Time: the late arrival at the Geocomputation party and the need for considered approaches to spatio-temporal analyses

Alexis Comber[1], Paul Harris*[2], Narumasa Tsutsumida[3]

[1]School of Geography, University of Leeds, Leeds, LS2 9JT, UK
[2]Rothamsted Research, North Wyke, Okehampton, Devon, EX20 2SB, UK
[3]Graduate School of Global Environmental Studies, Kyoto University, Kyoto 606-8501
*Email: a.comber@leeds.ac.uk

Abstract

This study uses exploratory approaches for the robust development of a *Geographical and Temporal Weighted Regression* (GTWR) model in the context of limited theoretical constructs for modelling spatio-temporal processes via a kernel-based model. These include establishing the appropriateness of a GTWR by considering autocorrelation and relationship heterogeneity effects, and also visualising all potential space-time kernel bandwidths, so that the GTWR model is correctly specified, relative to the problem being investigated. For demonstration, this study uses a 16-year livestock population data set for Mongolia linked to environmental, climatic, socio-economic and agricultural covariates in order to predict livestock populations using GTWR.

Keywords: GTWR, Spatio-temporal Analyses, Space-time Bandwidth Optimisation.

1. Introduction

A number of recent papers have sought to develop and apply methods for spatio-temporal analyses with geographically weighted (GW) models, and a subset have been published under the banner of *Geographical and Temporal Weighted Regression* (GTWR), building upon GW regression (GWR) (Brunsdon et al., 1996). The real advance is that GTWR models propose spatially and temporally weighted kernels (Huang et al., 2010; Yu, 2014; Wrenn and Sam, 2014; Fotheringham et al., 2015). The methodology of a GTWR can be considered the first step to expanding any GW approach to the space-time case within the generic GW framework (Gollini et al., 2015).

In brief, GW models calculate a series of local statistics or local models using data falling under a moving kernel, with the data weighted by their distance to the kernel centre. They reflect Tobler's first law of geography and allow spatial non-stationarity to be explored and handled. Kernel bandwidth selection is a critical step GW modelling and bandwidth selection. It can be optimised by minimising some measure of model fit such as AIC or leave-one-out cross validation. The elegance of the GW framework is for two main reasons. First, it reflects the spatial pattern of many observed anthropogenic and environmental processes: birds of a feather *do* flock together. Second, it is underpinned by well-established and universally agreed theoretical frameworks describing spatial non-stationarity of many process and distance decay, that is rarely challenged by observations and measurements.

A number of GW methods exist for handling spatio-temporal data including various GW panel regression formulations (Yu, 2010; Lin, 2011; Cai et al., 2012; Tabak et al., 2013). However, the critical issue in GTWR is how to calibrate the kernel bandwidth. In contrast to space, the temporal properties of many processes and associated relationships do not simply decay *with* time as it were (i.e. as time increases) in contrast to many spatial processes. Rather than a simple decay, they may have a cyclic nature – the daily commute, weekly shopping, seasonal weather patterns – and have some degree of regularity. Additionally, the distribution of many processes in time at any given location, may exhibit alternate, abrupt or cyclic behaviours. Universally agreed geographical frameworks such as Tobler's first law and the MAUP (Openshaw, 1984) have not yet been established for spatio-temporal processes. As yet there is no well-developed geographical philosophy about how spatio-temporal processes behave, or ways to overcome the issue that distances in space and time are measured in different metrics. Although much theoretical work can be found within various sophisticated statistical frameworks, the techniques are commonly suited to continuous processes, such as those resulting from air pollution monitoring (Cressie and Wikle, 2011).

In the context of GTWR, there are two critical problems in bandwidth selection that have not been satisfactorily overcome in previous research. The first is the space-time distance metric: space and time are measures in different dimensions (Huang et al, 2010). The second is computational. Fotheringham et al. (2015) note that to fully investigate all possible bandwidth combinations in a space-time space would require an extremely large number of potential bandwidths to be computed and evaluated. We deal with both of these issues in this paper. Previous work has applied different approaches to optimise the GTWR space-time bandwidth. Huang et al. (2010) and Yu (2014) used different methods to determine a single parameter to weight and relate spatial and time kernels. Huang et al. (2010) incorporated a "temporal heterogeneity" measure to weight the distance in time between data for two locations. Yu (2014) modelled the salience of past observations to the question being asked. Wrenn and Sam (2014) used the Mahalanobis distance to calculate the distance from each observation in space and time. A third approach is to optimize with respect to time and space simultaneously, which despite computationally demanding (Fotheringham et al, 2015) is transparent and makes less assumptions about the processes under investigation. The latter is the approach adopted in this research.

However, rather than just proceeding with a GTWR analysis, this paper advocates taking a fresh look at space and time by stepping back and reflecting on the *nature* of the process under consideration and specifying temporal kernels that reflect their temporal properties. We use a Mongolian livestock data set covering a 16-year period (described in Tsutsumida et al., in press) to examine changes in cattle in relation to changes in environmental, climatic, socio-economic and agricultural covariates.

2. Data and Analysis

Annual data of cattle numbers for 1990-2006 for 341 soums (second-level administrative units) in Mongolia were linked to annual data describing mean NDVI, annual rainfall, the number of households working with livestock and the annual number reported cattle losses.

An extensive year on year initial investigation was undertaken to test for the presence of spatial autocorrelation in the global regression errors via Moran's *I,* together with the likelihood of relationship spatial heterogeneity via the reporting of optimised GWR bandwidths. AIC values for each annual global regression, annual simultaneous autoregressive regression and annual GWR were also found. This analysis is reported elsewhere, but clear spatial effects were observed both for autocorrelation and for heterogeneity, for all 16 years. The AIC values strongly suggested that a relationship heterogeneity regression was appropriate for each year, rather than one employing autocorrelation effects. This initial, purely spatial set of analyses, when considered as a whole, also alluded to likely temporal dependencies. Thus moving to a GTWR model appears worthwhile. This exploratory analysis is missing in much of the reported research on GTWR, which seems to have inadvertently adopted the spirit of McNoleg (1998).

In order to deal with the bandwidth problems of dimension and dimensionality, an optimal (adaptive) space time kernel was found via GTWR fits to all possible combinations of bandwidths from 5 to 100% in steps of 1%, for each of the 16 years, sequentially adding data from previous years. From this the optimum temporal and spatial bandwidth combination could be selected via a minimised AIC approach. Thus the first model included all data from 2006 (i.e. simply a 'GWR' fit), the second model included all data from 2006 plus temporally weighted data from 2005 (i.e. the 'first GTWR' fit), the third included all data from 2006 plus temporally weighted data from 2005 and 2004 (i.e. 'subsequent GTWR' fits), etc. In the temporal dimension, data were weighted using a bi-square function such that data from the earliest years were weighted the least. A bi-square function was also employed in the spatial dimension.

3. Results and Discussion

The resultant AIC map is presented in Figure 1, where an optimal temporal bandwidth was identified at 12 years (i.e. data are temporally weighted, starting from 2005 and ending in 1994) and an optimal spatial bandwidth was identified as the nearest 35% of the 341 soums.

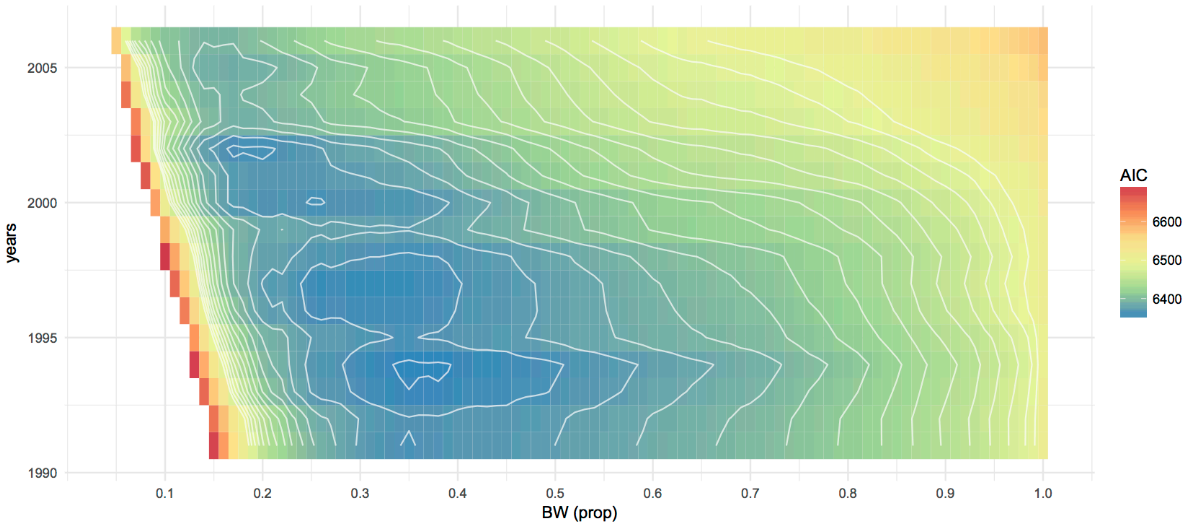

Figure 1: the AIC values for bandwidths from 1 to 100% (x axis) with data from 1990 to 2006 (y axis). The contours and shading indicate the local minima and AIC values greater than 7000 have been removed in each case.

As an example application, GTWR was used to predict cattle numbers for each soum in 2006. The scatterplot of predicted against observed is shown in Figure 2 and the prediction errors are mapped in Figure 3.

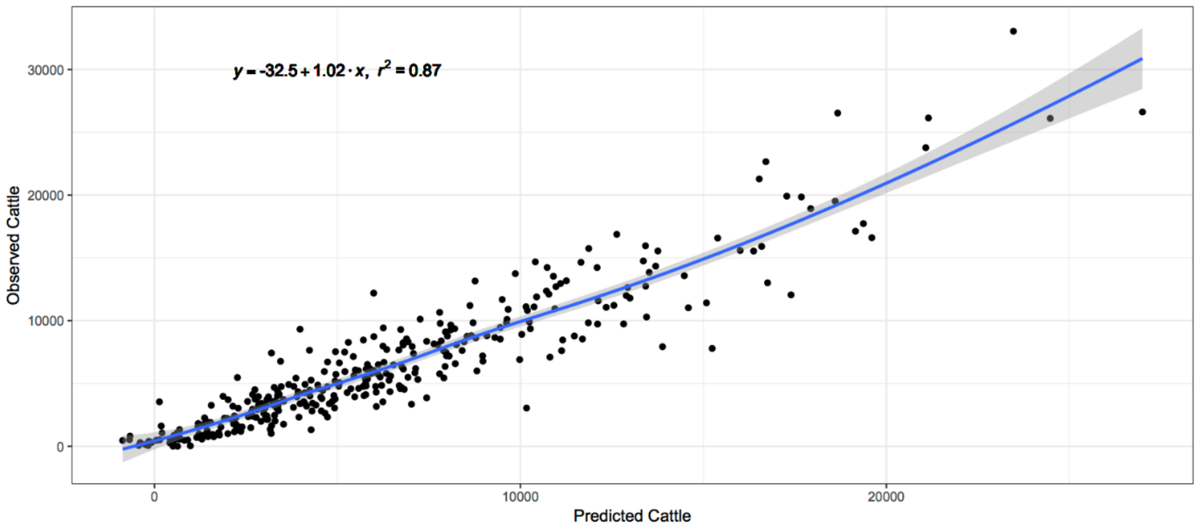

Figure 2: The predicted cattle values (from a GTWR model incorporating temporally and spatially weighted data) against observed values for each soum in Mongolia, for 2006.

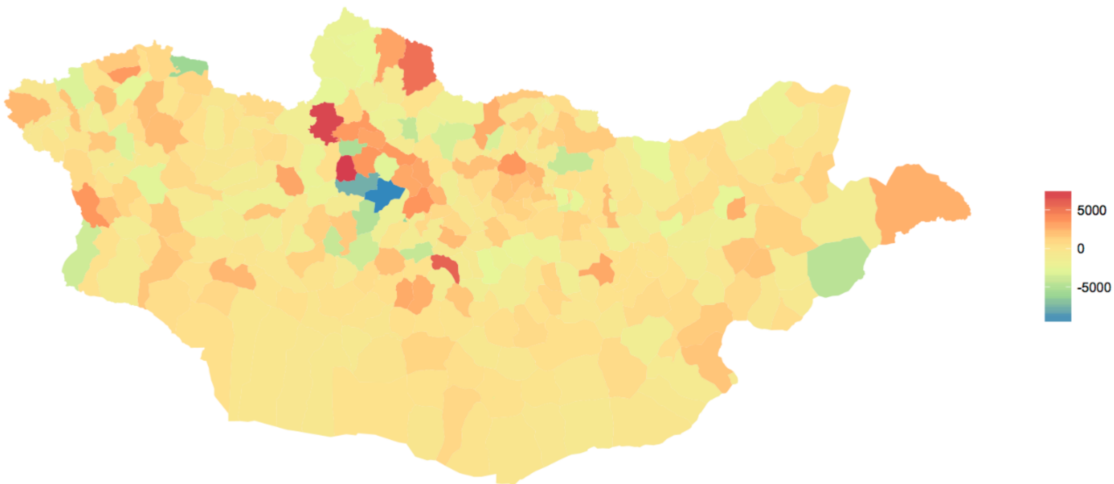

Figure 3: The mapped errors for GTWR predicted versus observed cattle numbers in 2006.

The main advances made in this paper are as follows:

- It emphasises the importance of initial analyses to assess which (if any) spatial effects (autocorrelation or heterogeneity) are more dominant, before proceeding to a space-time regression analysis.
- Then, if GTWR is to be undertaken, the optimum spatial bandwidth and temporal bandwidths can be simultaneously determined, overcoming the issues of *dimension* and *dimensionality* identify by Huang et al (2010) and Fotheringham et al (2015), respectively.
- The shape and form of the bandwidths should reflect the logic of the question being asked of the data – in this case to use GTWR as a predictor. Here, both time and space were weighted using bi-square kernel, but these were of very different shapes: the spatial kernel incorporating data that were within the bandwidth in 2 dimensions around the kernel location, and the temporal kernel considered only data backwards in time from the time point under consideration.

This highly exploratory approach is missing in much of the reported research on GTWR. It allowed this research to: a) establish the appropriateness of a GTWR; and then b) to determine an appropriate bandwidth shape and size for the predictive task in hand. This emphasises the need for investigation and reflection when developing a GTWR analyses: don't just plug all the data in and press the button.

4. Acknowledgements

This research was supported the Natural Environment Research Council Newton Fund grant (NE/N007433/1), a UK Biotechnology and Biological Sciences Research Council grant (BB/J004308/1), Sinfonica Statistical GIS Research Grants, JSPS KAKENHI (15K21086) and KU SPIRITS project. The authors would like to thank the National Statistics Office of Mongolia for providing the livestock population data.

5. References

Brunsdon C, Fotheringham AS, Charlton M (1996). Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, 28, 281–289.

Cai, R., Yu, D., & Oppenheimer, M. (2012). Estimating the Effects of Weather Variations on Corn Yields using Geographically Weighted Panel Regression. In *Annual Meeting of the Agricultural and Applied Economics Association*. Seattle, http://ideas.repec.org/p/ags/aaea12/124627.html

Cressie & Wilke (2011) *Statistics for Spatio-Temporal Data*. Wiley.

Fotheringham, A. S., Crespo, R., & Yao, J. (2015). Geographical and temporal weighted regression (GTWR). *Geographical Analysis*, *47*(4), 431-452.

Gollini, I., Lu, B., Charlton, M., Brunsdon, C., & Harris, P. (2015). GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. *Journal of Statistical Software*, 63 (17), 1–50.

Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, *24*(3), 383-401.

Lin, Z. (2011). ML Estimation of Spatial Panel Data Geographically Weighted Regression Model. In *International Conference on Management and Service Science (MASS)*. Wuhan University.

McNoleg O (1998). Professor Oleg McNoleg's guide to the successful use of Geographical Information Systems (Ten ways to say nothing with GIS). *International Journal of Geographical Information Science*, 12: 429-430.

Openshaw, S. (1984). The modifiable areal unit problem. Geo Abstracts University of East Anglia.

Tabak, B. M., Miranda, R. B., & Fazio, D. M. (2013). A geographically weighted approach to measuring efficiency in panel data: The case of US saving banks. *Journal of Banking & Finance*, *37*(10), 3747–3756

Tsutsumida N., Harris P., and Comber A (in press). The application of a geographically weighted principal components analysis for exploring 23 years of goat population change across Mongolia. Paper accepted for publication in *the Annals of the American Association of Geographers*.

Wrenn, D. H., & Sam, A. G. (2014). Geographically and temporally weighted likelihood regres- sion: Exploring the spatiotemporal determinants of land use change. *Regional Science and Urban Economics*, *44*(1), 60–74.

Yu, D. (2010). Exploring spatiotemporally varying regressed relationships: the geographically weighted panel regression analysis. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *38*(2), 134–139.

Yu, D. (2014). Understanding regional development mechanisms in Greater Beijing Area, China, 1995–2001, from a spatial–temporal perspective. *GeoJournal*, *79*(2), 195-207.