GEOGRAPHICALLY WEIGHTED REGRESSION

Stewart Fotheringham
Chris Brunsdon
Martin Charlton

GIS and Spatial Analysis

- GIS are very useful for the storage, manipulation and display of spatial data
- They are less useful for the analysis of spatial data
- Have been repeated calls for this to change
- In some cases the link between GIS and spatial analysis has been a step backwards
- One important way the situation can be improved is to develop better spatial analytical tools that can take advantage of the features of GIS

An important catalyst for the better integration of GIS and spatial analysis has been the development of local spatial statistical techniques

Chief among these has been the development of Geographically Weighted Regression (GWR)

Local versus Global Statistics

Local statistics are spatial disaggregations of global statistics

- **Global**
  - similarities across space
  - single-valued statistics
  - non-mappable
  - GIS "unfriendly"
  - search for regularities
  - aspatial

- **Local**
  - differences across space
  - multi-valued statistics
  - mappable
  - GIS "friendly"
  - search for exceptions
  - spatial

Local versus Global

- **Local** versus **global data**: *the example of US climate data*

- **Local** versus **global relationships**: *the example of house price determinants*

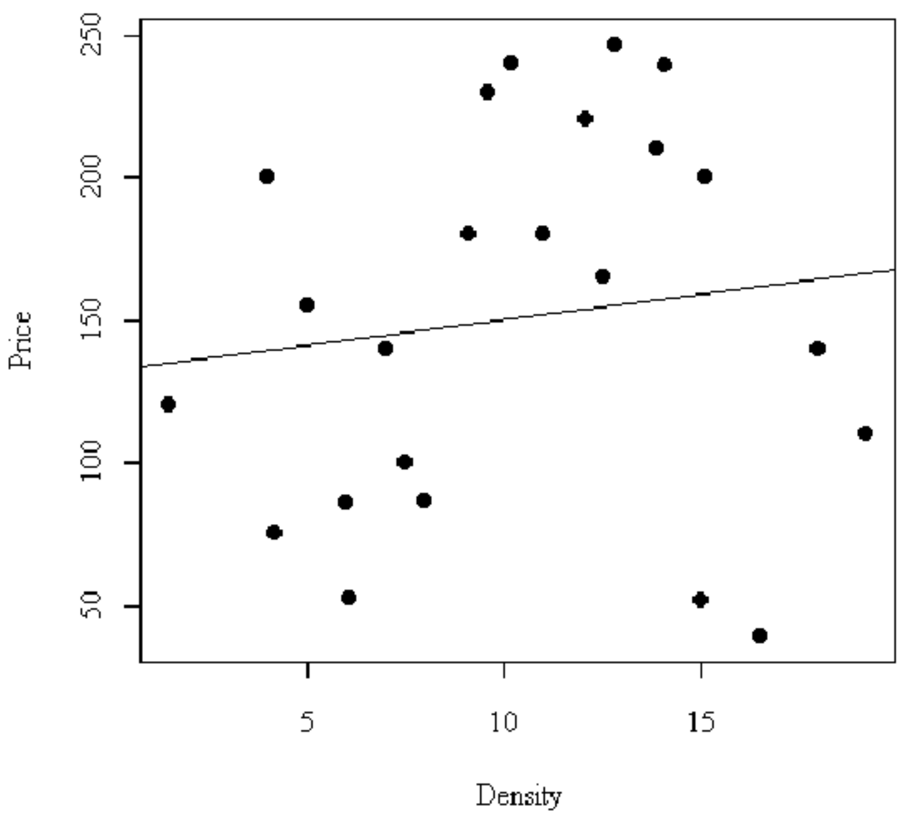
- **Local** versus **global models**: *the example of regression*

Why might relationships vary spatially?

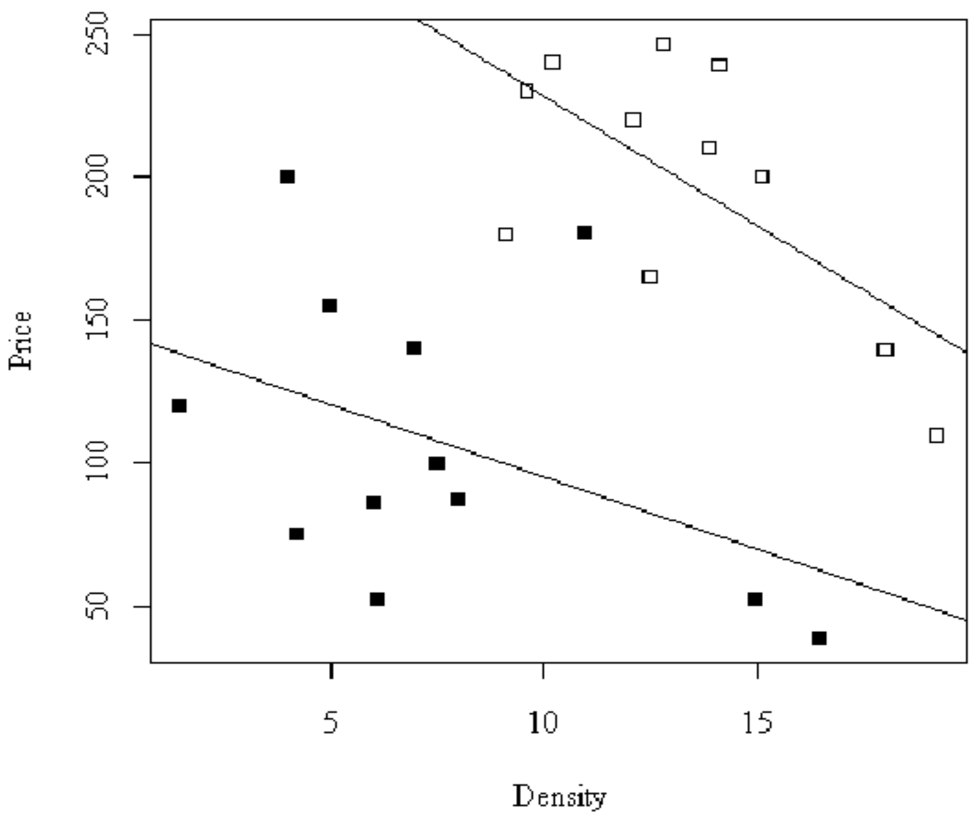
- Sampling variation
- Relationships intrinsically different across space *e.g. differences in attitudes, preferences or different administrative, political or other contextual effects produce different responses to the same stimuli - a post-modernist view*
- Model misspecification - *suppose a global statement can ultimately be made but models not properly specified to allow us to make it. Local models good indicator of how model is misspecified - a positivist view*
- Can all contextual effects ever be modelled?
- Can all significant variations in local relationships be removed?

Another reason for local modelling - Simpson's Paradox

Spatially aggregated data



Spatially disaggregated data



GEOGRAPHICALLY WEIGHTED REGRESSION

- The mechanics of GWR
- Software for GWR
- GWR in practice: an example of the determinants of London house prices
- Won’t discuss the math of GWR in much detail

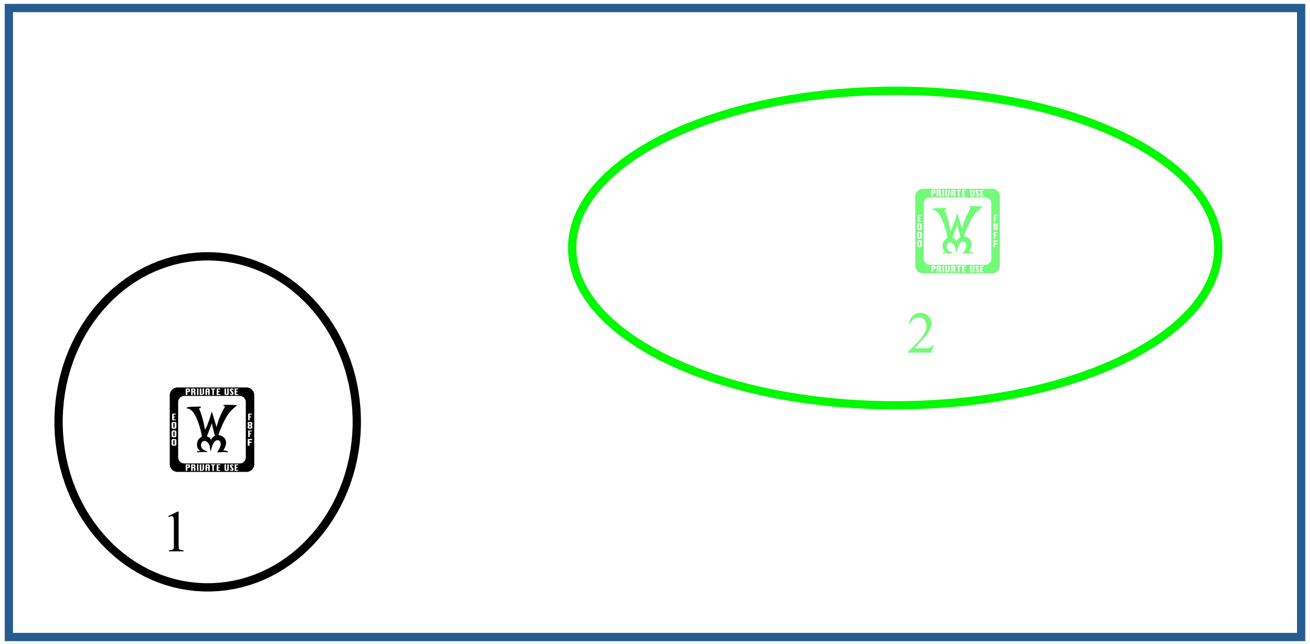
Regression

In a typical linear regression model applied to spatial data we assume a stationary *(the same stimulus provokes the same response in all parts of the study region)* process:

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

The assumption of stationarity in regression

$$y_i = \alpha + \beta\ x_i$$



Assumption is that the values of are the same everywhere.

Consequently…if there is spatial non-stationarity,

- We only see it through the residuals
- The residuals from a global model applied to a spatial non-stationary process will exhibit a marked spatial pattern
- Spatially dependent residuals violate the regression assumption of error independence and invalidate any inferences from the model

GWR and Spatial Autocorrelation

Suppose we have a non-stationary process that can be modelled by:

$$y_i = \alpha + \beta_i x_i$$

but we model it incorrectly with a global model of the form:

$$y_i = \alpha + \beta x_i$$

Real values of β_i

.9	.8	.8	.7	.5
.8	.7	.6	.5	.4
.7	.6	.5	.4	.4
.6	.5	.4	.3	.2
.5	.4	.3	.2	.1

Estimated value of β_i from global model

.5	.5	.5	.5	.5
.5	.5	.5	.5	.5
.5	.5	.5	.5	.5
.5	.5	.5	.5	.5
.5	.5	.5	.5	.5

Residuals (y_i - y_i')

+	+	+	+	0
+	+	+	0	-
+	+	0	-	-
+	0	-	-	-
0	-	-	-	-

To examine spatial dependency in the residuals,

- We might map the residuals from the regression to determine whether there are any spatial patterns.
- Or compute an autocorrelation statistic for the residuals
- We might even try to ‘model’ the error dependency with various types of spatial regression models e.g.Spacestat

However…

Why not address the issue of spatial nonstationarity directly and allow the relationships we are measuring to vary over space?

This is the essence of GWR

$$y_i = a_0(u,v) + \sum_{j=1 \dots m} a_m(u_i, v_i) x_{ij} + \epsilon_i$$

Where (u,v) refers to a location at which data on y and x are measured and at which local estimates of the parameters are obtained

… with the estimator

$$\hat{\boldsymbol{\beta}}(u,v) = (\mathbf{X}^T\mathbf{W}(u,v)\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}(u,v)\mathbf{y}$$

where **W(u,v)** is a matrix of weights specific to location (u,v) such that observations nearer to (u,v) are given greater weight than observations further away.

$$
\mathbf{W}(u,v) = \begin{pmatrix} w_1(u,v) & 0 & \cdots & 0 \\ 0 & w_2(u,v) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n(u,v) \end{pmatrix}
$$

where $w_i(u,v)$ is the weight given to data point i for the estimate of the local parameters at location (u,v)

Weighting schemes

Numerous weighting schemes can be used. They can be either **fixed** or **adaptive**.

Two examples of a fixed weighting scheme are the Gaussian function:

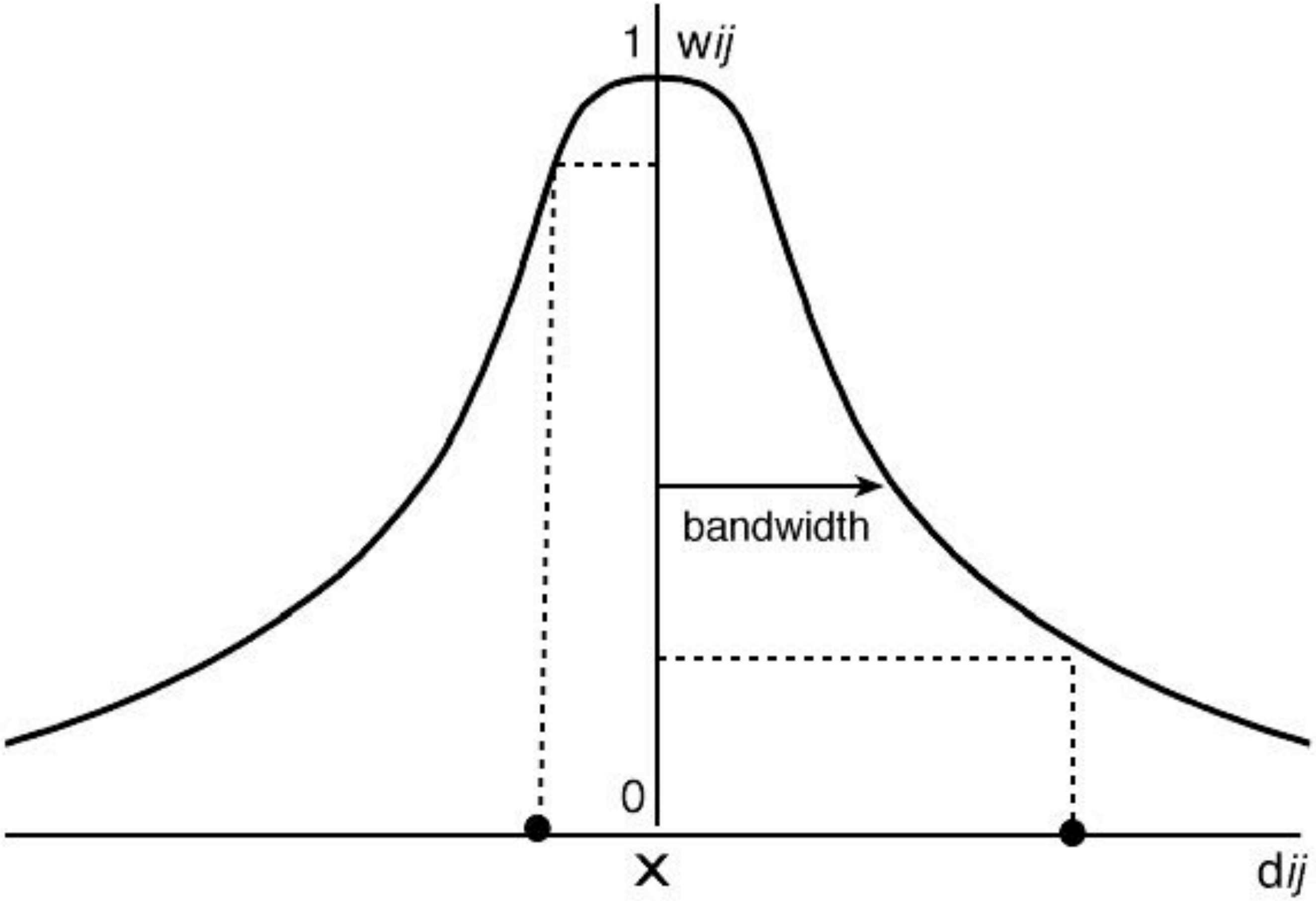
$$w_{ij} = \exp[-(d_{ij}^2 / h^2)/2]$$

where h is known as the bandwidth and controls the degree of distance-decay

and the bisquare function:

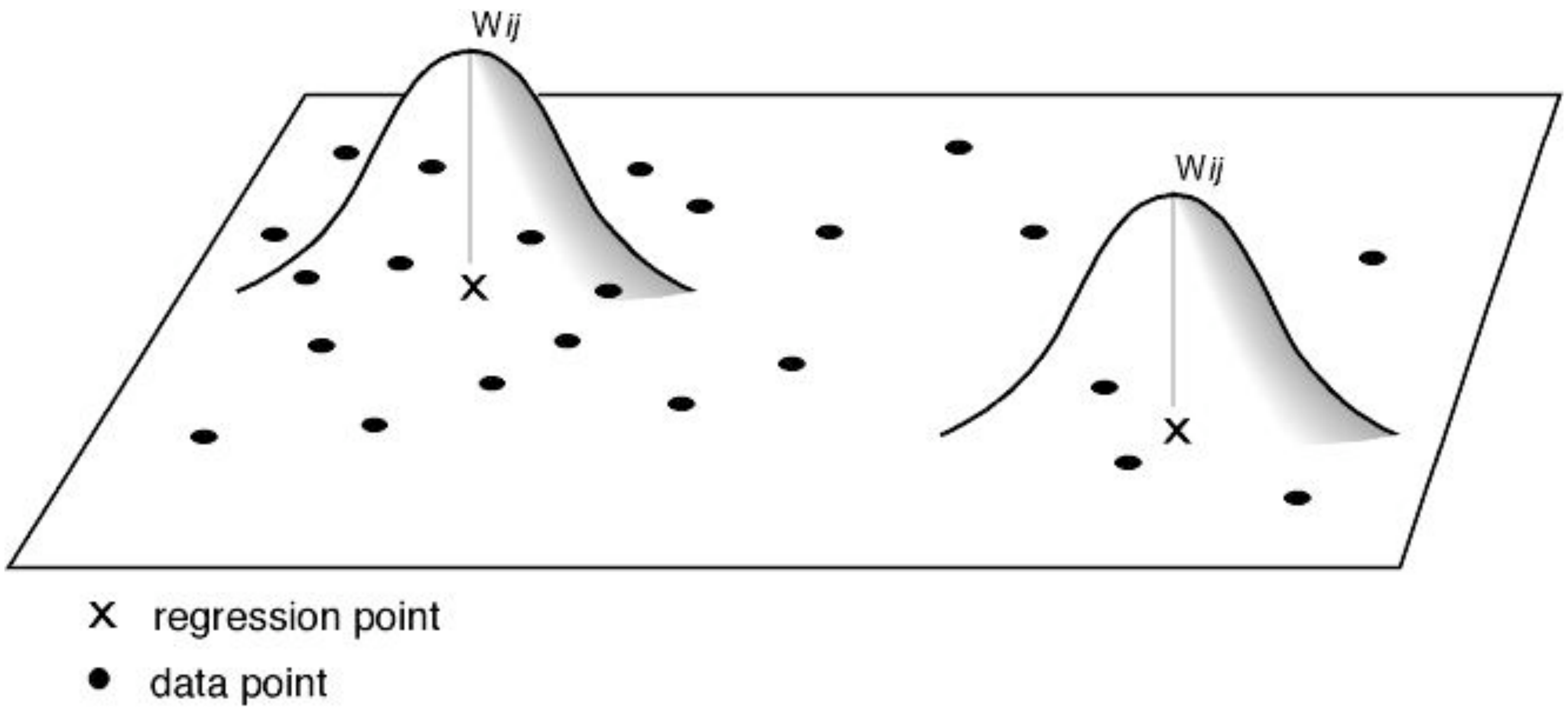
$$w_{ij} = [1-(d_{ij}^2 / h^2)]^2 \quad \text{if } d_{ij} < h$$

$$= 0 \qquad \text{otherwise}$$



x regression point wij is the weight of data point j at regression point i

• data point dij is the distance between regression point i and data point j


Wij
Wij
x regression point
• data point

Perhaps better…

Is to use a spatially adaptive weighting function such as:

$W_i\,(u,v) = \exp(-R_i(u.v)/\,h)$

where R is the ranked distance

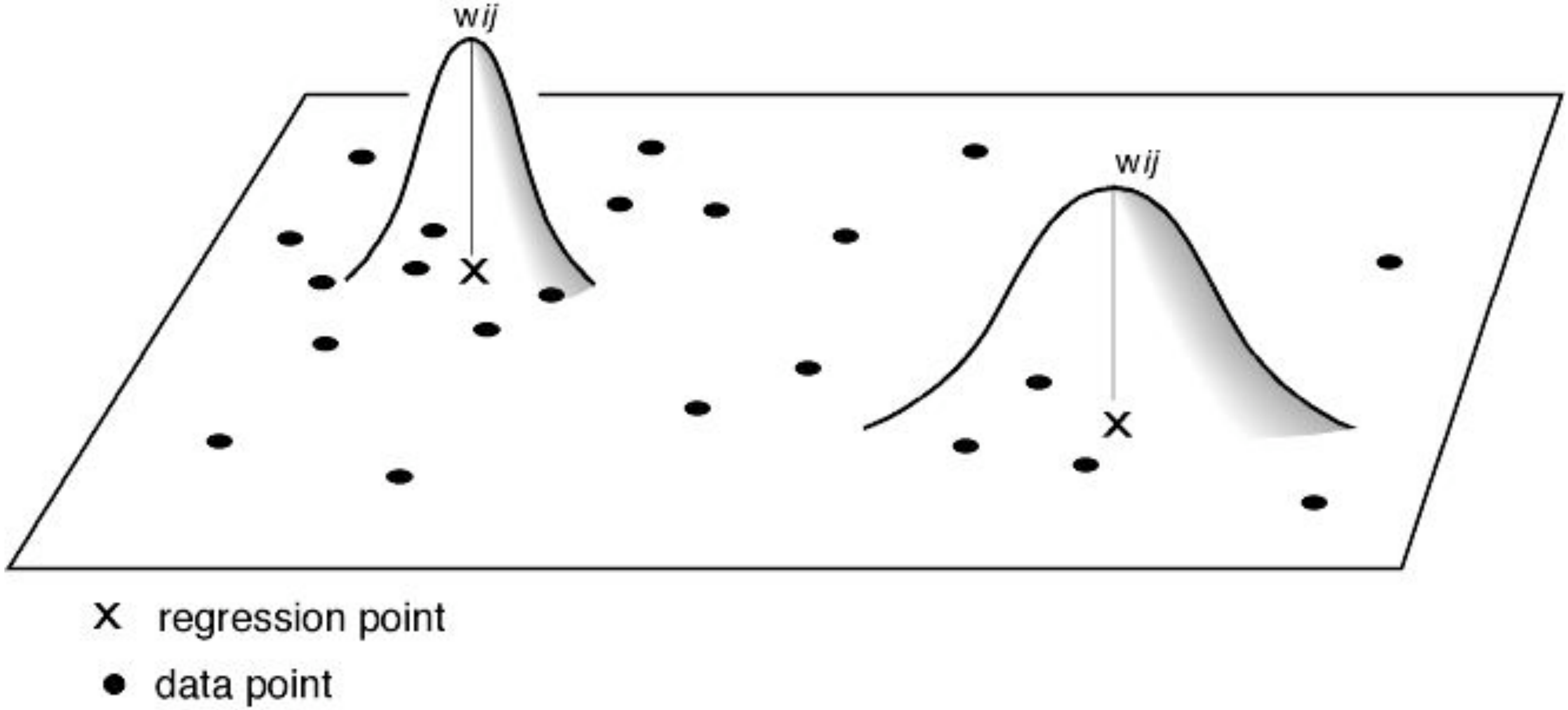
or

$W_i\,(u,v) = [1-(d_i(u,v)^2 \,/\, h^2)]^2$

if j is one of the Nth nearest neighbours of i

$= 0$ otherwise

In the latter, we estimate an optimal value of N in the GWR routine


wij
wij
x
x
x regression point
data point

Calibration

- The results of GWR appear to be relatively insensitive to the choice of weighting function **as long as it is a continuous distance-based function**

- Whichever weighting function is used, the results will, however, be sensitive to the degree of distance-decay.

- Therefore an optimal value of either h or N has to be obtained. This can be found by minimising a crossvalidation score or the Akaike Information Criterion

where…

$$\mathrm{CV} = \sum_i \left(y_i - \hat{y}_{-i}(h)\right)^2$$

Where $\hat{y}_{-i}$ is the fitted value of y_i with data from point i omitted from the calibration

$$\mathrm{AIC} = 2n\log(\hat{\sigma}) + n\log(2\pi) + \frac{n(n + \mathrm{Tr}(\mathbf{S}))}{n - 2 + \mathrm{Tr}(\mathbf{S})}$$

where n is the number of data points, $\hat{\sigma}$ is the estimated standard deviation of the error term, and Tr(**S**) is the trace of the hat matrix.

$$\hat{\mathbf{y}} = \mathbf{S}\mathbf{y}$$

GWR Jargon

- **Data points**
  - locations at which your data are measured
- **Regression points**
  - locations at which you require parameter estimates

These need **not** be the same locations

This can be handy if you want to map the results from very large data sets

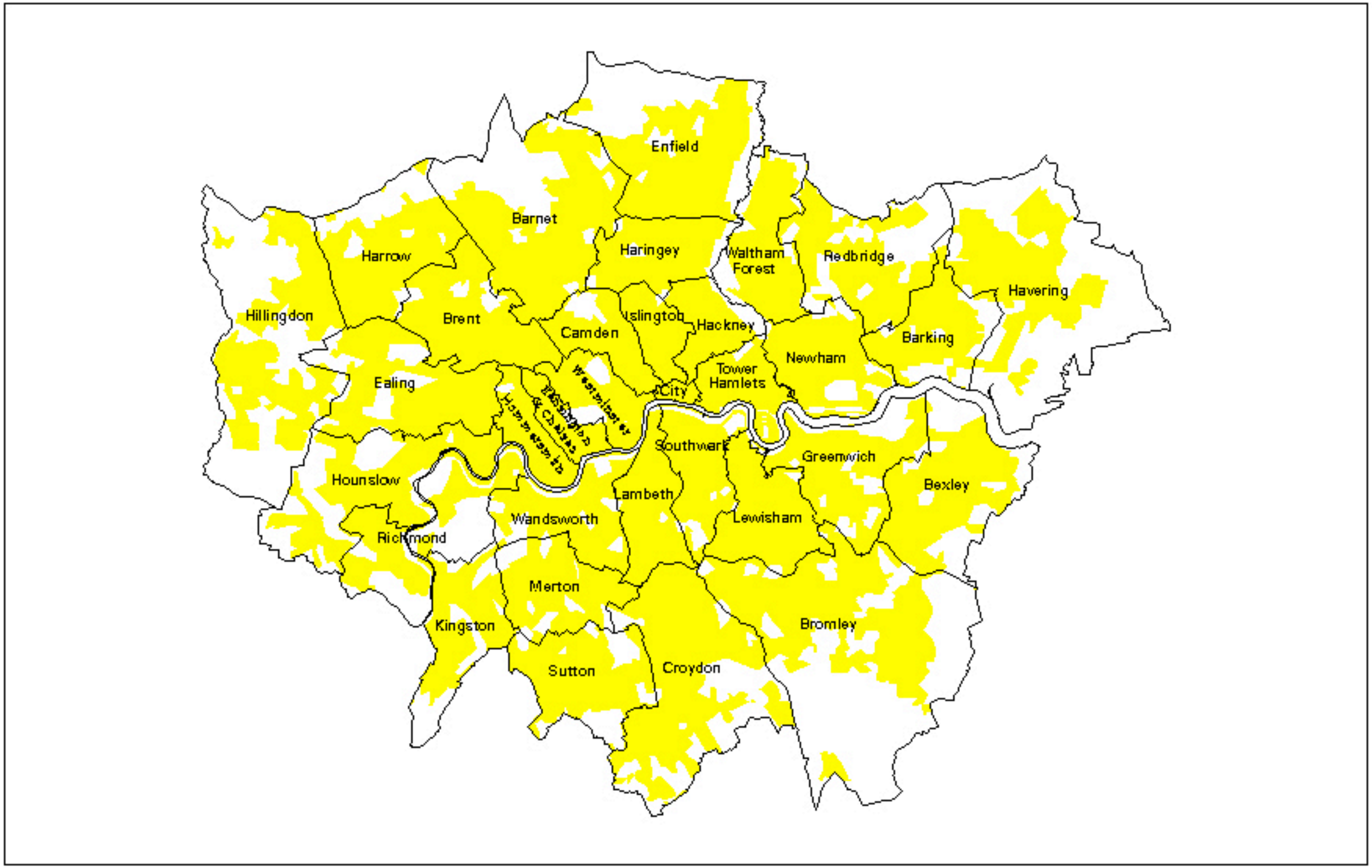
In GWR, we can also …

- estimate local standard errors
- calculate local goodness-of-fit measures
- calculate local leverage measures
- perform tests to assess the significance of the spatial variation in the local parameter estimates
- perform tests to determine if the local model performs better than the global one, accounting for differences in degrees of freedom

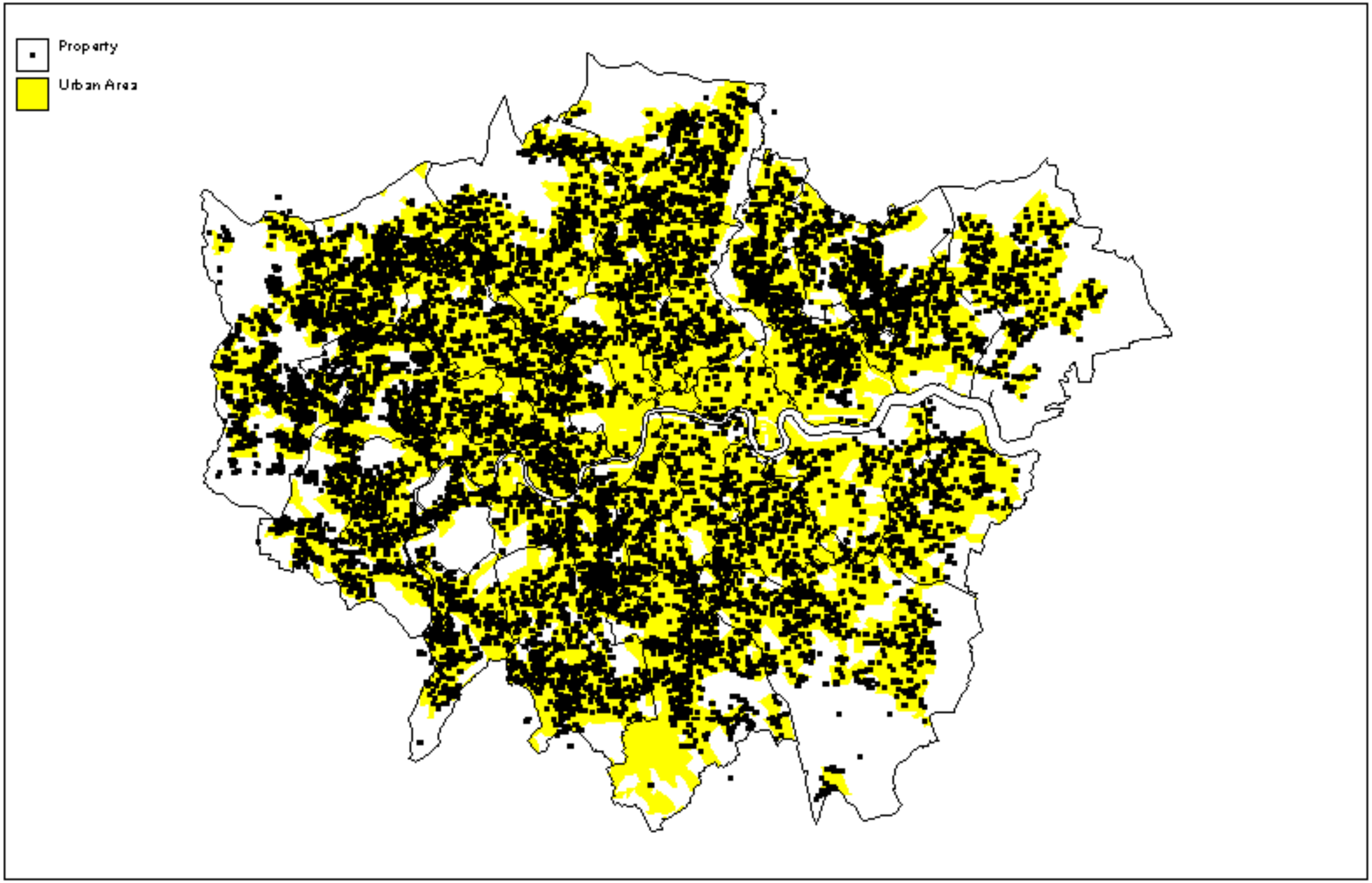
An empirical example - house prices in London

- 1990 sales price data for 12,493 houses in London *(excludes houses sold below market value)*
- along with various attributes of each property and a *postcode* so locations down to 100m can be obtained via the Central Postcode Directory
- neighbourhood data obtained for enumeration districts *(via postcode-to-ED LUT)*

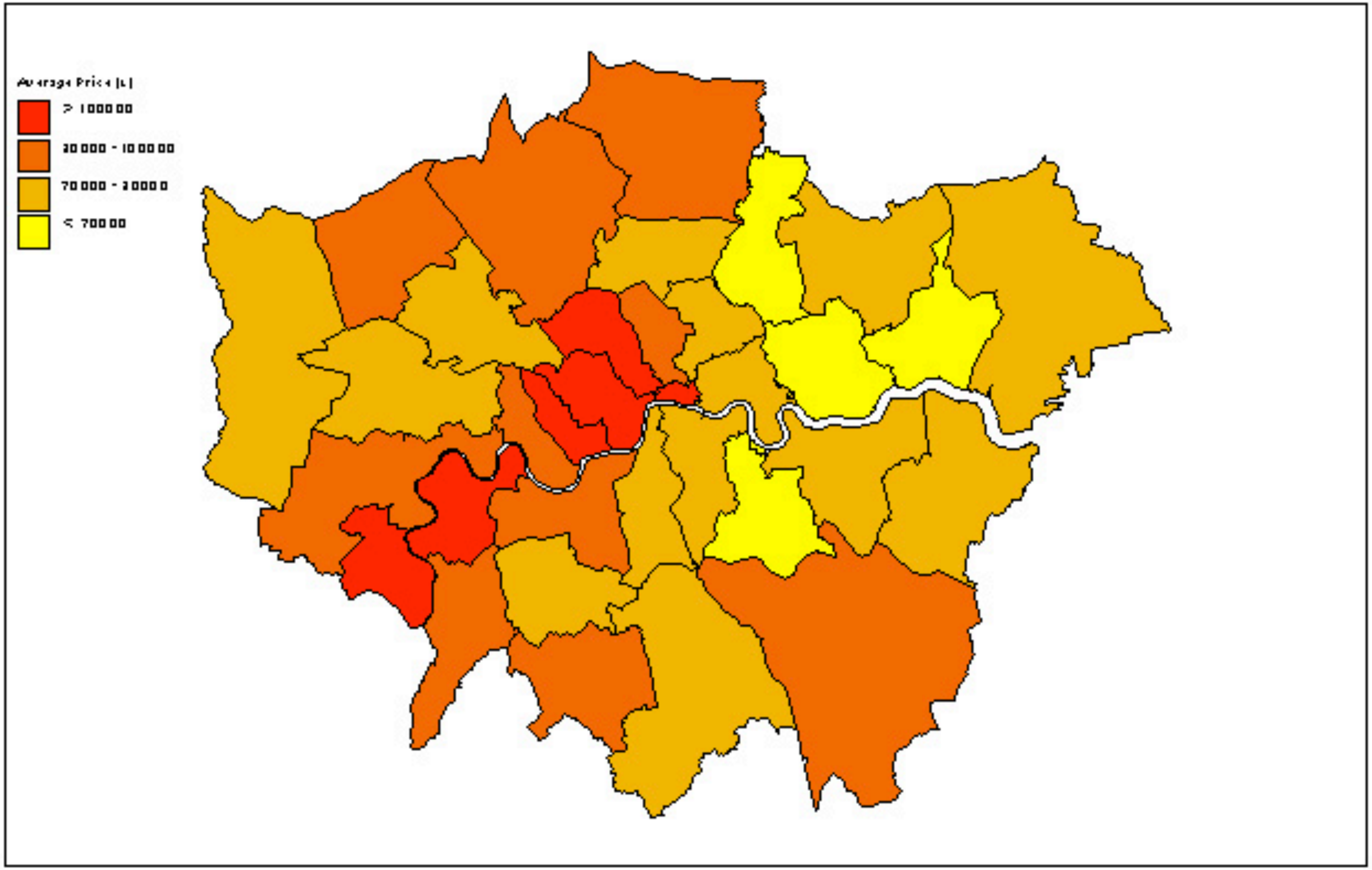
London Boroughs and Urban Area



Locations of house sales in data set



Average House Prices by District



Average House Prices

1990 and 2001

Hedonic Price Modelling

Basic premise:

$$P_i = f\,[\mathbf{S}(i), \mathbf{N}(i)]$$

Lancaster (1966*) J. Political Economy*

Overviews: (very popular technique)

Meen and Andrew (1998) *Modelling Regional House Prices: A Review of the Literature DETR*

Orford (1999) *Valuing the Built Environment: GIS and House Price Analysis Ashgate: Aldershot.*

Issues:

- Almost all applications are *global,* implying no coefficient variation over space whereas several authors have argued that the assumption of uniform price coefficients is unrealistic even within a single metropolitan area.

Global Regression Parameter Estimates

Variable	Parameter Estimate	T value
Intercept	58,900	23.3
FLRAREA	697	49.3
FLRDETACH*	205	7.5
FLRFLAT*	-123	-5.6
FLRBNGLW*	-87	-1.4
FLRTRRCD*	-119	-6.2
BLDPWW1**	-2,340	-3.9
BLDPOSTW**	-2,786	-3.1
BLD60S**	-5,177	-5.0
BLD70S**	-2,421	-2.1
BLD80S**	6,315	6.9
GARAGE	5,956	10.6
CENHEAT	7,777	12.4
BATH2+	22,297	19.1
PROF	72	3.0
UNEMPLOY	-211	-5.5
ln(DISTCL)	-18,137	-30.1

$R^2 = 0.60$

* Excluded house type is Semi-detached
** Excluded age is Inter-war 1914-1939

Price / Square Metre of Various House Types Estimated from the Global Regression Results

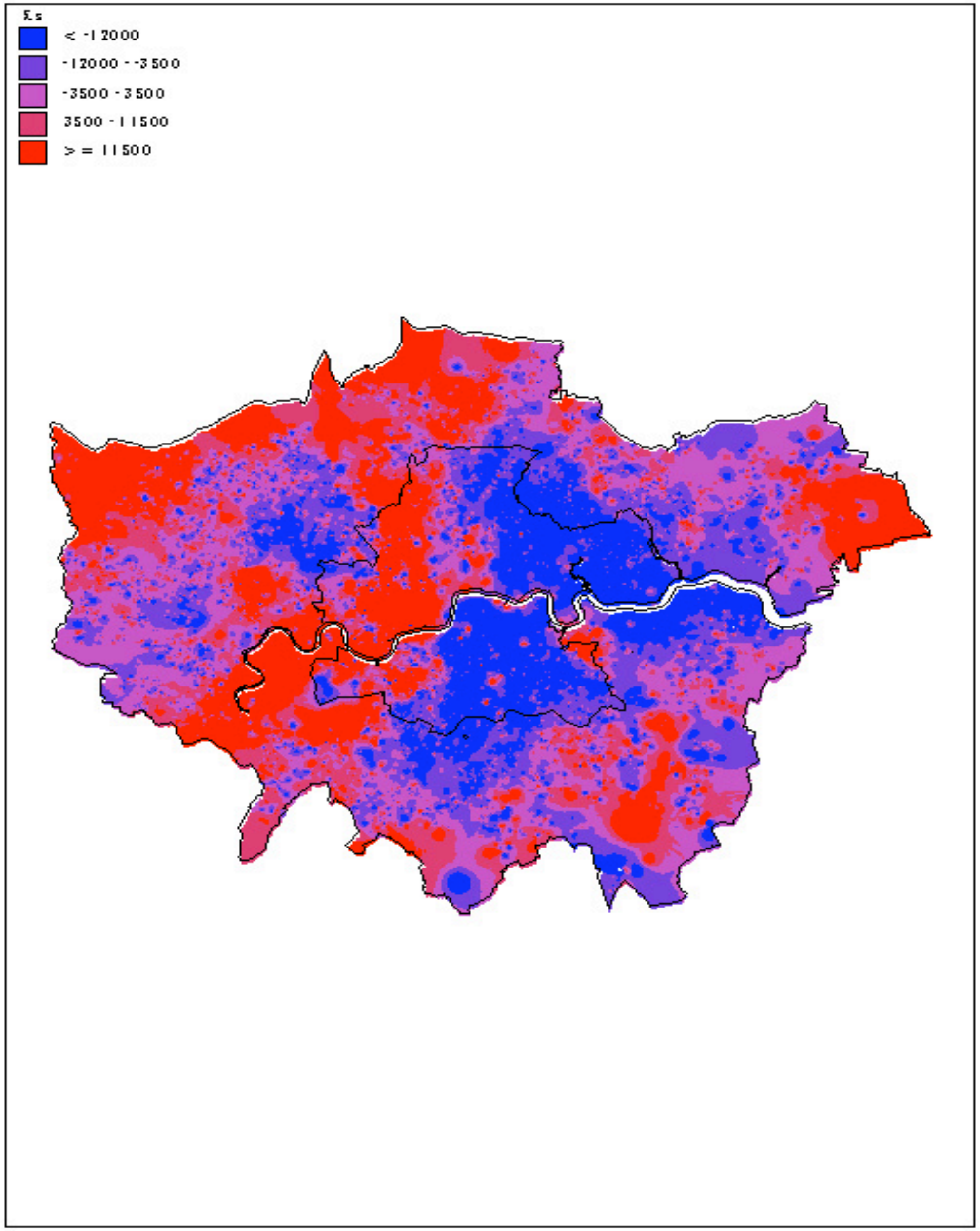
House Type	Price / Sq. M. (£)
Detached	902
Semi-Detached	697
Bungalow	610
Terraced	578
Flat	574

Price Comparisons of equivalent houses by age built

Period of Housing	Pre-1914	1914-1939	1940-1959	1960-1969	1970-1979	1980-1989
Pre-1914	-	-2,340	446	2,837	81	-8,655
1914-1939	2,340	-	2,786	5,177	2,421	-6,315
1940-1959	-446	-2,786	-	2,391	-365	-9.101
1960-1969	-2,837	-5,177	-2,391	-	-2,756	-11,492
1970-1979	-81	-2,421	365	2,756	-	-8,736
1980-1989	8,655	6,315	9,101	11,492	8,736	-

Residuals from Global Model

Figure 3.4 Residuals from the global model



An Alternative

- Calibrate separate hedonic price models for each of the London boroughs
- Map results or present in table form
- Example of the value of flatted properties and terraced properties

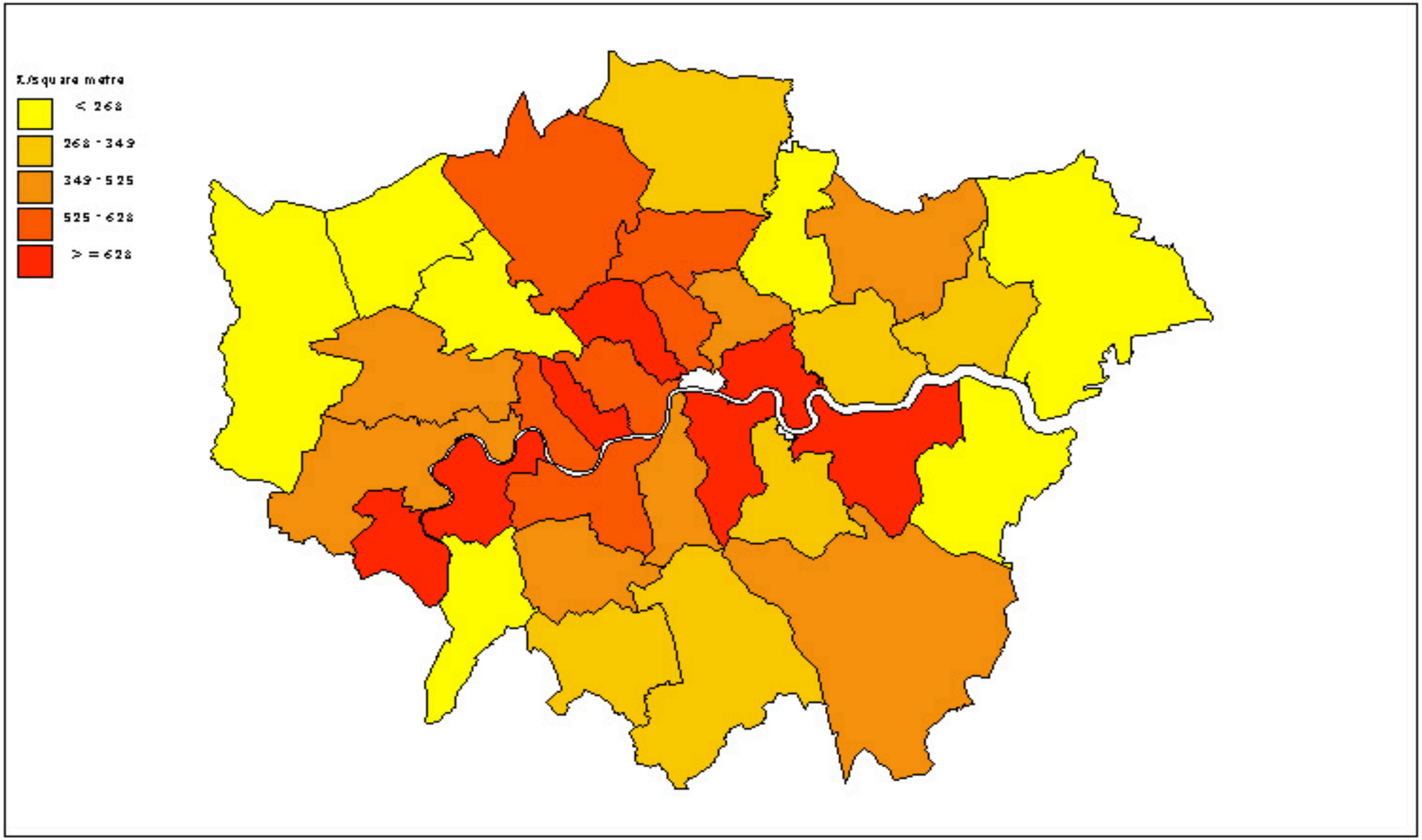
Table 2.5

Price /m² of Flats and Terraced Housing in each London Borough from Separate Calibrations of the Global Hedonic Model

Borough	Price/m^2 (£) Flat	Terraced	Ratio Terrace/Flat	R^2
Barking	310	609	1.96	.70
Barnet	528	579	1.10	.75
Bexley	106	80	0.75	.86
Brent	263	310	1.18	.73
Bromley	399	427	1.07	.83
Camden	897	179	0.20	.69
City	***	***	***	***
Croydon	329	216	0.66	.83
Ealing	464	350	0.75	.63
Enfield	326	615	1.89	.85
Greenwich	629	611	0.98	.53
Hackney	432	612	1.42	.71
Hammersmith	524	1272	2.43	.82
Haringey	543	623	1.15	.73
Harrow	233	444	1.91	.47
Havering	104	555	5.34	.67
Hillingdon	265	270	1.02	.71
Hounslow	513	733	1.43	.65
Islington	595	889	1.49	.80
Kensington	1574	2019	1.28	.75
Kingston	141	605	4.29	.81
Lambeth	350	606	1.73	.72
Lewisham	268	513	1.91	.76
Merton	517	554	1.07	.64
Newham	267	249	0.93	.56
Redbridge	420	518	1.23	.77
Richmond	866	713	0.82	.75
Southwark	667	498	0.75	.72
Sutton	311	572	1.84	.82
Tower Hamlets	628	381	0.61	.79
Waltham Forest	257	320	1.25	.80
Wandsworth	563	780	1.39	.68
Westminster	626	1672	2.67	.64

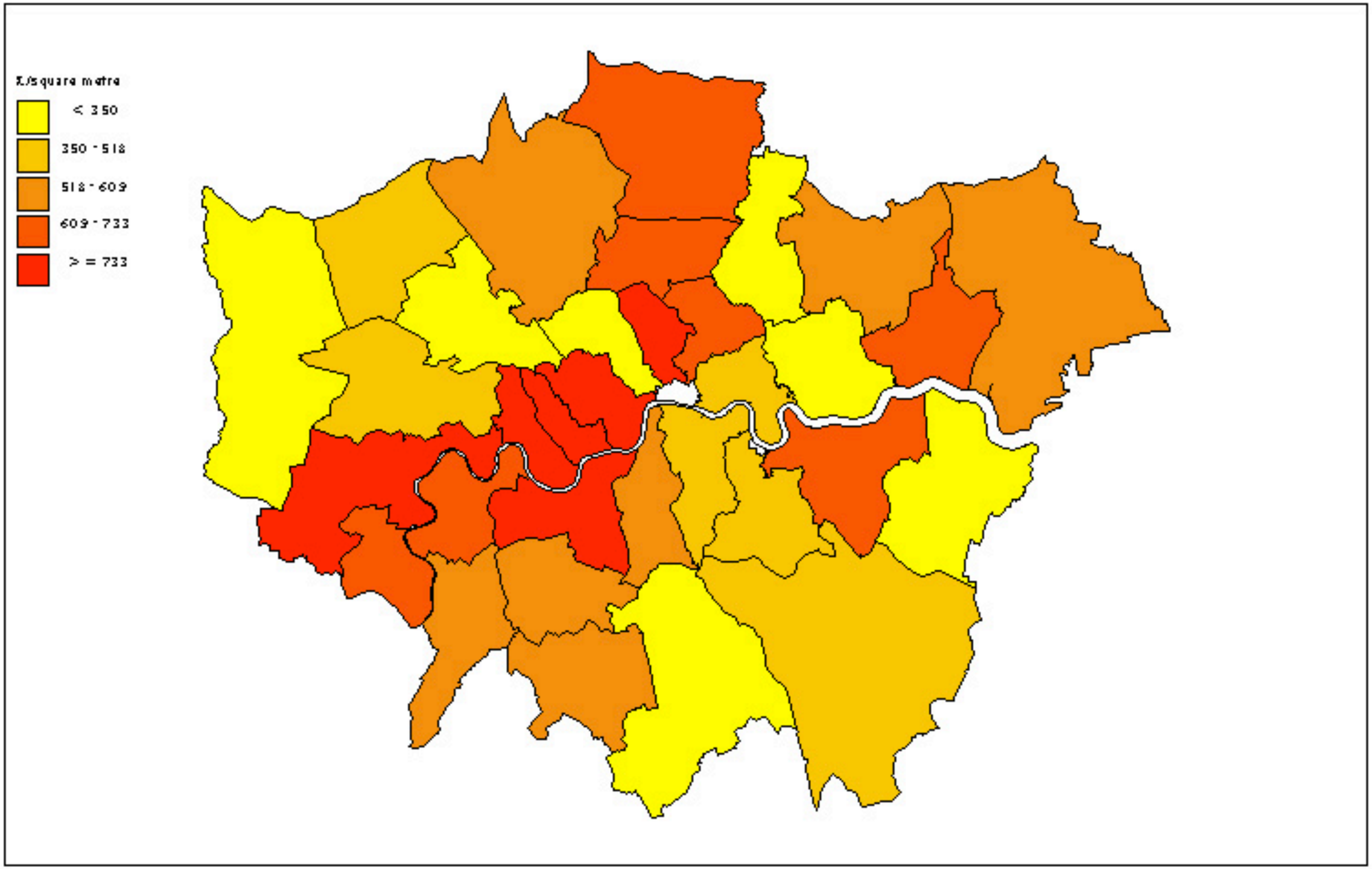
Value of Flatted Property £/m^2

(Figure 2.5) Flat+Floorspace Parameter



Value of Terraced property £/m²

(Figure 2.6) Terrace+Floorspace Parameter



Problems with this approach

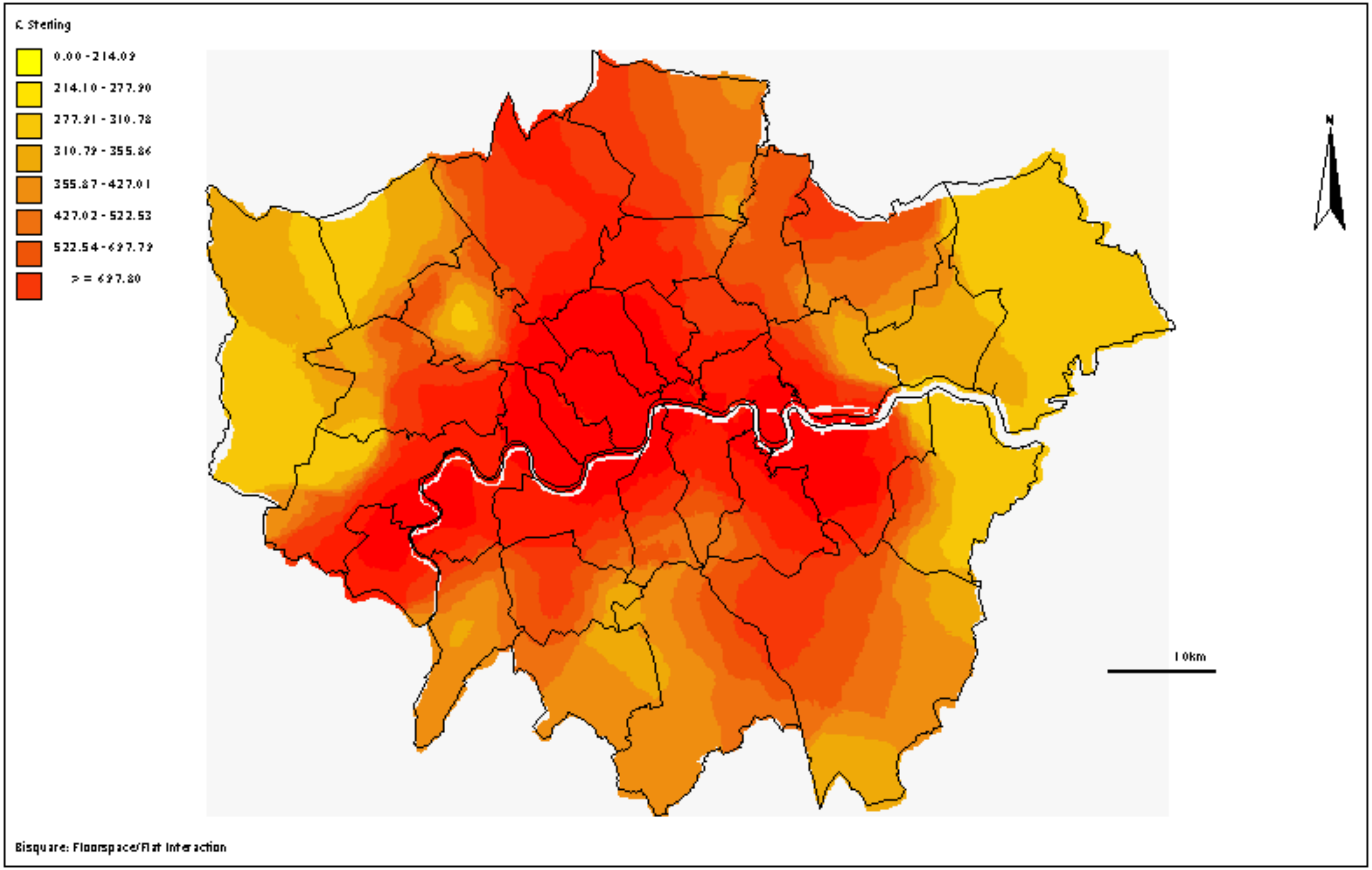
- There is a statistical issue in that some areas do not have sufficient data to support independent calibrations
- It is similar to a global model in that the processes being examined are assumed to be stationary across each borough *(yet are assumed to vary between boroughs!)*
- The process is assumed to be discrete and discontinuities coincide exactly with the boundaries of the boroughs. However, most spatial processes are continuous and unrelated to the location of administrative boundaries

Better to use GWR

- Models a continuous change in local parameter estimates
- In this case an adaptive kernel is used - a bisquare function
- Calibration yielded an optimal number of nearest neighbours = 931
- Results presented in a series of parameter surfaces - those shown all have significant spatial variation

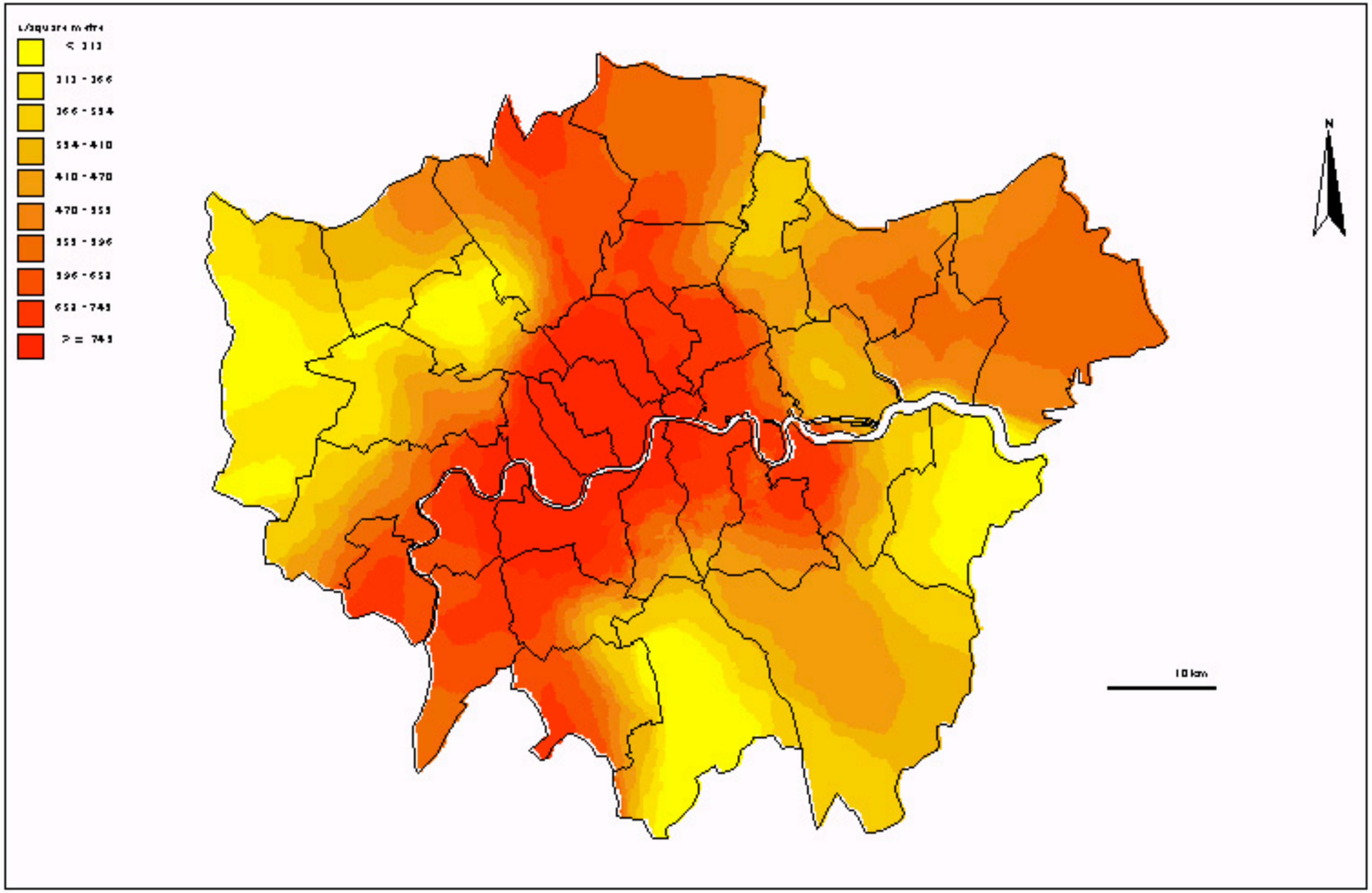
Value of flatted property £/m^2

(global estimate = 574)

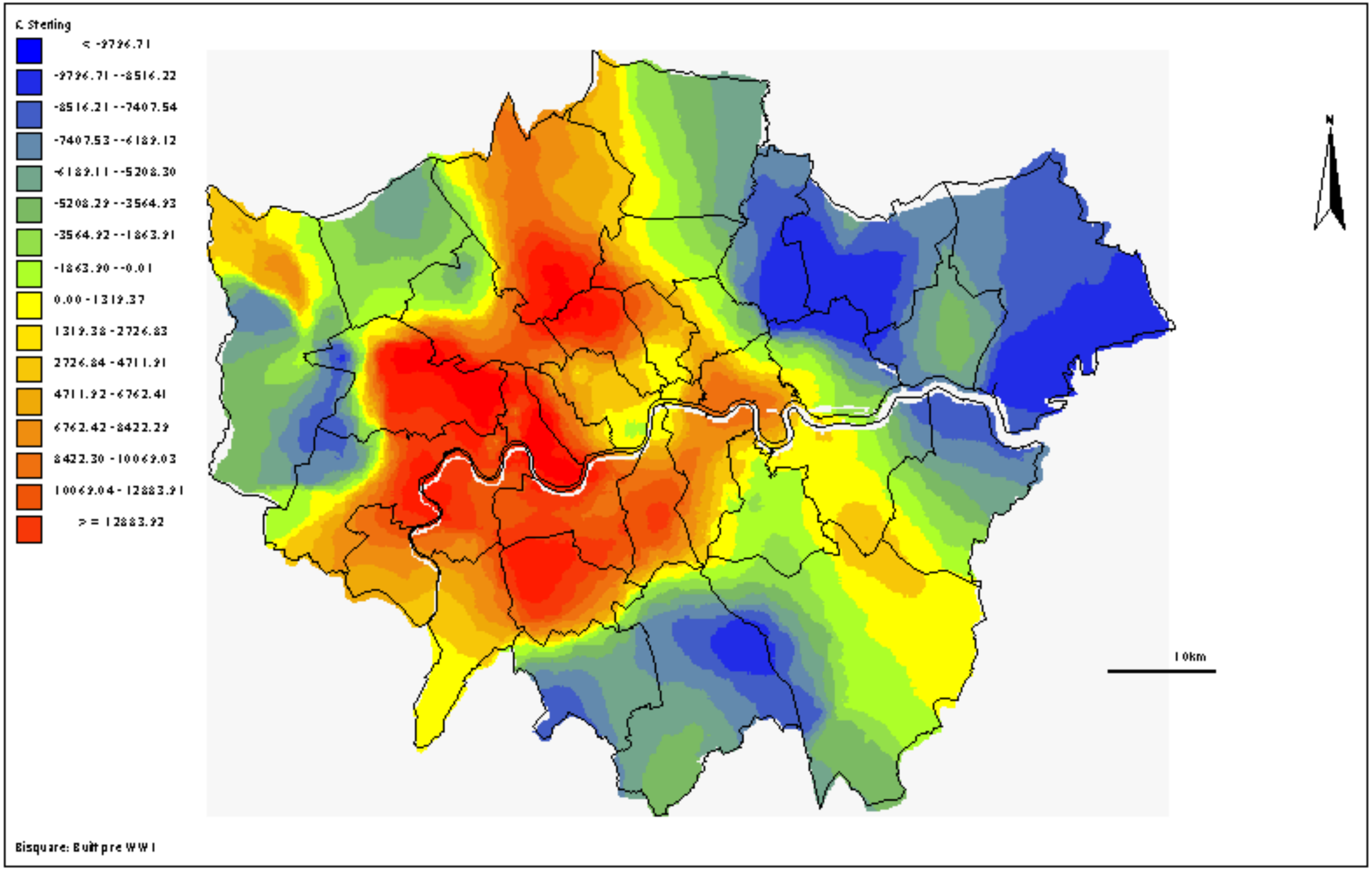


Value of terraced property £/m²
(global estimate = £578)

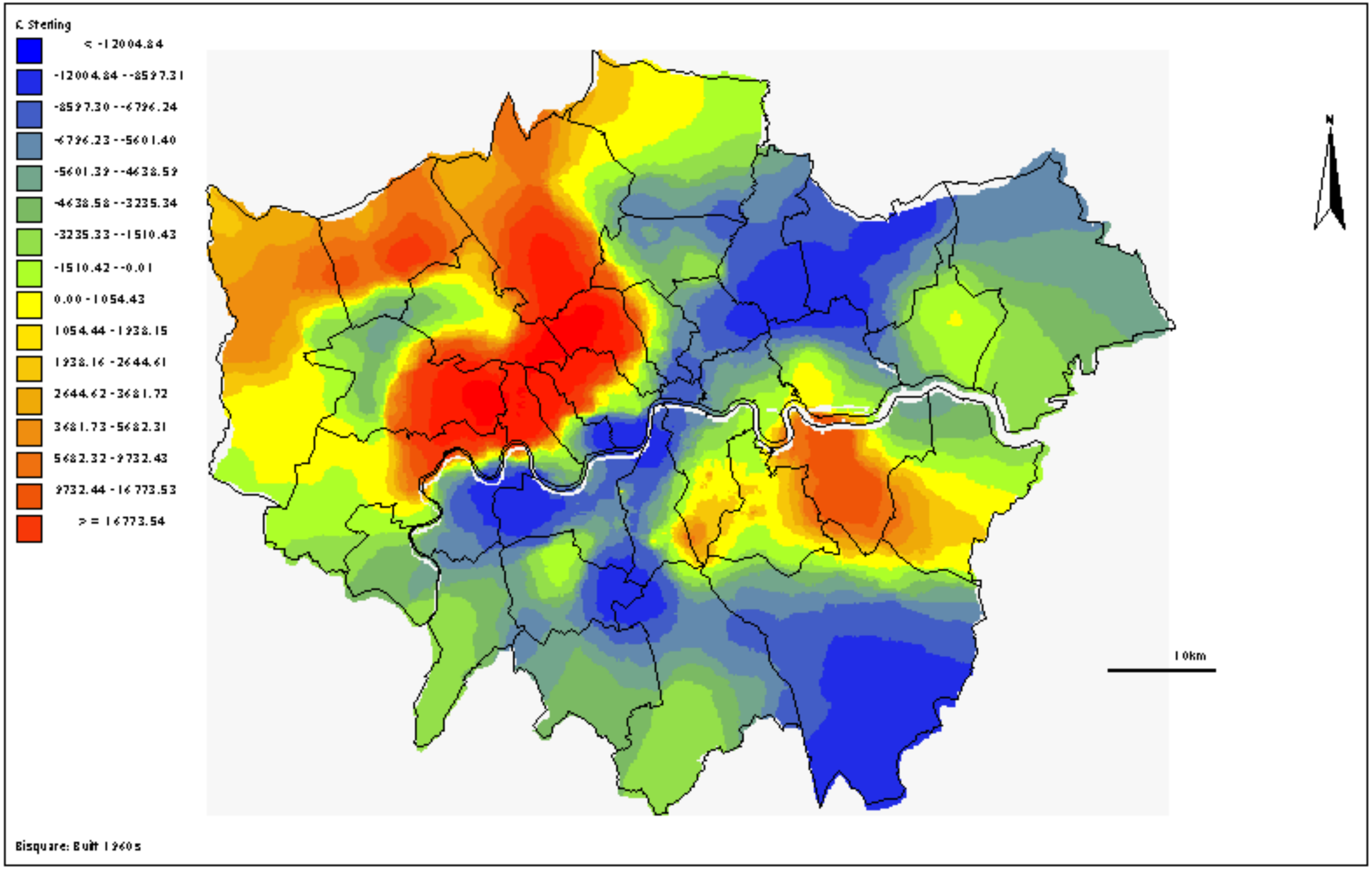
fig2_15.eps Bisquare: Floorspace/Terraced Interaction



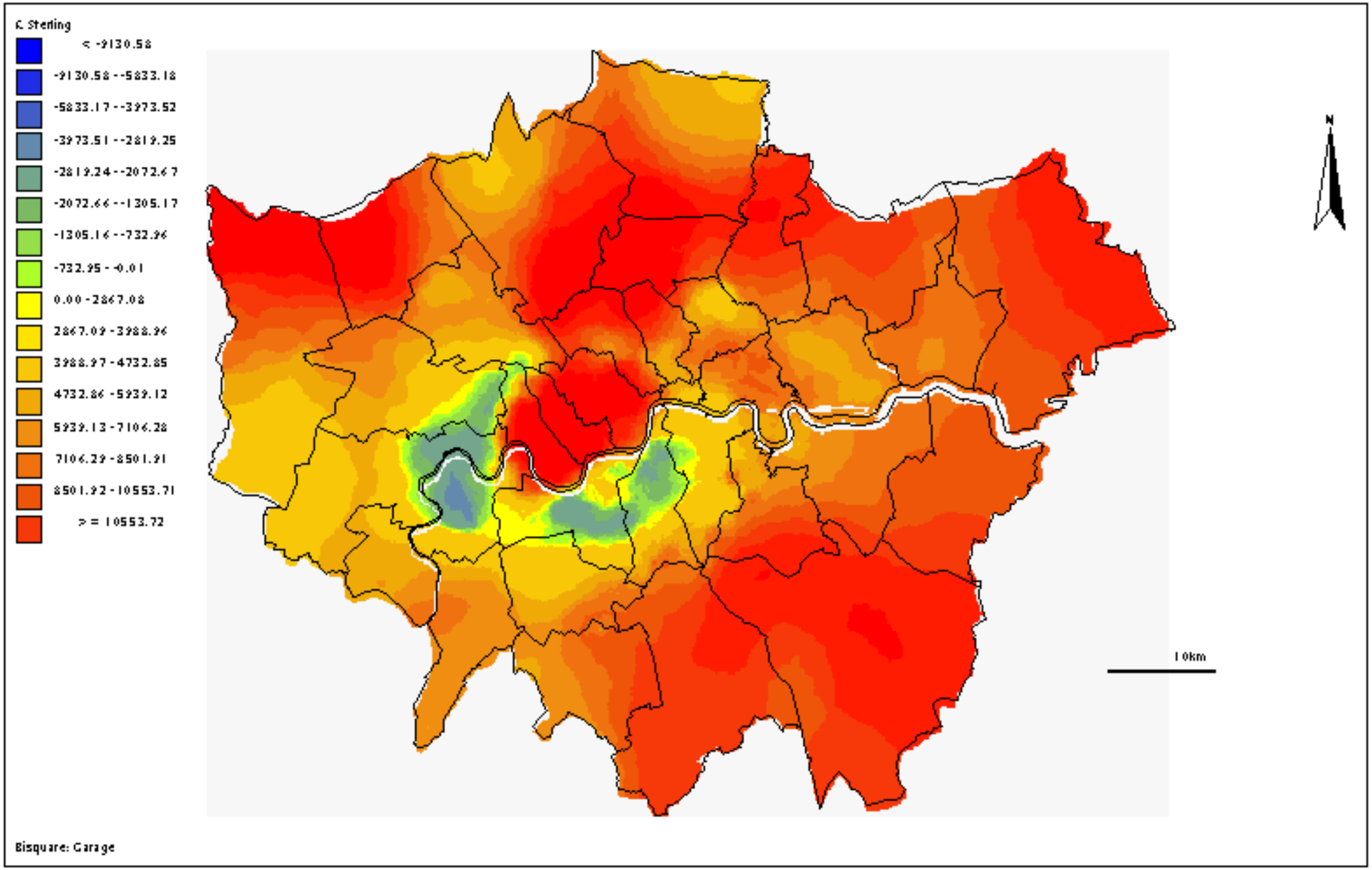
Pre-1914 housing compared to inter-war
(global estimate = £-2,340)



1960s housing compared to inter-war
(global estimate = £-5,177)

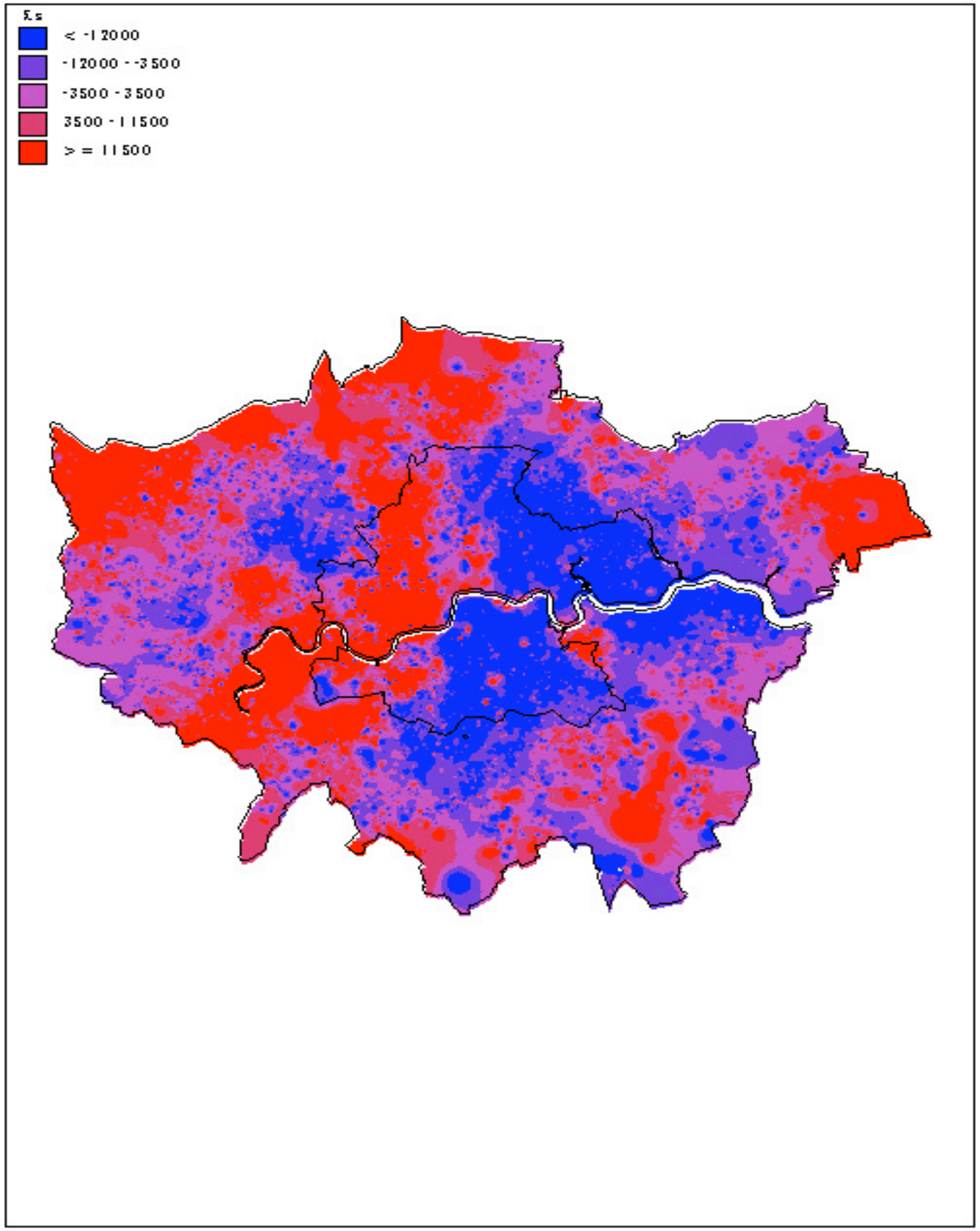


Value of a garage
(global estimate = £5,956)



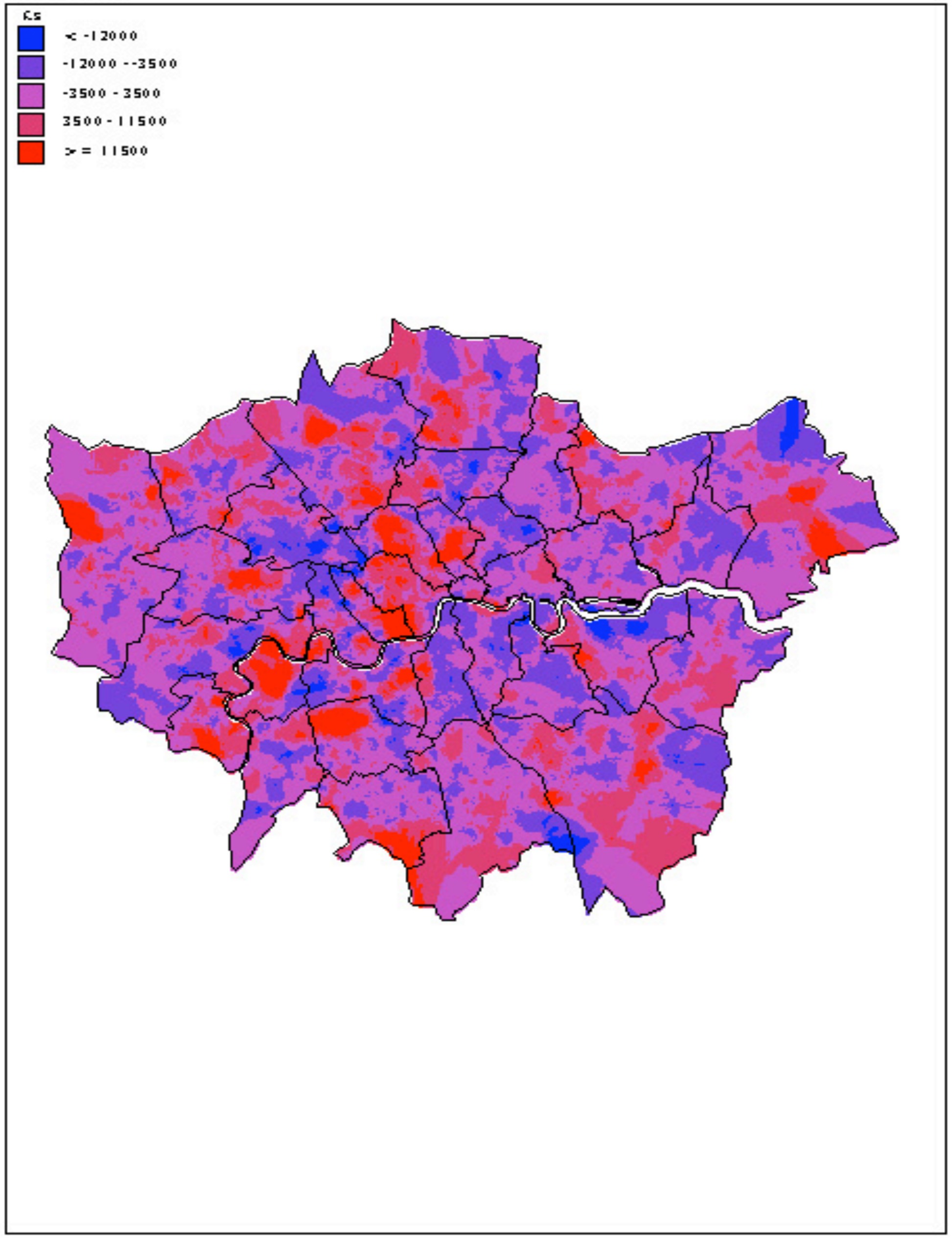
Residuals from Global Model

Figure 3.4 Residuals from the global model



Residuals from GWR Model

Figure 3.19 Residuals from the GWR model



Summary of Lecture

- GWR is a useful method to investigate spatial non-stationarity - simply assuming relationships are stationary over space is no longer tenable
- GWR is a genuine spatial statistical technique that is GIS friendly in that it is designed to take advantage of locational information as well as attribute information
- GWR can be likened to a ‘spatial microscope’ - allows us to see patterns in relationships that were previous unobservable
- Can use GWR either to aid model development or identify interesting areas for further investigation.