GWmodel: an R package for exploring spatial heterogeneity

Binbin Lu[1], Paul Harris[1], Isabella Gollini[1], Martin Charlton[1], Chris Brunsdon[2]

[1]National Centre for Geocomputation, National University of Ireland Maynooth, Maynooth, Co.kildare, Ireland

Tel. (353) 1 7086208 Fax (353) 1 7086456

binbin.lu@nuim.ie, http://ncg.nuim.ie/

[2]School of Environmental Sciences, University of Liverpool, Liverpool, UK

KEYWORDS: geographically weighted regression, geographically weighted principal components analysis, spatial prediction, spatial statistics, R package.

1. Introduction

In the very early developments of quantitative geography, statistical techniques were invariably applied at a ‘global’ level, where moments or relationships were assumed constant across the study region (Fotheringham and Brunsdon, 1999). However, the world is not an “average” space but full of variations and as such, statistical techniques needed to account for different forms of spatial heterogeneity or non-stationarity (Goodchild, 2004). Consequently, a number of local methods were developed for modelling spatial heterogeneity. Many such methods model non-stationary relationships via some regression adaptation, such as: the expansion method (Casetti, 1972), random coefficient modelling (Swamy et al., 1988), multilevel modelling (Duncan and Jones, 2000) and space varying parameter models (Assunção, 2003).

Among these local regression techniques, geographically weighted regression (GWR) (Brunsdon et al., 1996) has become increasingly popular and has been broadly applied in many disciplines outside of its quantitative geography roots, including: regional economics, urban and regional analysis, sociology, ecology and environmental science. Currently, there are several toolkits available for applying GWR, such as **GWR** 3.x (Charlton et al., 2007), **GWR** 4.0 (Nakaya et al., 2009), the **GWR toolkit** in ArcGIS (ESRI, 2009), the R packages **spgwr** (Bivand and Yu, 2006) and **gwrr** (Wheeler, 2011), and **STIS** (Arbor, 2010). However, most focus on the fundamental functions of GWR or some specific issue - e.g. **gwrr** provides tools to diagnose collinearity. Therefore as a major extension, we are currently developing an integrated framework for handling spatially varying structures via a wide range of geographically weighted (GW) models, not just GWR. All functions are included in an R package named **GWmodel**. This package is also mirrored with a set of GW modelling tools for ESRI’s ArcGIS written in Python.

2. GWmodel package

The **GWmodel** package is developed under the open source **R** software coding environment (R Development Core Team, 2011). This package includes all common GW modelling approaches as well as some newly developed ones. It will provide an integrated and comprehensive toolset for modelling spatially heterogeneous processes. Currently the package consists of the following four core components:

- GW Summary Statistics (GWSS)

Functions to calculate GWSSs (Brunsdon et al., 2002; Harris and Brunsdon, 2010): including the GW mean, GW standard deviation, GW skewness and GW correlation.

- GW Principal Components Analysis (GWPCA)

Functions to implement the GWPCA methodology (Harris et al., 2011a) for investigating the changing local structure in multivariate spatial data sets, including: i) automatic bandwidth selection, ii) randomisation tests for its application and iii) visualisation techniques for its output.

- GW Regression and GW Generalized Linear Models (GWGLM)

Functions to implement GWR including statistical tests for relationship non-stationarity, model specification and visualization tools of its results. Functions are also included to implement GW Poisson regression (Nakaya et al., 2005) and GW logistic regression (Atkinson et al., 2003). Furthermore, a selection of newly-developed, locally-compensated GWR models is available to combat issues of local collinearity in GWR(Brunsdon et al., 2012).

- GW Prediction Models:

Functions to implement GWR as a spatial predictor (Harris et al., 2011b) and to also implement a selection of new hybrids where kriging is combined with some form of GW approach (Harris et al., 2010a; Harris et al., 2010b; Harris and Juggins, 2011).

Common to all four core components is the ability to choose from:

1. a range of kernel functions (Gaussian, bi-square, tri-cube and box-car)
2. fixed and adaptive bandwidths
3. a flexible choice of distance metric (Lu et al., 2011; Lu et al., 2012)
4. basic and robust GW model forms. For example, robust GWSS (Brunsdon et al., 2002; Harris and Brunsdon, 2010); robust GWR (Fotheringham et al., 2002; Ghosh and Manson, 2008; Harris et al., 2010c); and robust GWPCA (Harris et al., 2012).

Further GW model functions not listed here, will be integrated accordingly.

#

Acknowledgements

We gratefully acknowledge support from a Strategic Research Cluster grant (07/SRC/I1168) by Science Foundation Ireland under the National Development Plan.

References

ARBOR, A. (2010) TerraSeer Space-Time Analysis of Health Data Workshop. Michigan, http://www.directionsmag.com/pressreleases/terraseer-space-time-analysis-of-health-data-workshop-may4-5-2010/120433.

ASSUNÇÃO, R. M. (2003) Space varying coefficient models for small area data. *Environmetrics,* 14**,** 453-473.

ATKINSON, P. M., GERMAN, S. E., SEAR, D. A. & CLARK, M. J. (2003) Exploring the Relations Between Riverbank Erosion and Geomorphological Controls Using Geographically Weighted Logistic Regression. *Geographical Analysis,* 35**,** 58-82.

BIVAND, R. & YU, D. (2006) spgwr: Geographically weighted regression. http://cran.r-project.org/web/packages/spgwr/index.html.

BRUNSDON, C., CHARLTON, M. & HARRIS, P. (2012) Living with collinearity in Local Regression Models. *Spatial Accuracy 2012.* Brazil.

BRUNSDON, C., FOTHERINGHAM, A. S. & CHARLTON, M. (2002) Geographically weighted summary statistics -- a framework for localised exploratory data analysis. *Computers, Environment and Urban Systems,* 26**,** 501-524.

BRUNSDON, C., FOTHERINGHAM, A. S. & CHARLTON, M. E. (1996) Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis,* 28**,** 281-298.

CASETTI, E. (1972) Generating Models by the Expansion Method: Applications to Geographical Research. *Geographical Analysis,* 4**,** 81-91.

CHARLTON, M., FOTHERINGHAM, A. S. & BRUNSDON, C. (2007) Geographically Weighted Regression: Software for GWR. National Centre for Geocomputation.

DUNCAN, C. & JONES, K. (2000) Using Multilevel Models to Model Heterogeneity: Potential and Pitfalls. *Geographical Analysis,* 32**,** 279-305.

ESRI (2009) ArcGIS 9.3: Interpreting GWR results. http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Interpreting_GWR_results.

FOTHERINGHAM, A. S. & BRUNSDON, C. (1999) Local Forms of Spatial Analysis. *Geographical Analysis,* 31**,** 340-358.

FOTHERINGHAM, A. S., BRUNSDON, C. & CHARLTON, M. (2002) *Geographically Weighted Regression: the analysis of spatially varying relationships,* Chichester, Wiley.

GHOSH, D. & MANSON, S. M. (2008) Robust Principal Component Analysis and Geographically Weighted Regression: Urbanization in the Twin Cities Metropolitan Area (TCMA). *URISA Journal,* 20**,** 15-25.

GOODCHILD, M. F. (2004) The Validity and Usefulness of Laws in Geographic Information Science and Geography. *Annals of the Association of American Geographers,* 94**,** 300-303.

HARRIS, P. & BRUNSDON, C. (2010) Exploring spatial variation and spatial relationships in a freshwater acidification critical load data set for Great Britain using geographically weighted summary statistics. *Comput. Geosci.,* 36**,** 54-70.

HARRIS, P., BRUNSDON, C. & CHARLTON, M. (2011a) Geographically weighted principal components analysis. *International Journal of Geographical Information Science,* 25**,** 1717-1736.

HARRIS, P., BRUNSDON, C. & CHARLTON, M. (2012) Multivariate spatial outlier detection: a comparison of techniques. *geoENV 2012.* Valencia, Spain.

HARRIS, P., BRUNSDON, C. & FOTHERINGHAM, A. (2011b) Links, comparisons and extensions of the geographically weighted regression model when used as a spatial predictor. *Stochastic Environmental Research and Risk Assessment,* 25**,** 123-138.

HARRIS, P., CHARLTON, M. & FOTHERINGHAM, A. (2010a) Moving window kriging with geographically weighted variograms. *Stochastic Environmental Research and Risk Assessment,* 24**,** 1193-1209.

HARRIS, P., FOTHERINGHAM, A., CRESPO, R. & CHARLTON, M. (2010b) The Use of Geographically Weighted Regression for Spatial Prediction: An Evaluation of Models Using Simulated Data Sets. *Mathematical Geosciences,* 42**,** 657-680.

HARRIS, P., FOTHERINGHAM, A. S. & JUGGINS, S. (2010c) 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**,** 286-306.

HARRIS, P. & JUGGINS, S. (2011) Estimating Freshwater Acidification Critical Load Exceedance Data for Great Britain Using Space-Varying Relationship Models. *Mathematical Geosciences,* 43**,** 265-292.

LU, B., CHARLTON, M. & 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.

LU, B., CHARLTON, M. & HARRIS, P. (2012) Geographically Weighted Regression using a non-euclidean distance metric with simulation data. *2012 First International Conference on Agro-Geoinformatics.*

NAKAYA, T., FOTHERINGHAM, A. S., BRUNSDON, C. & CHARLTON, M. (2005) Geographically weighted Poisson regression for disease association mapping. *Statistics in Medicine,* 24**,** 2695-2717.

NAKAYA, T., FOTHERINGHAM, A. S., CHARLTON, M. & BRUNSDON, C. (2009) Semiparametric geographically weighted generalised linear modelling in GWR 4.0. IN LEES, B. & LAFFAN, S. (Eds.) *10th International Conference on GeoComputation.* UNSW Sydney, Australia.

R DEVELOPMENT CORE TEAM (2011) R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing.

SWAMY, P. A. V. B., CONWAY, R. K. & LEBLANC, M. R. (1988) The stochastic coefficients approach to econometric modeling, part 1: a critique of fixed coefficients models. Board of Governors of the Federal Reserve System (U.S.).

WHEELER, D. (2011) Package gwrr: Geographically weighted regression with penalties and diagnostic tools. http://cran.r-project.org/web/packages/gwrr.

Biography

Binbin Lu *is a postdoctoral research fellow at the National Centre for Geocomputation (NCG). He received his PhD on "Geographically weighted regression with non-Euclidean distance metrics" from the NCG in 2012.*

Paul Harris *is a postdoctoral research fellow in the NCG. Paul was formally employed as a Geostatistician for Goldfields, South Africa; a research associate in the Dept. of Sociology, Durham; and a research associate in the Dept. of Geography, Manchester. Paul has degrees in Statistics and in Geostatistics.*

Isabella Gollini *received an Italian-French double degree in Mathematical and Computational Statistics in 2006, and completed her MSc degree in Statistics in 2008. In 2012, Isabella completed her PhD in Statistics at the University College Dublin. Since March 2012, Isabella has been a postdoctoral research fellow at the NCG.`*

Martin Charlton *is an expert in the use of Geographical Information Systems and has been a leading international researcher in this area for over 25 years. Together with Chris Brunsdon and Stewart Fotheringham, Martin is one of the originators of Geographically Weighted Regression.*

Chris Brunsdon *is Professor of Human Geography at the University of Liverpool. Chris is a leading international researcher in geographical information science.*