Model Selection Issues in Geographically Weighted Regression

A. Stewart Fotheringham[1], Martin Charlton[1], Chris Brunsdon[2] and Tomoki Nakaya[3]

1 National Centre for Geocomputation, National University of Ireland, Maynooth, Co Kildare, Ireland

2. School of Computing, University of Glamorgan, Wales

3. Department of Geography, Ritsumeikan University, Kyoto, Japan

Bio of First Author

Professor Stewart Fotheringham is a Science Foundation Ireland Research Professor and Director of the National Centre for Geocomputation (NCG) at the National University of Ireland, Maynooth. He has long-standing research interests in GIS-based spatial analysis and spatial modelling and is one of the co-developers of Geographically Weighted Regression.

National Centre for Geocomputation
John Hume Building
National University of Ireland, Maynooth
Maynooth
Co. Kildare
IRELAND

Tel: +353 708 6455
Fax: +353 708 6456
Web: ncg@nuim.ie
email: stewart.fotheringham@nuim.ie

Abstract

The objective of ordinary least squares (OLS) regression is to produce a single set of parameter estimates given data on a dependent variable and one or more independent variables. When applied to spatial data, the single set of relationships described by the parameter estimates are assumed to apply equally to all parts of the region from which the data are drawn. That is, the relationship between the dependent variable and any independent variable is assumed to be stationary over space. However, non-stationarity can occur for two reasons: i) there are intrinsic differences in relationships over space; and ii) the regression equation is not perfect and includes either incorrect functional forms of relationships between variables (such a non-linear relationship between two variables being described by

a linear one) and/or excludes relevant variables. In the case of the latter, these relevant variables may be unknown to the researcher or unmeasurable. Whatever, the cause of non-stationarity, it is possible to measure its intensity and to map it using geographically weighted regression (GWR). In GWR, the data are weighted geographically around a point in space so that neighbouring data weight more heavily than data further away. In this way, different parameter estimates are produced for each point in space so that the resulting set of local parameter estimates can be mapped and spatial variations in relationships explored.

Although GWR is now an established technique with increasing numbers of applications of the technique appearing in the literature, almost all GWR models have been calibrated by allowing all of the relationships in the model to vary spatially. In some instances, however, it is difficult to justify why some relationships should be allowed to vary spatially. In others, empirical results may suggest that some relationships are stationary over space while others vary significantly. In these instances, 'mixed' GWR models, where some relationships are allowed to vary spatially while others are held constant, would seem to be more appropriate. The question then becomes how one decides on whether a relationship should be fixed globally or allowed to vary locally. Theory often cannot help here and we need to rely on empirical findings. One method of doing this is to perform an analogous procedure to ordinary stepwise regression within a GWR framework. In 'stepwise' GWR, we allow all possible combinations of stationary and spatially varying parameters to exist and then select the model form yielding the optimal fit to the data using Akaike Information statistics as a guide. We demonstrate an example of this procedure using 2002 Irish census data. Hopefully, we will also be able to demonstrate the use of the latest version of the GWR software, GWR 4.0, which allows mixed GWR models to be calibrated.

In addition to describing the technical issues surrounding the use of 'stepwise' GWR procedures and describing the application of this technique to census data, the paper will also demonstrate the application of the technique to a data set with known properties. We will describe a simulated data set in which we allow some relationships to vary spatially and others to be stationary and we demonstrate how well stepwise GWR performs in uncovering these known relationships. This simulated data set has the advantage in that we know that whether or not the relationships uncovered by GWR are spurious or real.

Key words: Geographically Weighted Regression, spatial non-stationarity, mapping relationships, Stepwise GWR.