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, Ricardo Crespo

ICA Visualisation Workshop at the AGILE 2008 conference, “Geovisualisation of Dynamics, Movement and Change”

5th May 2008

National Centre for Geocomputation
National University of Ireland, Maynooth
urska.demsar@nuim.ie
**Spatial Statistics + Geovisual Analytics**

**Research topic:**
Can Geovisual Analytical exploration help with interpretation of results of a Spatial Statistical method and therefore facilitate understanding of complex spatio-temporal processes?

**Spatial Statistical method**
Geographically Weighted Regression

**Geovisual Analytical exploration**
Computational + Visual Data Mining methodology

1. **Space**
   ICA WS 2007, Helsinki

2. **Space and time**
   ICA WS 2008, AGILE

Real data on housing prices for the Greater London Area 1980-1998

Hedonic regression model

Dependent variable

house price

Independent variables

floor area

time of construction

property type

percentage of professionals

2 or more bathrooms

central heating

presence of garage

percentage of unemployment

Model with Geographically Weighted Regression: Local statistical technique - spatial variations in relationships

Regression parameters $\beta_{j}$ are continuous functions of location $i$

$$y_i = \beta_0(i) + \beta_1(i) x_{1i} + \beta_2(i) x_{2i} + \ldots + \beta_n(i) x_{ni} + \epsilon_i$$

with the estimator $\beta'(i) = (X^T W(i) X)^{-1} X^T W(i) Y$
Experiment

**Input data:**
house pricing data for 2500 property transactions for each year 1980-1998

Run one GWR analysis for each year, $t=1980,...,1998$

**GWR results:**
parameter estimate surfaces, interpolated in the centres of a 1km grid covering the area of Greater London

Merge all GWR result spaces

**Exploration data:**
merged GWR results = high-dimensional spatio-temporal dataset

Explore the data in a visual exploratory environment

**Questions:**
What do the patterns we find tell us about the dynamics of the London housing price market? Can we connect found discovered patterns with known events? Can we find evidence for different social trends (studentification, gentrification) in different areas?
Preliminary results
Temporal PCPs for separate parameter-defined subspaces

Time series of the influence of the respective parameter on the house price

- floor-area of semi-detached housing
- presence of garage
- percentage of unemployment

Stable development
Range expands
Stable again
Range expands even more

Similar pattern present in all PCPs. What is going on?

1990 Market crash
1997 UK general elections

National Centre for Geocomputation
An Ionzad Naolaitse Geofreannachta
Preliminary results

Parameter-defined subspace for presence of garage

Very different non-spatial clusters in the SOM, also at different spatial locations in the map.

Areas with different temporal behaviour

Affluent central area

workers’ housing in East London

selection
Main comments of the reviewers

"Are these patterns valid/consistent?"

"Some kind of validation analysis is missing."

Test the approach on a simulated artificial dataset

Create a time series of GWR models with known and controlled spatio-temporal behaviour of regression parameters

Question:
Can the patterns describing the known behaviour of parameters be identified through geovisual exploration in the spatio-temporal dataset of GWR results?
We repeat the same procedure as with real data:

**Input data:**
50x50 grid points in unity square [0,1]x[0,1] with 11 attributes:
\[ X_1(x,y,t), \ldots, X_{10}(x,y,t) \]
and \[ Y(x,y,t) = \sum_{j=1}^{10} \beta_j(x,y,t) X_j(x,y,t) \]

**Run one GWR analysis for each time, \( t=1,\ldots,8 \)**

**GWR results:**
parameter estimates
\[ \hat{\beta}_1(x,y,t), \ldots, \hat{\beta}_{10}(x,y,t) \]
and \( \hat{\beta}_0(x,y,t) \) for intercept for each time, \( t=1,\ldots,8 \)

**Explore the data in a visual exploratory environment**

**Merge all GWR result spaces**

**Exploration data:**
merged GWR results = 88-dimensional spatio-temporal dataset

**Question:**
can we visually identify patterns in parameter estimates \( \hat{\beta}_j(x,y,t) \) that correspond to behaviour of original parameters \( \beta_j(x,y,t) \) in space and time?
Exploration framework

**3 exploration approaches**

**Approach I** – explore parameter-defined subspace: time series of GWR parameter estimate $\hat{\beta}_j(x,y,t)$

**Approach II** – explore subspace defined by a group of similar parameters

**Approach III** – explore entire dataset at once

**3 types of patterns, related to interpretation of GWR results**

**Type 1** - areas of temporal stability
- temporal stability = approximately constant (GWR noise)
- $\rightarrow$ Approach I

**Type 2** - areas that behave similarly in time
- $\rightarrow$ Approach II
- Approach III

**Type 3** - parameters with similar temporal behaviour.
- $\rightarrow$ Approach II
- Approach III
Approach I – parameter-estimate-defined subspace

Pattern type 1 - areas of temporal stability

- **GeoMap**
- **SOM**

Expected:
values in a narrow band around the central axis $x=1/2$ are relatively stable

Seen:
Yes, beige/orange areas in the map and their tracks in both PCPs

Pattern type 2 - areas that behave similarly in time

- **SOM PCP**
  - SOM cells, nested-means scaling
- **Temporal PCP**
  - grid cells, min-max-over-the-entire-temporal-range linear scaling

Expected:
areas near both sides of the square increase/decrease steadily

Seen:
Yes, violet and green clusters in the SOM ($C_1$, $C_2$), their locations in the map, and their tracks in both PCPs
Expected:
- $\hat{\beta}_7$, $\hat{\beta}_8$, $\hat{\beta}_9$ – have a similar increase trend, but $\hat{\beta}_8$ mirrors the other two
- $\hat{\beta}_{10}$ is different from all other parameters

Seen: Yes, this can be clearly seen by comparing tracks in all the PCPs.
Does combined GWR+VDM approach work in on the simulated dataset? YES!

Future plans: detailed exploration of the GWR results produced by hedonic spatio-temporal modelling of the real house price data

Open topics/questions for discussion:

Other temporal or spatio-temporal visualisations?
At present the results are limited to what’s available in GeoVISTA Studio. What other visualisations would be appropriate for GWR results? Visualising spatial processes/temporal behaviour?

Would a similar combined approach be useful for interpretation of results of other spatial statistical methods besides GWR?

Applicability in practice & usability:
How difficult is it to understand all statistics/computational results (SOM) for potential users, who are not experts in visualisation?