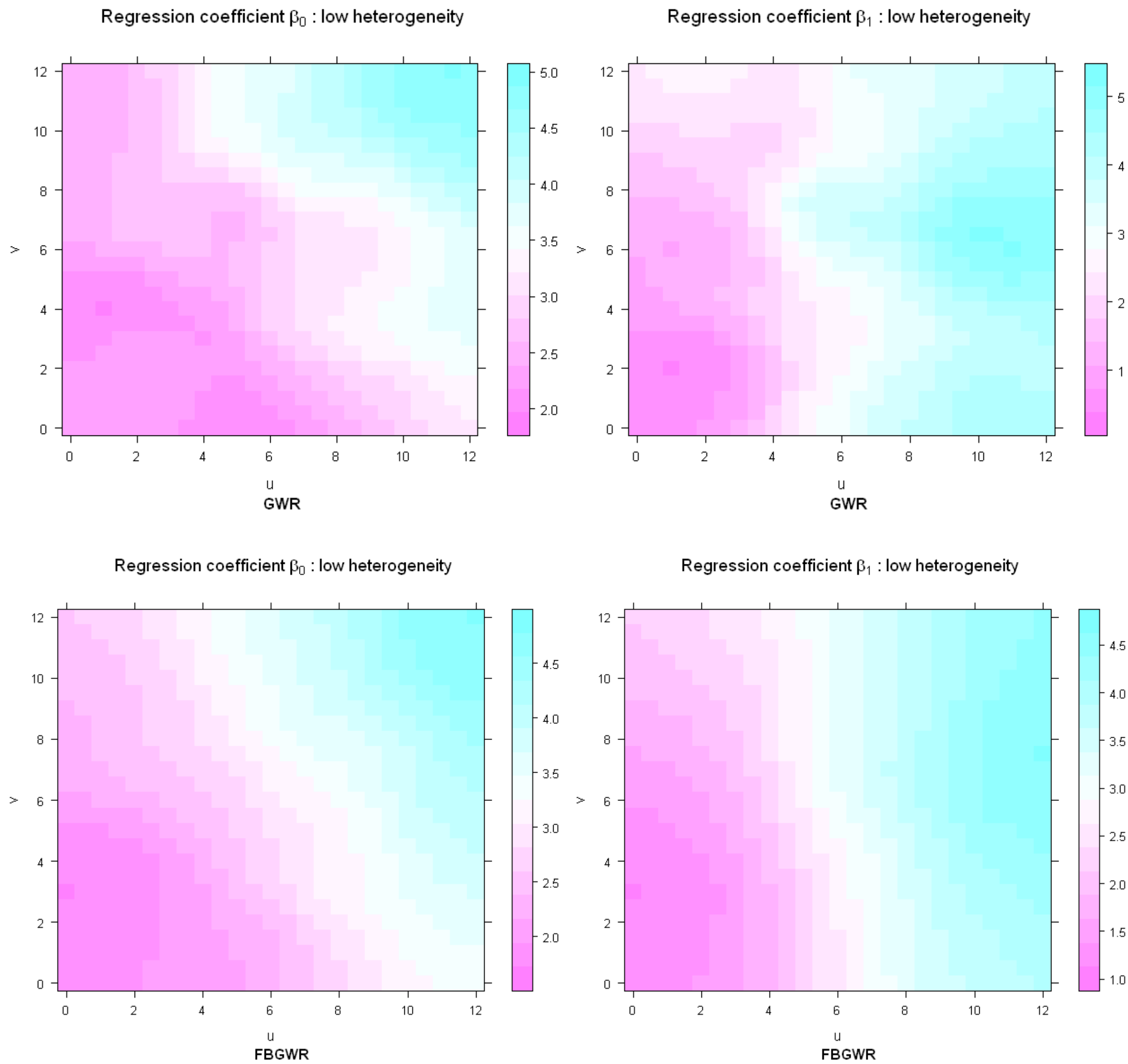

Figure 3. Estimated coefficient surfaces for case 2

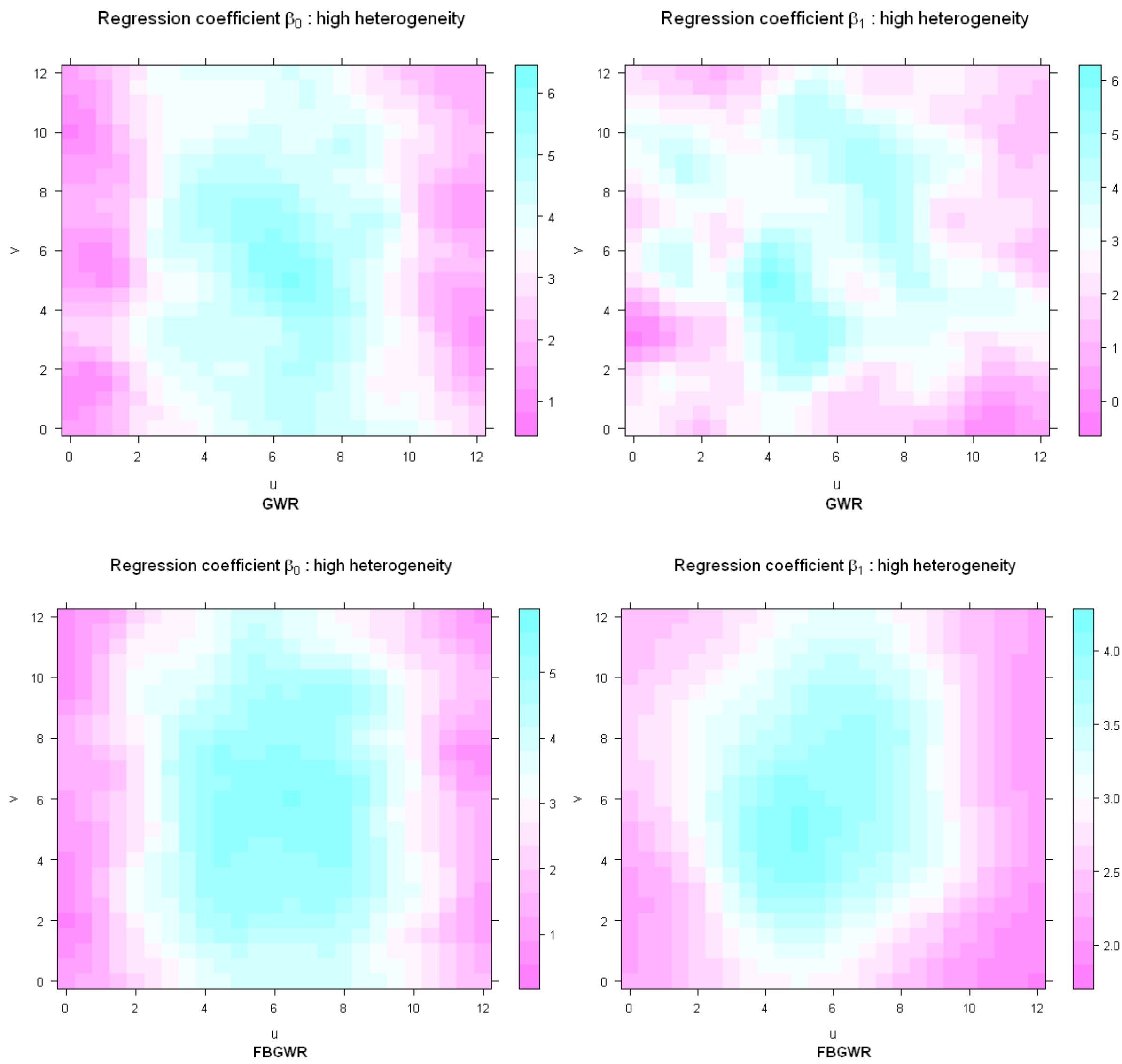

Figure 4. Estimated coefficient surfaces for case 3

5. Discussion

FBGWR enables an investigation of data relationships that may vary at different spatial scales, by allowing a different bandwidth to be selected for each coefficient. In doing so, FBGWR acts as a generalisation to simpler models – global linear regression, standard GWR and semi-parametric GWR. These preliminary simulation experiments suggest that FBGWR can provide an improvement over standard GWR, when the spatial variation of coefficients is complex. A more complete set of simulation experiments currently underway will investigate FBGWR with: (i) more independent variables; (ii) simulated data sets derived from coefficients with more complex levels of heterogeneity and (iii) multiple simulations to test various hypotheses. With respect to (ii), if the same surfaces are used for all the coefficients, FBGWR should reproduce the results of a standard GWR. Alternatively, if the coefficient surfaces reflect two very different levels of heterogeneity, zero and high heterogeneity, FBGWR should work equally to a semi-parametric GWR. The efficiency of backfitting algorithm needs to be tested on more complex models and techniques to accelerate the algorithm are under investigation.

6. Acknowledgements

Research presented in this article 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.

7. References

Brunsdon, C., Fotheringham, A.S. & Charlton, M., 1996. Geographically Weighted Regression: A method for Exploring Spatial Nonstationarity. *Geographical Analysis*, 28, pp.281-298.

Brunsdon, C., Fotheringham, A.S. & Charlton, M., 1999. Some Notes on Parametric Significance Tests for Geographically Weighted Regression. *Journal of Regional Science*, 39(3), pp.497-524.

Farber, S. & Páez, A., 2007. A systematic investigation of cross-validation in GWR model estimation: empirical analysis and Monte Carlo simulations. *Journal of Geographical Systems*, 9(4), pp.371-396.

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

Harris, P. et al., 2010. The Use of Geographically Weighted Regression for Spatial Prediction: An Evaluation of Models Using Simulated Data Sets. *Mathematical Geosciences*, 42(6), pp.657-680.

Hastie, T., Tibshirani, R., & Friedman, J., 2001 *The elements of statistical learning*, Springer

Wang, N., Mei, C.-L. & 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), pp.986-1005.

8. Biography

Wenbai Yang has a background in computer science and GIS, has worked in National Centre for Geocomputation, National University of Ireland Maynooth for two years, is now a PhD student in School of Geography & Geosciences, University of St Andrews, UK. Her interests include geocomputation, spatial analysis and computer science.

Stewart Fotheringham is Professor of Human Geography and Director of the Centre for GeoInformatics at the University of St Andrews。

Paul Harris is a research fellow at the National Centre for Geocomputation and has a PhD in Geostatistics from Newcastle University.