Combining Geographically Weighted Regression and Geovisual Analytics to investigate temporal variations in house price determinants across London in the period 1980-1998

Urška Demšar, Stewart Fotheringham, Martin Charlton and Ricardo Crespo

National Centre for Geocomputation, National University of Ireland, Maynooth, Co. Kildare, Ireland
urska.demsar@nuim.ie, stewart.fotheringham@nuim.ie, martin.charlton@nuim.ie, ricardo.crespo@nuim.ie

1. Introduction

Hedonic price modelling attempts to uncover information on the determinants of prices - in this case the prices are those of houses in the Greater London area for the period between 1980 and 1998. The determinants of house prices can include house attributes (such as size, type of building, age, etc.), neighbourhood attributes (such as proportion of unemployed people in the neighbourhood or local tax rates) and geographic attributes (such as distance from the city centre or proximity to various amenities) (Orford 1999).

Almost all applications of hedonic price models applied to housing are in the form of multiple linear regression models where price is regressed on various attributes. The parameter estimates from the calibration of this type of regression model are assumed to yield information on the relative importance of various attributes in influencing price. One major problem with this approach is that it assumes that the determinants of prices are the same in all parts of the study area. This seems particularly illogical in this type of application where there could easily be local variations in preferences and also in supply and demand relationships. Hence, it seems reasonable to calibrate local hedonic price models rather than global ones – that is, to calibrate a model form which is flexible enough to allow the determinants of house prices to vary spatially. Geographically Weighted Regression (GWR) (Fotheringham et al. 2002) is a statistical technique that allows local calibrations and which yields local estimates of the determinants of house prices. GWR was recently used to investigate spatial variations in house price determinants across London separately for each of the years between 1980 and 1998 (Crespo et al. 2007).

The result of the GWR analysis is a set of continuous localised parameter estimate surfaces which describe the geography of the parameter space. These surfaces are typically visualised with a set of univariate choropleth maps for each surface which are used to examine the plausibility of the stationarity assumption of the traditional regression and different possible causes of non-stationarity for each separate parameter (Fotheringham and al. 2002). The downside of these separate univariate visualisations is that multivariate spatial and non-spatial relationships and patterns in the parameter space can not be seen. In an attempt to counter this inadequacy, in a previous study we suggested to treat the result space of one single GWR analysis as a multivariate dataset and visually explore it (Demšar et al. 2007). The goal was to identify spatial and multivariate patterns that the separate univariate mapping could not recognise. In this paper we extend this approach with the temporal dimension: we use Geovisual Analytical exploration to investigate the spatio-temporal dynamics in a time series of GWR hedonic price models. The idea is to merge the time series of GWR result spaces (one space per year) into one single highly-dimensional spatio-temporal dataset, which we then visually explore in an attempt to uncover information about the temporal and spatio-temporal behaviour of parameter estimates of GWR and consequently of underlying geographical processes.

2. Geovisual Analytical exploration

The spatio-temporal dataset that we constructed consists of a time series of nineteen GWR result spaces – one for each of the years 1980 to 1998. For each of these years, the GWR was run on a sample of 2500 property transactions and the results were interpolated into centres of a 1km grid covering the area of Greater London. In each GWR model, the house price was regressed on the following variables (Crespo et al. 2007): floor area, time of construction of the property (before WW1, between wars, post WW2, in the 1960, in the 1970s, in the 1980s in the 1990s), property type (detached, semi-detached, terraced, bungalow, flat), presence of garage, central heating and 2 or more bathrooms, and finally percentage of professionals and of unemployment in the census area where the property was located. Each GWR result space consisted of parameter estimates in grid centres for all the listed variables plus the estimate for the regression intercept and the value of the local r^2 statistic. When merged, these result spaces formed a 408–dimensional spatio-temporal dataset with 1567 elements (cells of the 1 km grid).

To explore the temporal and the spatio-temporal dynamics in this dataset we developed two approaches. The first approach separated the dataset into parameter-defined subspaces. The idea was to investigate the behaviour of each parameter estimate separately by visually exploring the spatial time series of the GWR estimates for this parameter. We tried to identify two types of patterns in each of these series: temporal patterns and spatio-temporal patterns. Temporal patterns were observed using an appropriately scaled temporal parallel coordinates plot (PCP) (Inselberg 2002, Edsall 2003). A combination of computational modelling with a Self-Organising Map (SOM) (Kohonen 1997, Takatsuka 2001) with several other visualisations (such as a map and a temporal PCP) seemed appropriate for identification of spatio-temporal patterns. Exploration systems used were built using GeoVISTA Studio (Gahegan et al. 2002). Selected preliminary results of this approach are presented in the next section.

The idea behind the second approach was to merge spatial time series of each parameter into larger groups based on the similarity of the parameters in the hedonic price model. For example, one such group could consist of parameters that represent building type. Another group would be all variables related to when the house was built. Each of these groups of spatial time series could then be treated as a single multivariate space for identification of patterns over time, space and group parameters. Due to the complexity and large dimensionality of these spaces, this has not yet been tried at the time of writing, but the idea is that a combination of the computational and visual data mining methods could facilitate identification of multivariate patterns in these subspaces.

3. Results

This section presents selected preliminary results of the exploration of the parameter-defined subspaces, where we were looking for temporal and spatio-temporal patterns in the behaviour of each separate paremeter.

a)

b)

c)

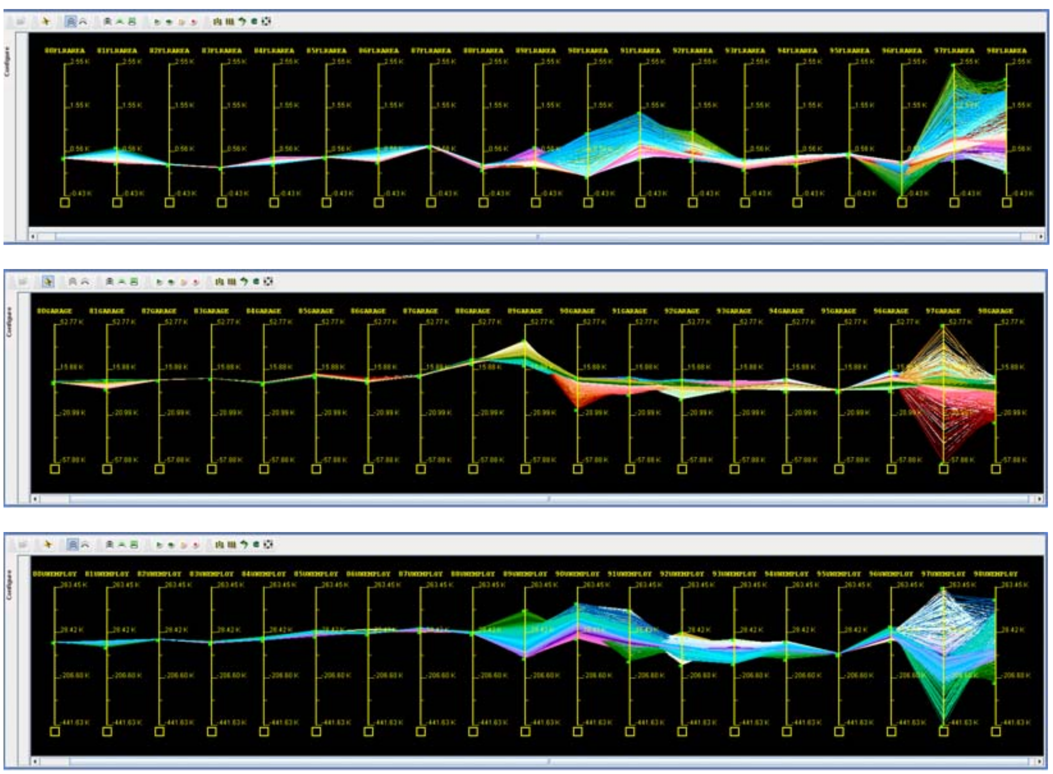

Figure 1. Time series of GWR estimates from 1980 to 1998 for three selected parameters: a) floor area for semi-detached houses, b) garage, c) percentage of unemployed.

Figure 1 shows a temporal PCPs of GWR estimates for three parameters: the floor area for semi-detached houses; the presence of a garage; and the percentage of unemployment in the neighbourhood of a property. The plots show the influence of each variable on housing price. Each axis represents one year from 1980-1998 and they are all scaled from min value across all years to max value across all years – this scaling variation is typically required in a PCP to visualise temporal trends. Each polygonal line represents one cell in the 1km grid covering the area of Greater London. All three plots exhibit a similar pattern: a relatively stable development of the influence of each of these three parameters during the time period 1980-1988 (represented by the first 9 axes), where the polygonal lines intersect the vertical axes within a relatively small range. Then there is an instability in the range around year 1990 (the range expands considerably), which can most probably be explained by the collapse of the housing market in London in 1990 (see the chart of the average housing prices in London in fig. 2). During 1992-1996 the influence becomes relatively stable again, but then something strange happens in 1996-1998, with the largest anomaly in 1997. This anomaly can not be explained with the chart of average prices in fig. 2. We hypothesise that it could perhaps be linked to the UK general election 1997, when buyers/sellers did not know how the change from Conservative to Labour politics would influence house prices and the market showed considerable uncertainty. Although here we only show the PCPs for three parameters in this abstract, similar trends with the 1990 and 1997 instabilities can be seen in temporal PCPs for all regression parameters. It should also be noted that the colours in fig. 1 are inherited from other visualisations and are not relevant for the identification of the described pattern.

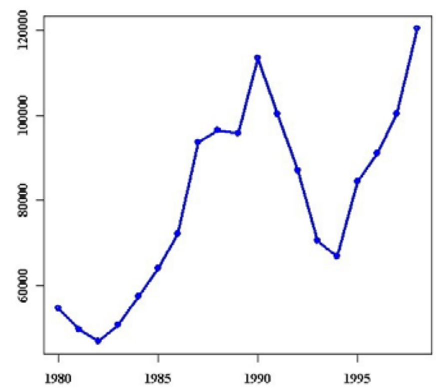

Figure 2. Average house prices (in GBP) in London during 1980-1998, updated by the inflation rate to year 1998.

Figure 3 shows an attempt to identify spatio-temporal patterns in the time series of GWR estimates for one particular regression parameter - the presence of a garage. The data was first clustered with a SOM (fig. 3b), which shows several clusters (light areas) separated with dark areas. The size of the circles overlaid over the SOM map shows how many elements are in each SOM cell and the colour of these circles is used for visual brushing to other visualisations (i.e. data elements that are in a cell of a particular colour in the SOM are shown with the same colour in other visualisations, for example, the map in fig. 3a). There are two clusters in the SOM that are very different from each other: the blueish-group in the lower left corner and the reddish group in the upper right corner. Figures 3c) and 3d) show a selection of these two clusters in the map and the temporal PCP respectively. Here each axis is scaled from min to max value of its respective attribute, which in contrast to consistent scaling of the axes (as in the previous example) hides temporal trends over the entire period, but on the other hand allows better visualisation of the dissimilarities of the two selected areas. From this visualisation it is clear that there are two completely different temporal trends at work in these two areas. Combining this observation from the temporal PCP with the locations of these two clusters in the map (central-west London and northeastern London) can give an indication of the influence of this particular variable on house prices in two different socio-economic environments – the affluent area of central London and a more deprived area in eastern London.

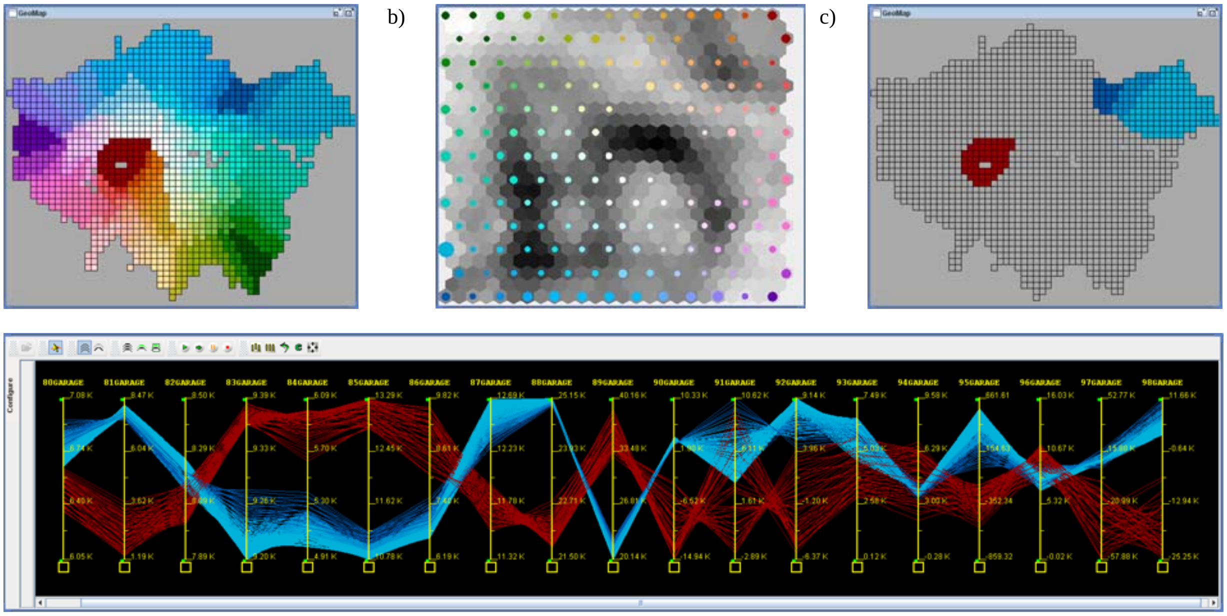

Figure 3. The temporal dataset of GWR estimates for the presence of the garage shown in a) the map and b) the SOM. Colours in the map and the PCP are inherited from the SOM. A selection of two areas in central and eastern London is shown c) on the map and d) in the PCP.

4. Conclusions and future work

In this abstract we present an example of how a geovisual exploration can be combined with a statistical method to obtain information about the spatio-temporal dynamics in hedonic house price modelling. The exploration provided new insights into the structure of the highly dimensional spatio-temporal dataset constructed from the GWR result spaces for each year. These insights have the potential to facilitate the interpertation of the temporal and spatio-temporal trends in the influence of each regression variable on house price.

The results presented in this abstract are selected preliminary observations from the parameter-based exploration approach. We intend to extend this by grouping the spatial time series of similar parameter estimates into larger subspaces. To identify the truly multivariate patterns, the entire spatio-temporal dataset could also be explored as a whole, without being divided into subsets of variables.

In the preliminary results presented in this abstract we identified patterns that most probably correspond to phenomena that occurred at a particular point in time, i.e. the collapse of the housing market in London in 1990 and the potential influence of the general election in UK in 1997. Alternatively we could look for evidence for longer lasting spatial processes or other spatio-temporal phenomena. For example, we could attempt to identify areas in London with particular socio-economic trends, such as gentrification or studentification. The question is what combination of visualisations and computational data mining methods would be most useful for such purposes – this is of course dependent on the type of exploration tasks. Therefore a typology of exploratory tasks for spatio-temporal exploration (such as for example Andrienko N et al. 2003) of the GWR results space should be considered for development of the most appropriate visualisation methodology.

Acknowledgements

The authors would like to thank Jan Rigby from The University of Sheffield for valuable suggestions regarding the development of London housing market and suggestions for interpretation of visually identified patterns.

References

Andrienko N., Andrienko G. and Gatalsky P. (2003) Exploratory spatio-temporal visualization: an analytical review. *Journal of Visual Languages and Computing*, 14:503-541.

Crespo R., Fotheringham A. S. and Charlton M. (2007) Application of Geographically Weighted Regression to a 19-year set of house price data in London to calibrate local hedonic price models. In: *Proceedings of the 9th International Conference on Geocomputation*, Sept 2007, Maynooth, Ireland.

Demšar U., Fotheringham A. S. and Charlton M. (2007) Combining Geovisual Analytics with Spatial Statistics: the example of Geographically Weighted Regression. Submitted for the special issue of *The Cartographic Journal* on Geovisual Analytics. Short version presented at the meeting of the ICA Commission on Visualization and Virtual Environments, "From geovisualization toward geovisual analytics", Helsinki, August 2007.

Edsall R. M. (2003) The parallel coordinate plot in action: design and use for geographic visualization. *Computational Statistics and Data Analysis*, 43:605-619.

Fotheringham A. S., Brunsdon C. and Charlton M. (2002) *Geographically Weighted Regression – the Analysis of Spatially Varying Relationships*. John Wiley & Sons Inc., Hoboken, New Jersey.

Gahegan M., Takatsuka M., Wheeler M. and Hardisty F. (2002) Introducing Geo-VISTA Studio: an integrated suite of visualization and computational methods for exploration and knowledge construction in geography. *Computers, Environment and Urban Systems*, 26:267-292.

Inselberg A. (2002) Visualization and data mining of high-dimensional data. *Chemometrics and Intelligent Laboratory Systems*, 60:147-159.

Kohonen T. (1997) *Self-Organizing Maps*, 2nd edition, Springer Verlag, Berlin-Heidelberg.

Orford S. (1999) *Valuing the Built Enviroment: GIS and house price analysis*. Ashgate, Aldershot.

Takatsuka M. (2001) ‘An application of the self-organizing map and interactive 3-D visualization to geospatial data’, in: *Proceedings of the 6th International Conference on Geocomputation*, Brisbane, Australia.