UBDC Glasgow 2016

# Local Modelling: An Introduction

**A. Stewart Fotheringham**

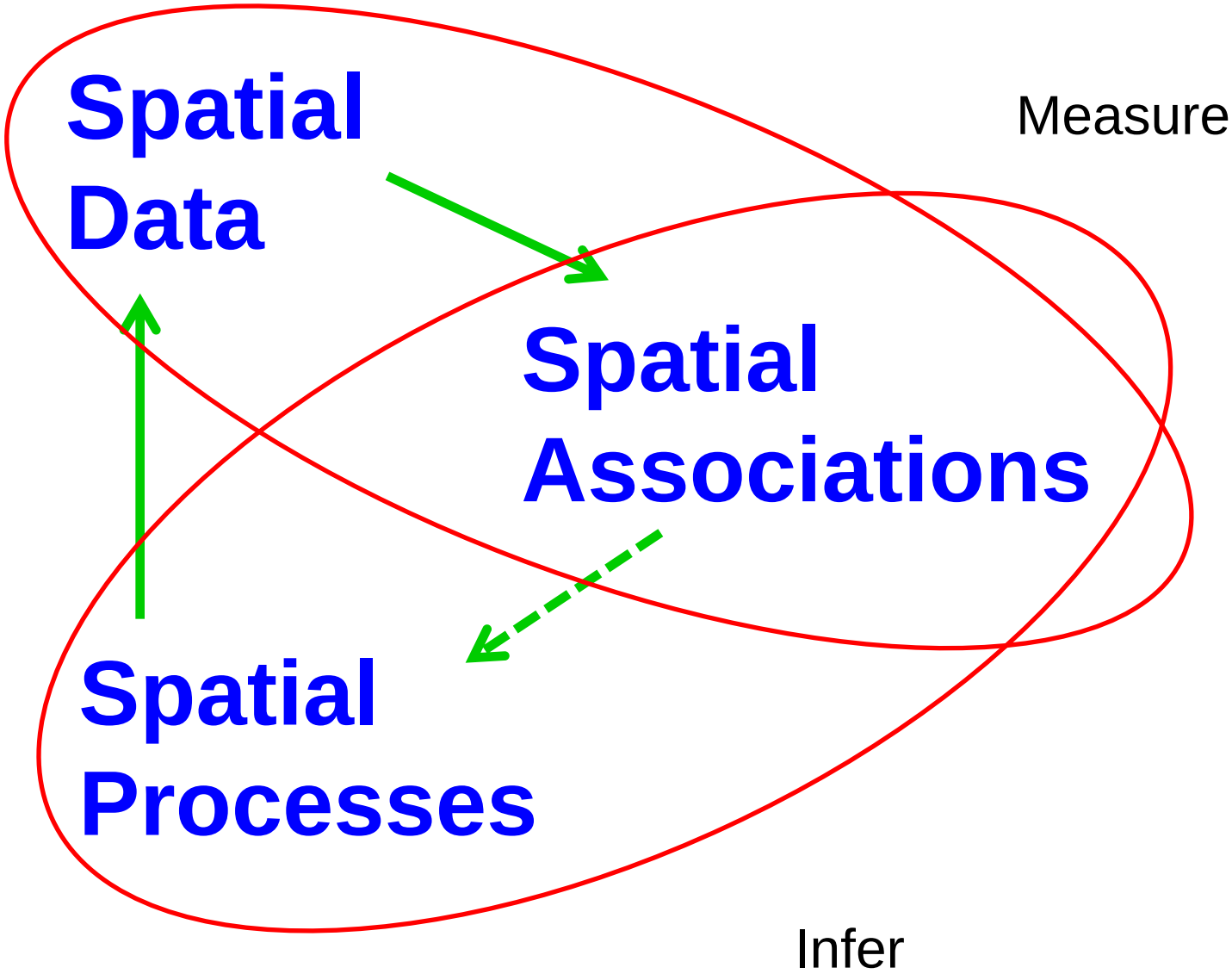
Professor of Computational Spatial Science
School of Geographical Sciences and Urban Planning
Arizona State University
stewart.fotheringham@asu.edu

## There are three main reasons why we employ models:

1. To predict – the future state of things; data we can't observe; or what would happen if something changes in a predictable way

2. To act as a benchmark by which to judge reality e.g. optimality in transportation systems

3. To obtain information on the processes that produce the data we observe about the world.


Spatial Data
Measure
Spatial Associations
Spatial Processes
Infer

The most common way of uncovering information on spatial associations is to calibrate a spatial model and obtain estimates of the model's **parameters**.

The more parameters we can **reliably** estimate, the more information on spatial associations we can generate to help us make inferences about spatial processes.

In a typical model (linear or non-linear) applied to spatial data we assume that the processes that generate the data we observe are stationary over space

- *That is, the same stimulus provokes the same response in all parts of the study region*:

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

# But suppose the processes aren't the same everywhere…

**For example…**

Suppose we regressed the proportion of people with cardio-vascular disease on a variety of socio-economic attributes for data zones across Scotland

Would we really expect these relationships to be the same for all data zones? Would we really expect these relationships to be the same across Glasgow?

If we calibrated a global model using data from across Scotland would we miss important differences?

**For example…**

Suppose we regressed the extra revenue raised on sales of surfing equipment on the amount spent on advertizing surfing equipment across cities in the US.

Would we really expect this relationship to be the same for **St Louis** and **Honolulu**?

**For example…**

Suppose we regressed the likelihood of voting Republican on income across the US

Would we really expect this relationship to be the same for **College Towns** and **Non-College Towns**?

Equally, if we calibrated a spatial lag model

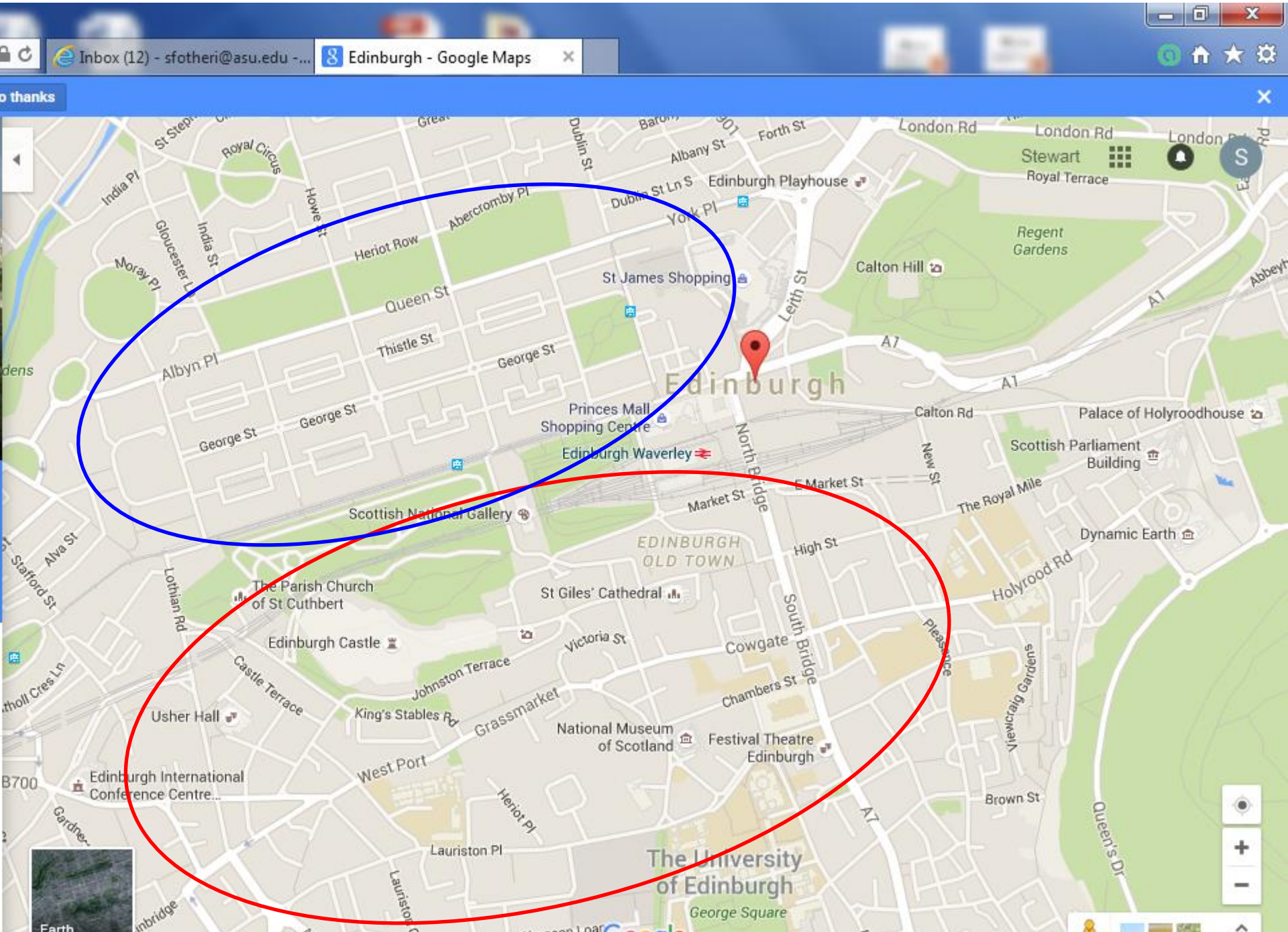
$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots \beta_n x_{ni} + \theta\, ylag_i$$

Given θ is a measure of spatial autocorrelation, why would we expect θ to be constant over space?

A variety of statistics have been developed to measure spatial variations in autocorrelation (e.g. LISA; Getis-Ord etc)

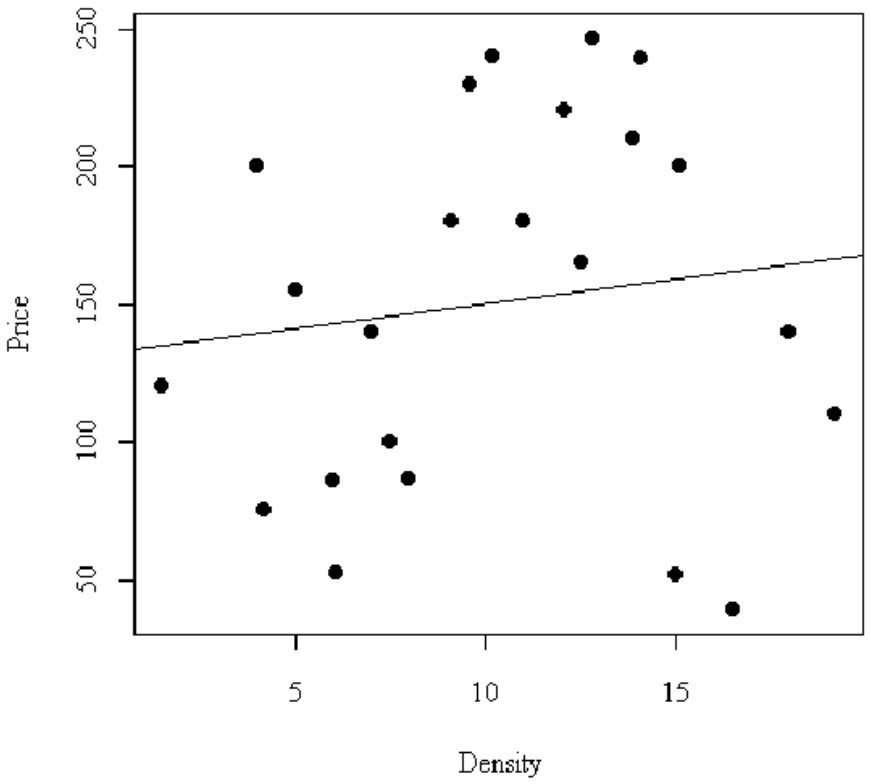
And if regressed actual road distances on distances generated from coordinates (say, Euclidean or Manhattan distances) in a system like this…

$$d_{ij}(p) = [(x_i - x_j)^p + (y_i - y_j)^p]^{1/p}$$

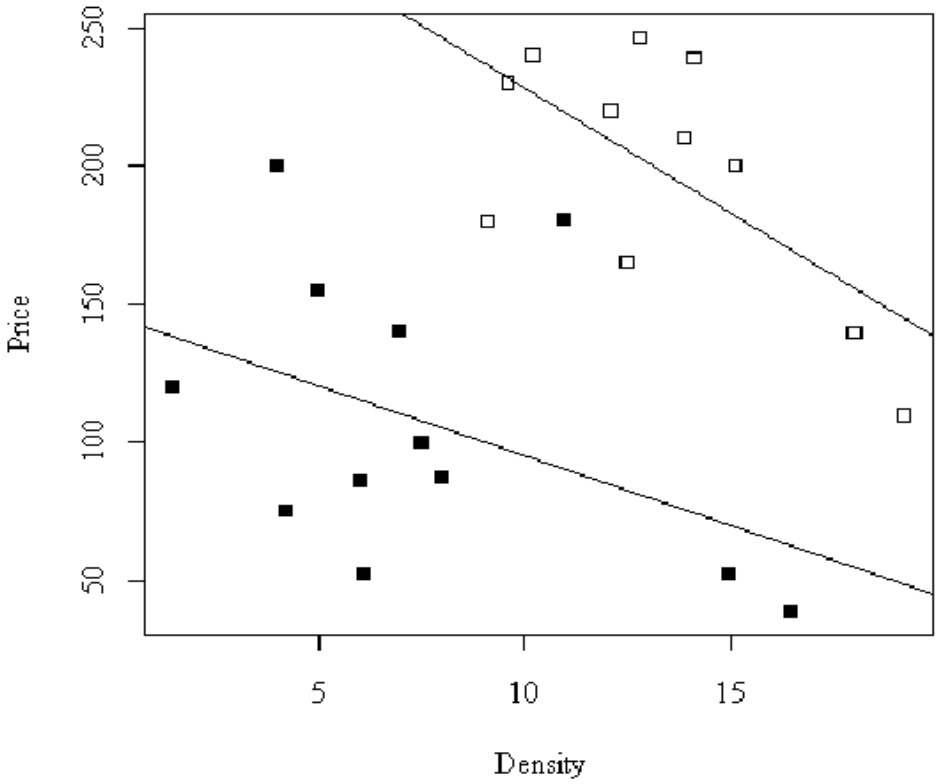

Inbox (12) - sfotheri@asu.edu -...
Edinburgh - Google Maps
o thanks
London Rd
Stewart
Royal Terrace
Edinburgh Playhouse
Regent Gardens
Calton Hill
St James Shopping
Queen St
Albyn Pl
Thistle St
George St
Leith St
A1
Edinburgh
Princes Mall Shopping Centre
Edinburgh Waverley
Calton Rd
Palace of Holyroodhouse
Scottish Parliament Building
North Bridge
E Market St
Market St
New St
The Royal Mile
Scottish National Gallery
EDINBURGH OLD TOWN
High St
Dynamic Earth
Holyrood Rd
The Parish Church of St Cuthbert
St Giles' Cathedral
South Bridge
Edinburgh Castle
Victoria St
Cowgate
Johnston Terrace
Castle Terrace
Chambers St
Usher Hall
King's Stables Rd
Grassmarket
National Museum of Scotland
Festival Theatre Edinburgh
West Port
Edinburgh International Conference Centre...
Heriot Pl
Brown St
Queen's Dr
Lauriston Pl
The University of Edinburgh
George Square
Earth

# Another indicator of the importance of local modelling - Simpson's Paradox

**Spatially aggregated data**




**Spatially disaggregated data**



In such situations, a global model is incorrect and the parameter estimates will represent ‘averages’ of underlying spatially varying relationships.

So we need a more flexible model form that allows the parameters to vary over space

In fact, there are several model forms that allow for **spatial nonstationarity** in processes:

# Models that allow for spatial nonstationary processes fall into two categories:

1. Those which require ‘regions of stability’ to be defined *a priori* e.g.

    i. Hierarchical or Multilevel models
    ii. Spatial regime models

2. Those which do not  e.g.

    1. Geographically Weighted Regression (GWR)
    2. Spatial Filtering Models
    3. Bayesian Spatially Varying Coefficients Models

# Geographically Weighted Regression

**The Problem**: If we believe the processes are examining might be spatially nonstationary and we had many repeated samples at every location, we could run separate regressions for each location. Unfortunately we rarely have such data and usually we have only a single set of observations at each location.

**The Solution:** Borrow data from nearby locations, weighted by the proximity of the location from data are being borrowed to the location for which the local regression is calibrated. Under the assumption of spatial dependency in processes, to minimise bias in the results, we weight data from nearby locations more heavily than from more distant ones.

*This is the essence of GWR*

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

Where the local parameter estimates are obtained with the estimator

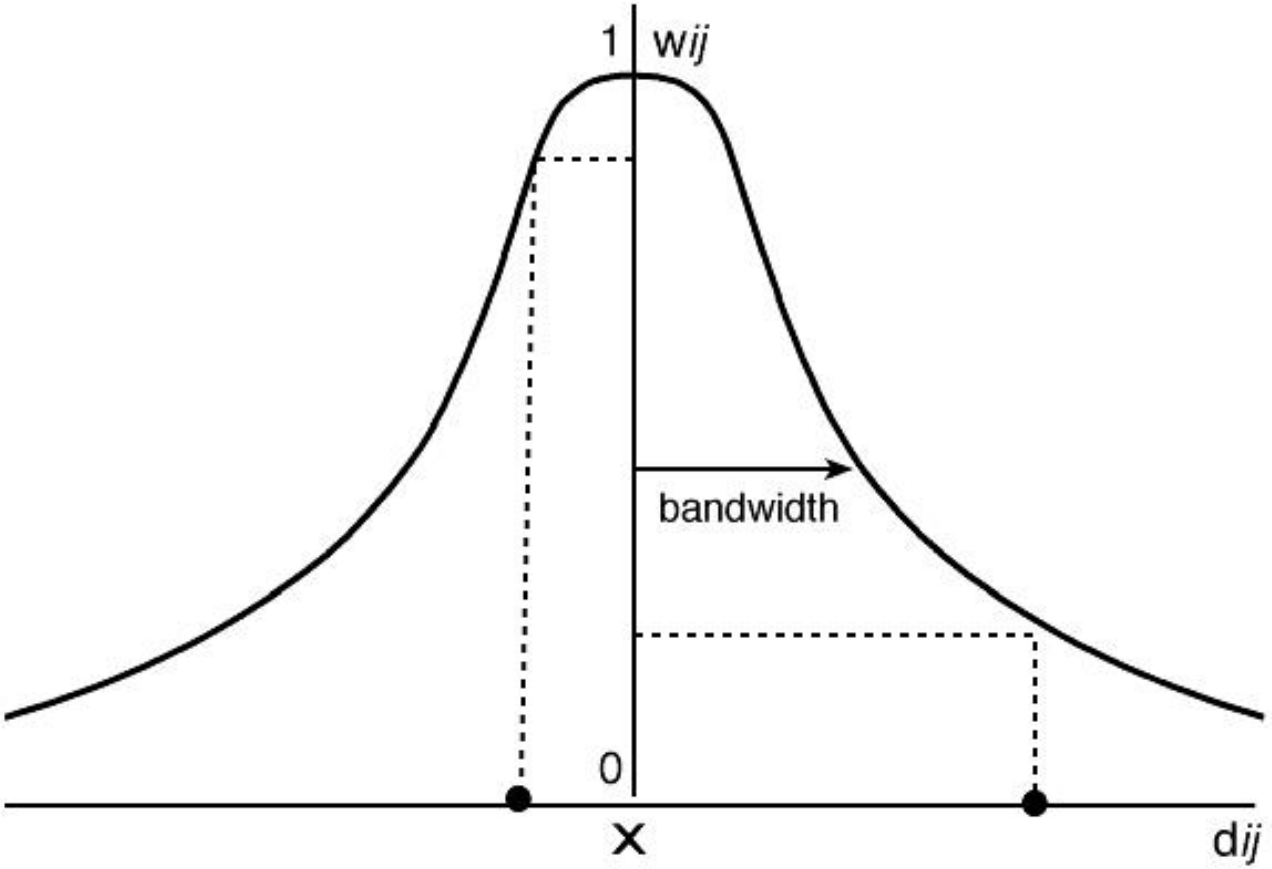
$$\beta'(i) = (X^T W(i) X)^{-1} X^T W(i) Y$$

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

$$\mathbf{W}(i) = \begin{matrix} w_{i1} & 0 & 0 & \cdots & 0 \\ 0 & w_{i2} & 0 & \cdots & 0 \\ 0 & 0 & w_{i3} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & w_{in} \end{matrix}$$

where $w_{in}$ is the weight given to data point n for the estimate of the local parameters at location i.

# A Typical Spatial Weighting Function



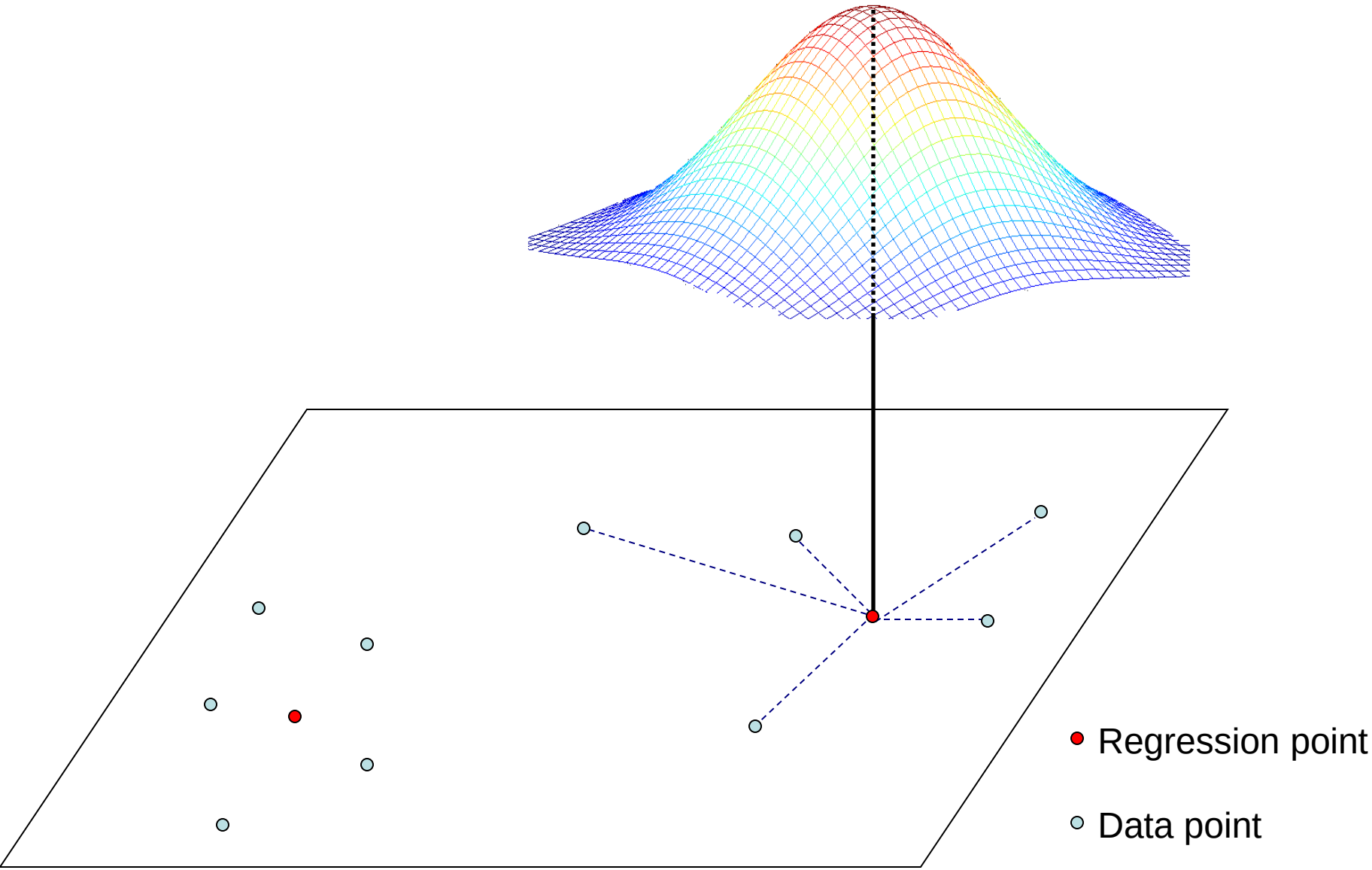

x regression point

● data point

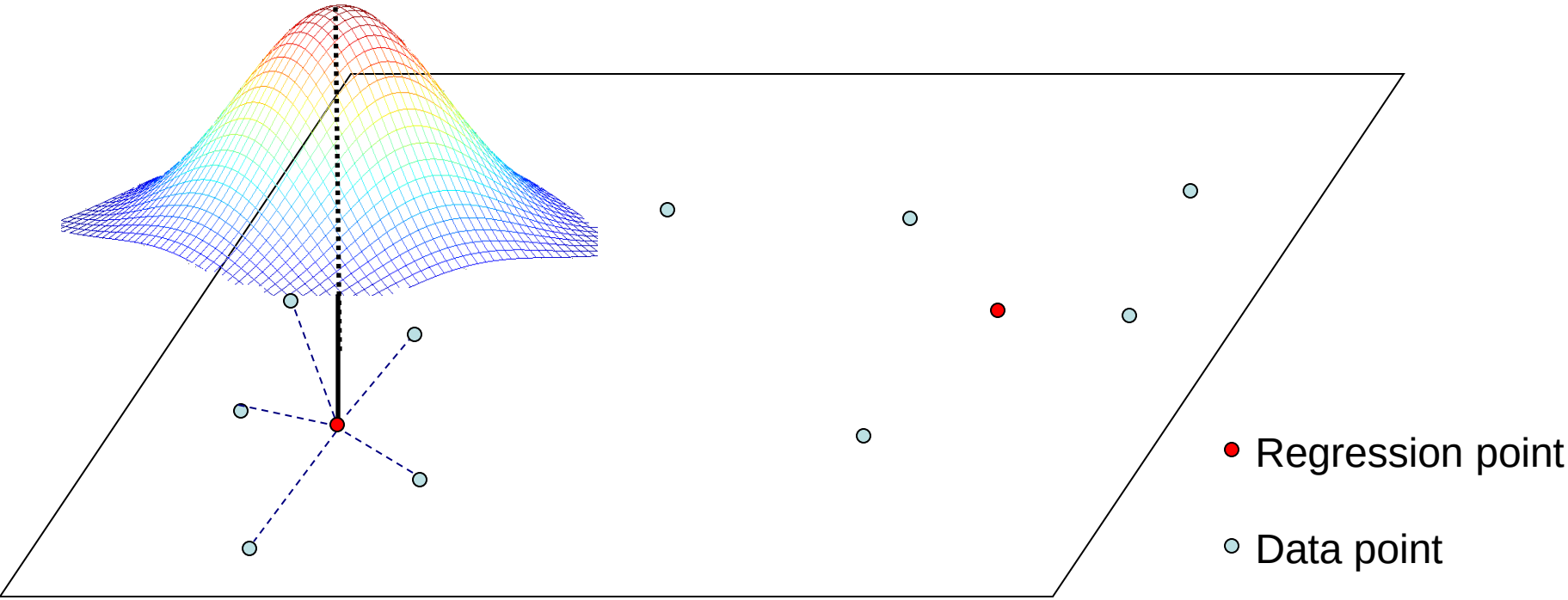
$w_{ij}$ is the weight of data point *j* at regression point *i*

$d_{ij}$ is the distance between regression point *i* and data point *j*

# Geographically Weighted Regression



# Geographically Weighted Regression

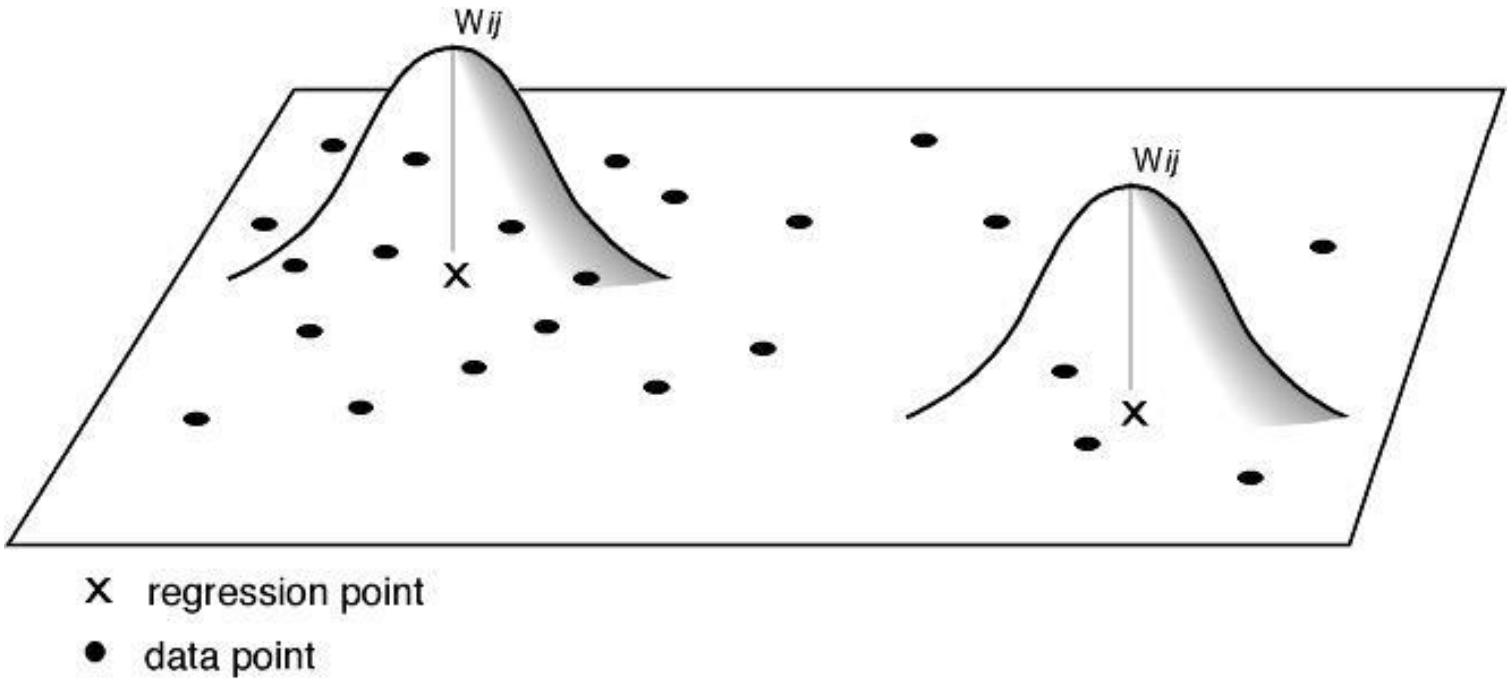


# Weighting schemes

Numerous weighting schemes can be used although they tend to be Gaussian or 'Gaussian-like' reflecting the type of dependency found in most spatial processes.

Weighting schemes can be either **fixed** or **adaptive**.

# Fixed Weighting Scheme



# Example of a Fixed Weighting Scheme

For each location i at which the local regression model is calibrated,

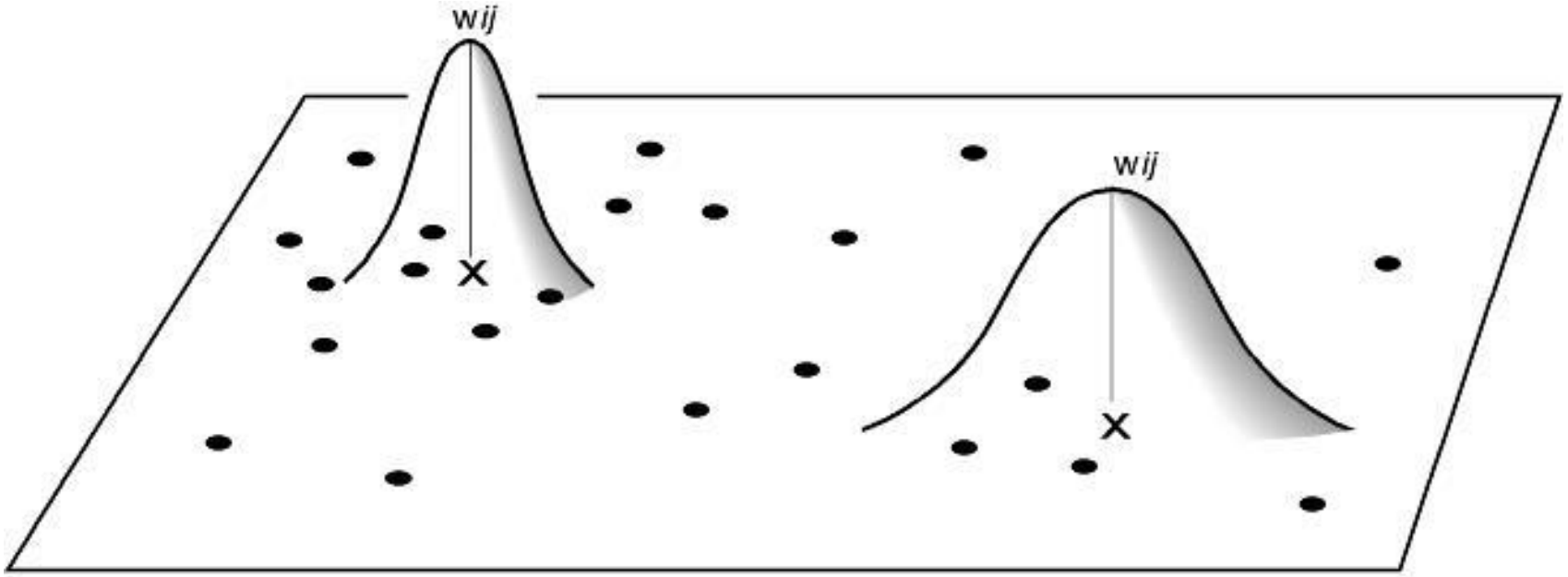
$$w_{ij} = \exp [ - ½ (d_{ij} / h)^2 ]$$

where

$d_{ij}$ is the distance between locations i and j

h is the bandwidth – as h increases, the gradient of the kernel becomes less steep and more data points are included in the local calibration. We need to find the optimal value of h in the GWR routine.

# Spatially Adaptive Weighting Scheme




X regression point

● data point

# Example of a Spatially Adaptive Weighting Scheme

$$w_{ij} = [1-(d_{ij}^2 / h^2)]^2$$ if j is one of the Nth nearest neighbours of i

$$= 0$$ otherwise

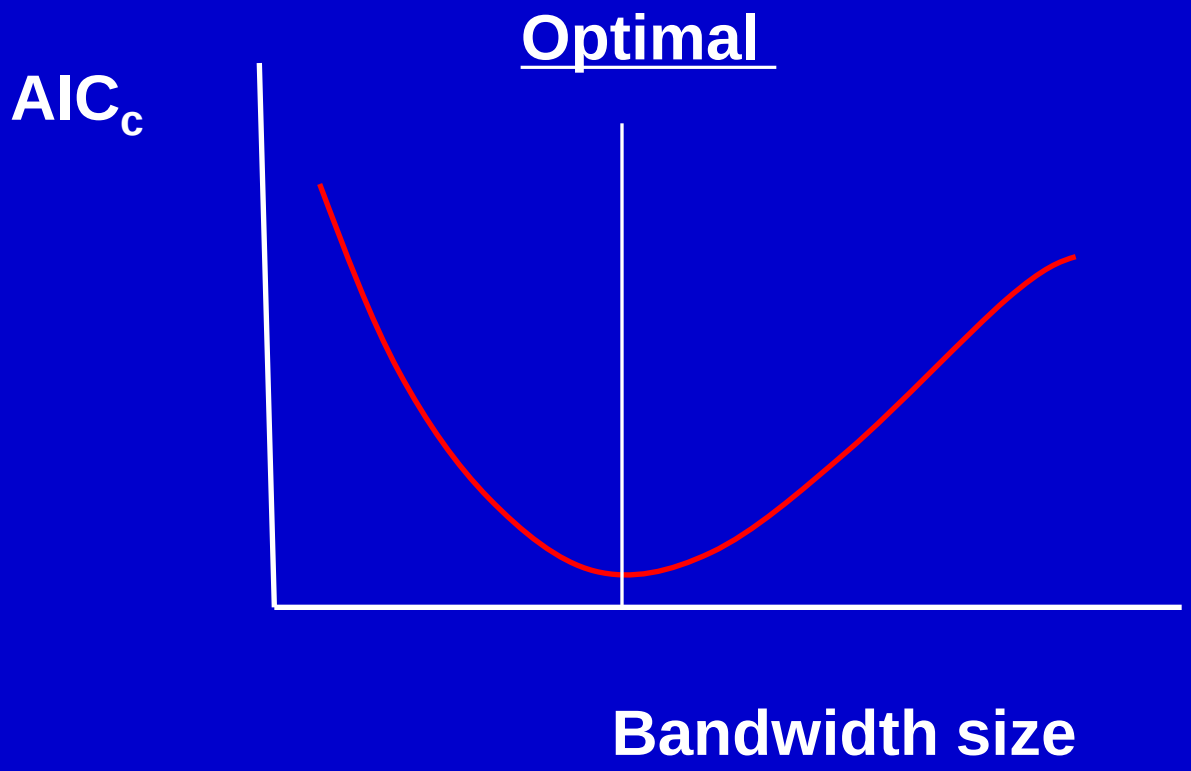
Here, we find the optimal value of N in the GWR routine

# 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 (CV) or the Akaike Information Criterion ($AIC_c$) ***or some other criterion!***


Optimal
AIC$_c$
Bandwidth size

# Bandwidth Selection

Optimal bandwidth selection is a trade-off between bias and variance

- Too small a bandwidth leads to large variance in the local estimates
- Too large a bandwidth leads to large bias in the local estimates

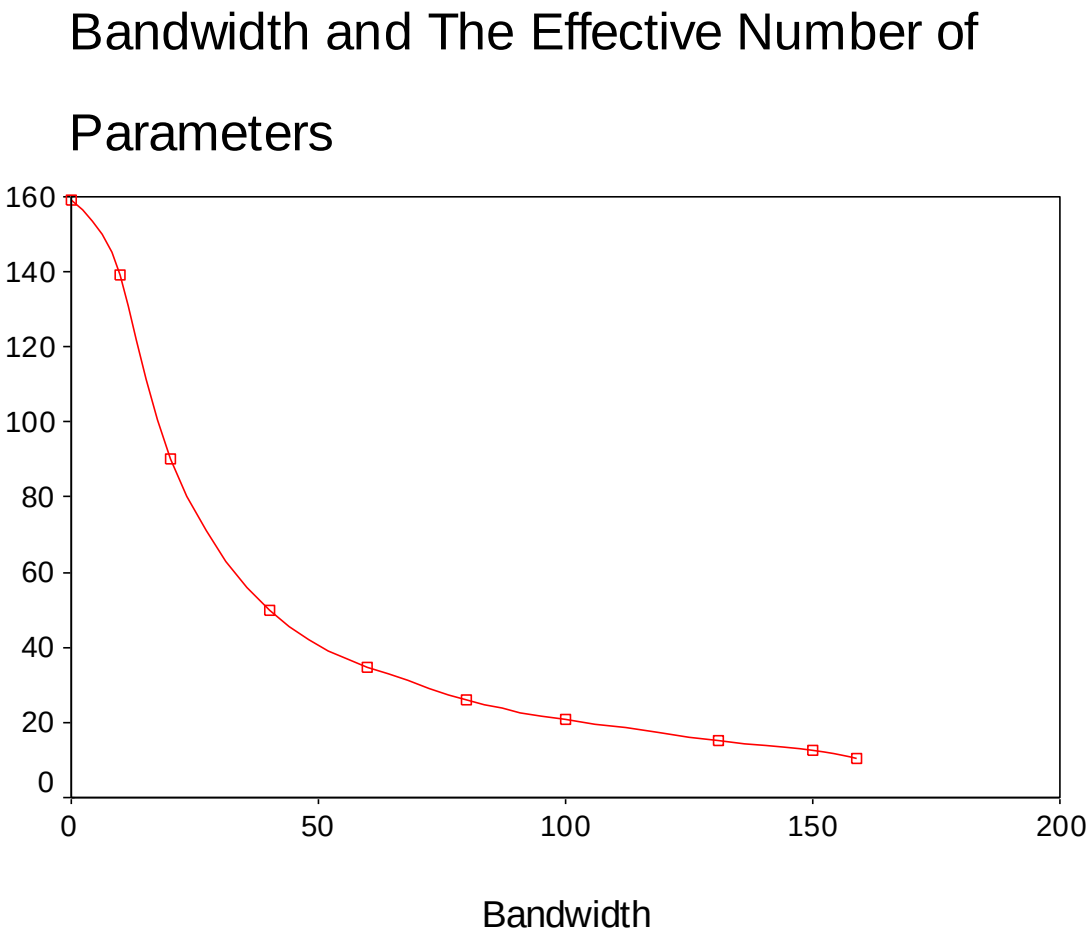
# Bandwidth and Effective Numbers of Parameters

As the bandwidth $\rightarrow \infty$, the local model will tend to the global model with number of parameters = k.

As the bandwidth $\rightarrow 0$, the local model 'wraps itself around the data' so the number of parameters = n

The number of parameters in local models therefore ranges between k and n and depends on the bandwidth. This number need not be an integer and we refer to it as **the effective number of parameters in the model**

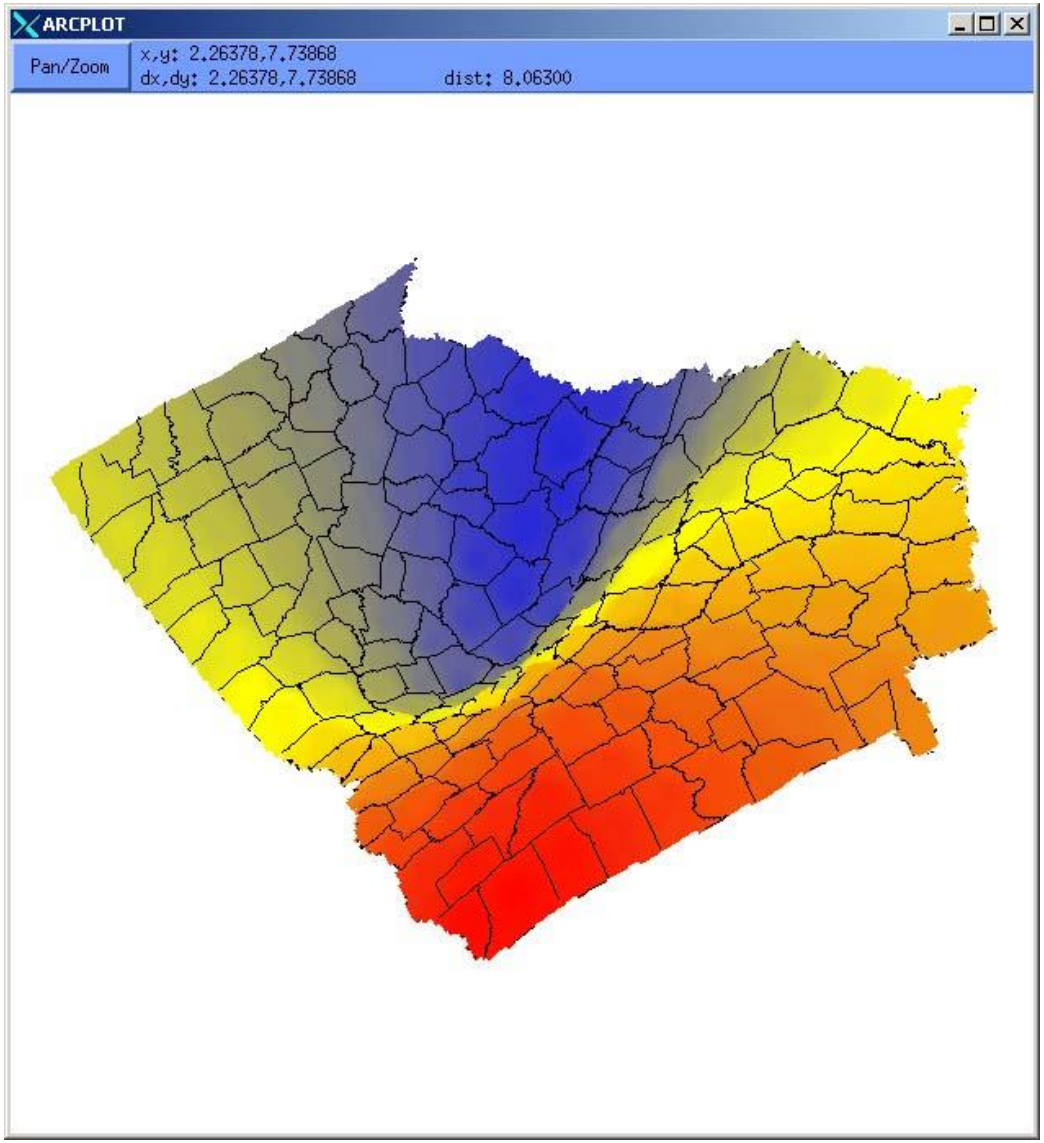
# An Example from the Georgia Data



# Output from GWR

Main output from GWR is a set of location-specific parameter estimates which can be mapped and analysed to provide information on spatial non-stationarity in relationships.

# An Example using Educational Attainment Data in Georgia



# In GWR, we can also ...

- estimate local standard errors
- derive local t statistics
- calculate local goodness-of-fit measures
- perform tests to assess the significance of the spatial variation in the local parameter estimates

**PLUS** an advantage of this method of borrowing data from nearby locations is that an optimal bandwidth, or decay rate, is determined - tells us something about the spatial scale over which the processes operate.

**If the processes are local, the opt. bandwidth will be small;**

**If the processes are regional, the opt. bandwidth will be large;**

**If the processes are global, the opt. bandwidth will tend to infinity.**

An extension to GWR, Multiscale GWR, or MGWR, takes this further and allows the spatial scale over which processes take place to vary by process. (more on this in second lecture…)

# A Simulation Experiment

$$Y_i = \alpha_i + \beta_{1i} X_{1i} + \beta_{2i} X_{2i}$$

Data on $X_1$ and $X_2$ drawn randomly for 2500 locations on a 50 x 50 matrix s.t. r($X_1$, $X_2$) is controlled. *Results shown to be independent of r($X_1$,$X_2$)*

## Experiment 1: (parameters spatially invariant)

**$\alpha_i$ = 10 for all i**

**$\beta_{1i}$ = 3 for all i**

**$B_{2i}$ = -5 for all i**

**$Y_i$ obtained from above**

**Data used to calibrate model by global regression and by GWR using an adaptive bandwidth**

# Results…

## Global:

**Adj. $R^2$ = 1.0  AIC = -59,390  K = 3**

**$\alpha$ (est.) = 10; $\beta_1$ (est.) = 3; $\beta_2$ (est.) = -5**

## GWR:

**Adj. $R^2$ = 1.0  AIC = -59,386  K = 6.5**

**N = 2,434**

**$\alpha_i$ (est.) = 10 for all i**

**$\beta_{1i}$ (est.) = 3 for all i**

**$\beta_{2i}$ (est.) = -5 for all i**

## Conclusion:

**GWR does NOT indicate any spurious nonstationarity when relationships are constant**

## Experiment 2: (parameters spatially variant)

$0 \le i \le 50 \qquad 0 \le j \le 50$

$\alpha_i = 0 + 0.2i + 0.2j$ 0 to 20

$\beta_{1i} = -5 + 0.1i + 0.1j$ -5 to 5

$B_{2i} = -5 + 0.2i + 0.2j$ -5 to 15

$Y_i$ obtained in same way

Data used to calibrate model by global regression and by GWR using an adaptive bandwidth

# Results…

## Global:

**Adj. $R^2$ = 0.04  AIC = 17,046  K = 3**

**$\alpha$ (est.) = 10.26; $\beta_1$ (est.) = -0.1; $\beta_2$ (est.) = 5.28**

***These are close to the averages of the local estimates (10;0;5)***

## GWR:

**Adj. $R^2$ = 0.997  AIC = 2,218  K = 167**

**N = 129**

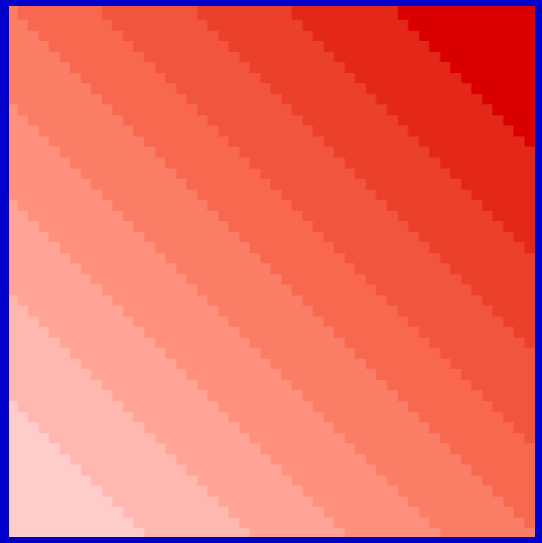
**$\alpha_i$ (est.) range = 2 to 18.6**

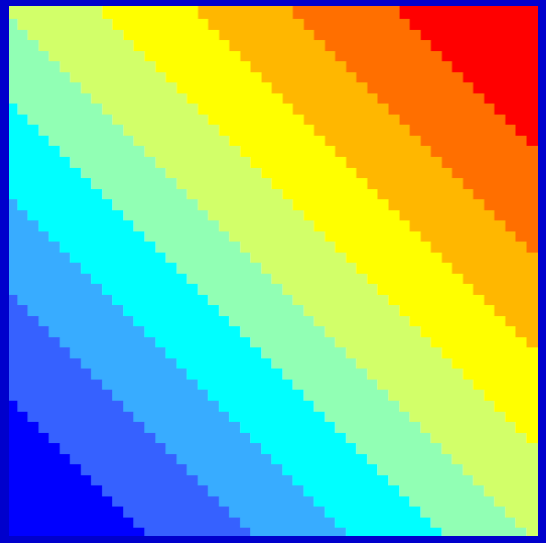
**$\beta_{1i}$ (est.) range = -4.3 to 4.7**

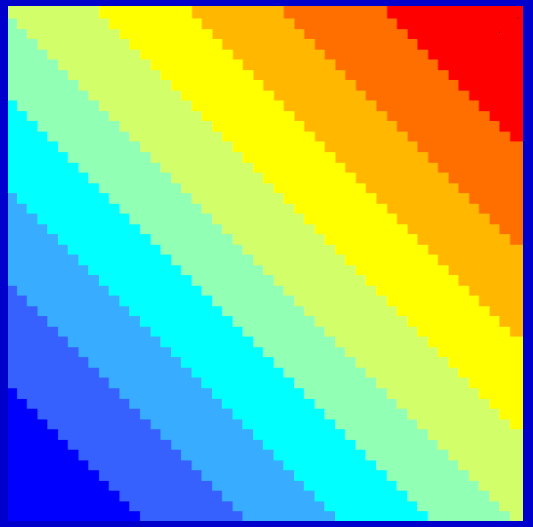
**$\beta_{2i}$ (est.) range = -3.9 to 13.6**

## Conclusion:

**GWR identifies spatial nonstationarity in relationships; global model fails completely.**



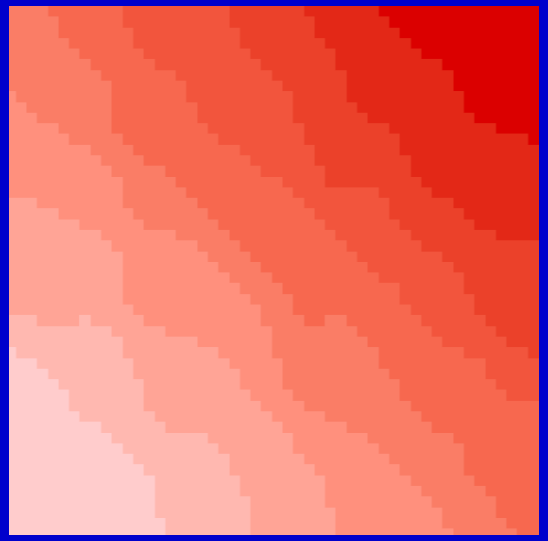


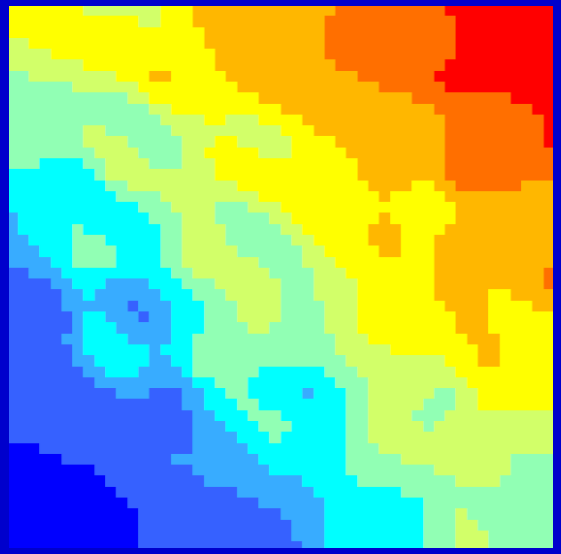


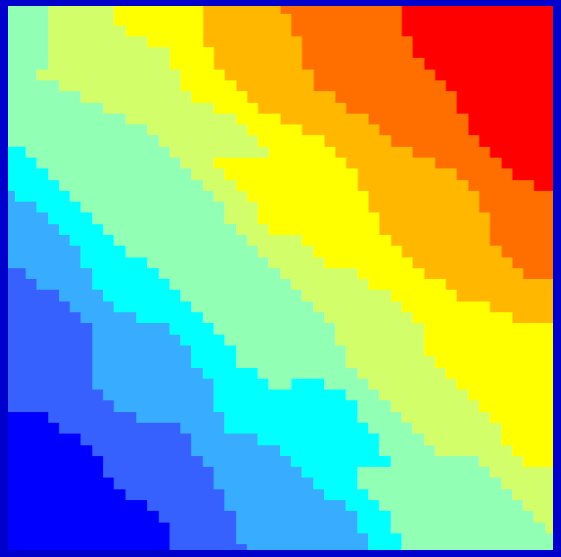
$0 \leq \alpha(i) \leq 20$

$-5 \leq \beta 1(i) \leq 5$

$-5 \leq \beta 2(i) \leq 15$



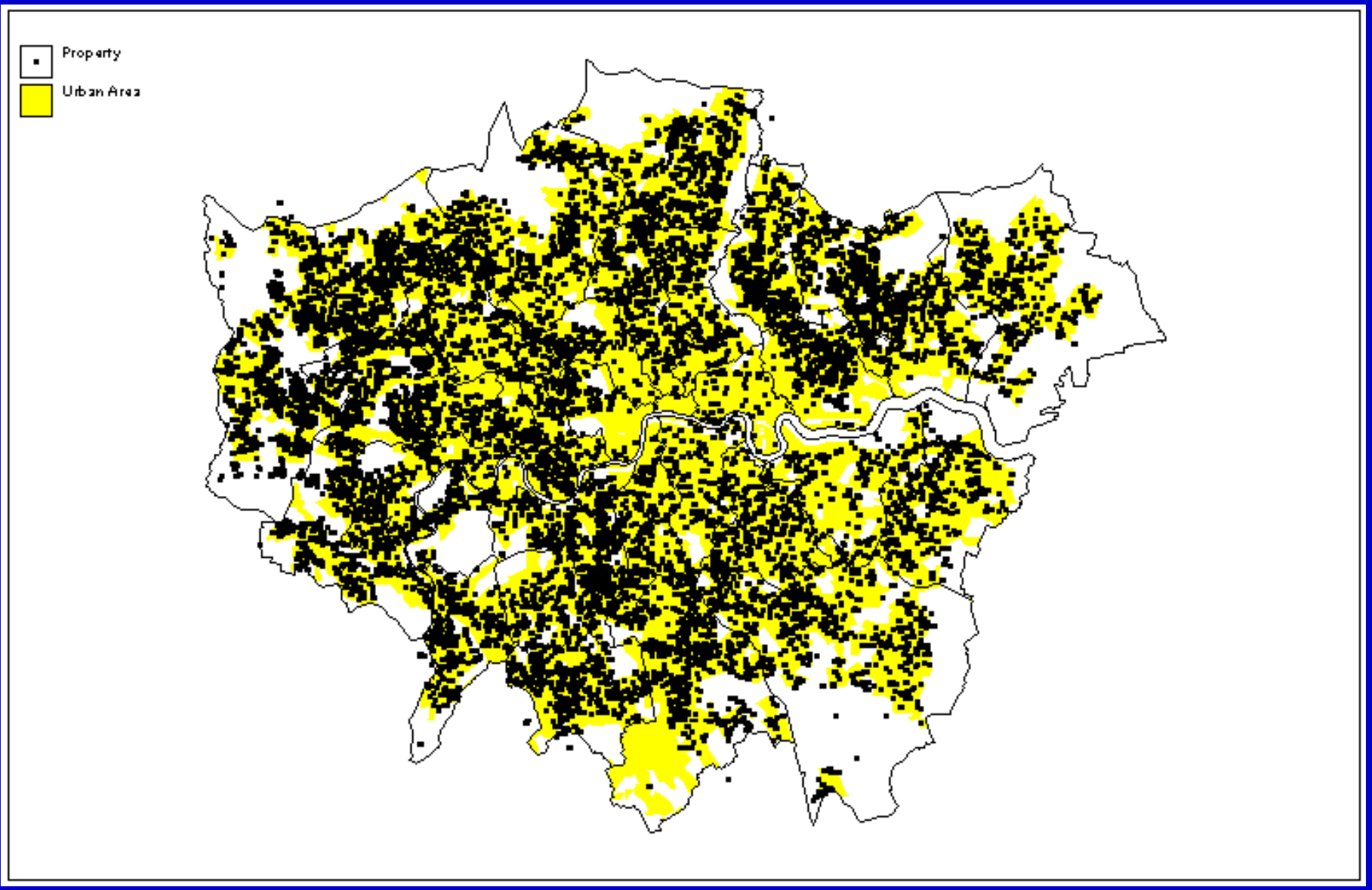




# An Empirical Example - House Prices in London

- 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)*

# Locations of house sales in data set



**To what extent are differences in average house prices a function of differences in the intrinsic value associated with different areas and to what extent are they due to different mixes of properties?**

To answer this, we need regression techniques to account for variations in housing attributes so that we can derive a comparable value per sq.m.

## 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

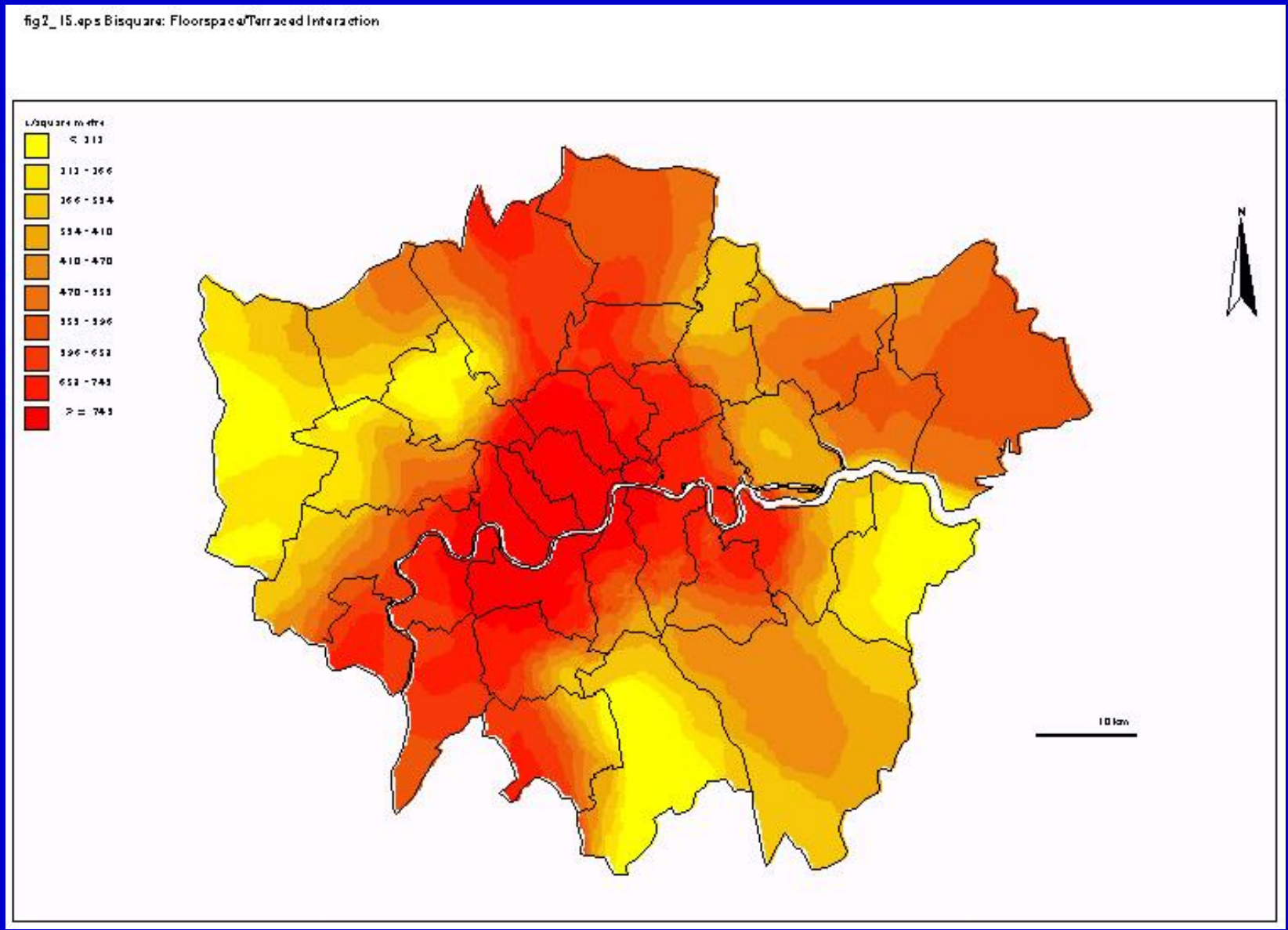
**However, these are all global results, i.e. averages over the whole of London.**

**Might there be differences across London in some of these relationships?**

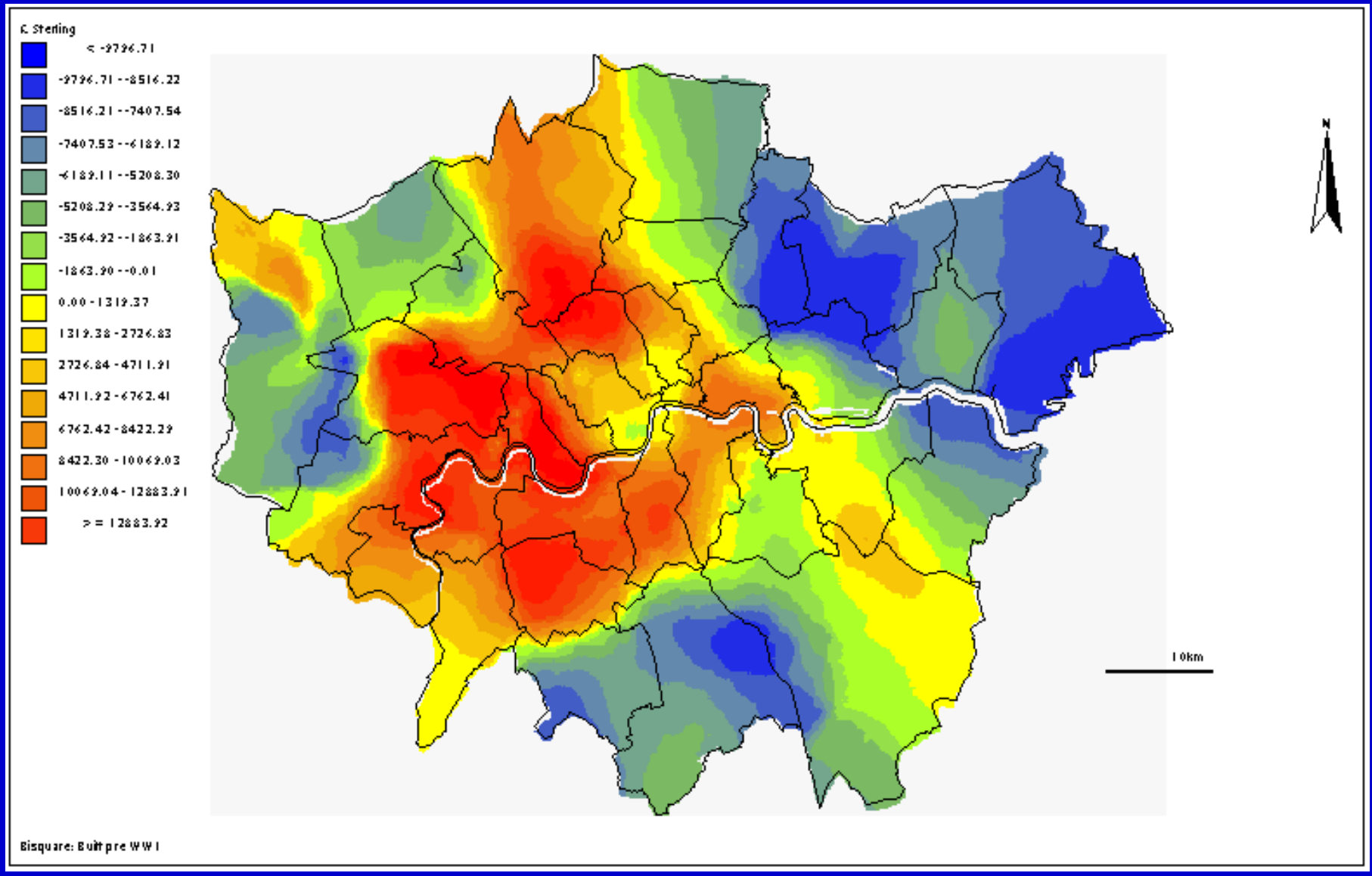
# Using GWR

- In this case an adaptive kernel was 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

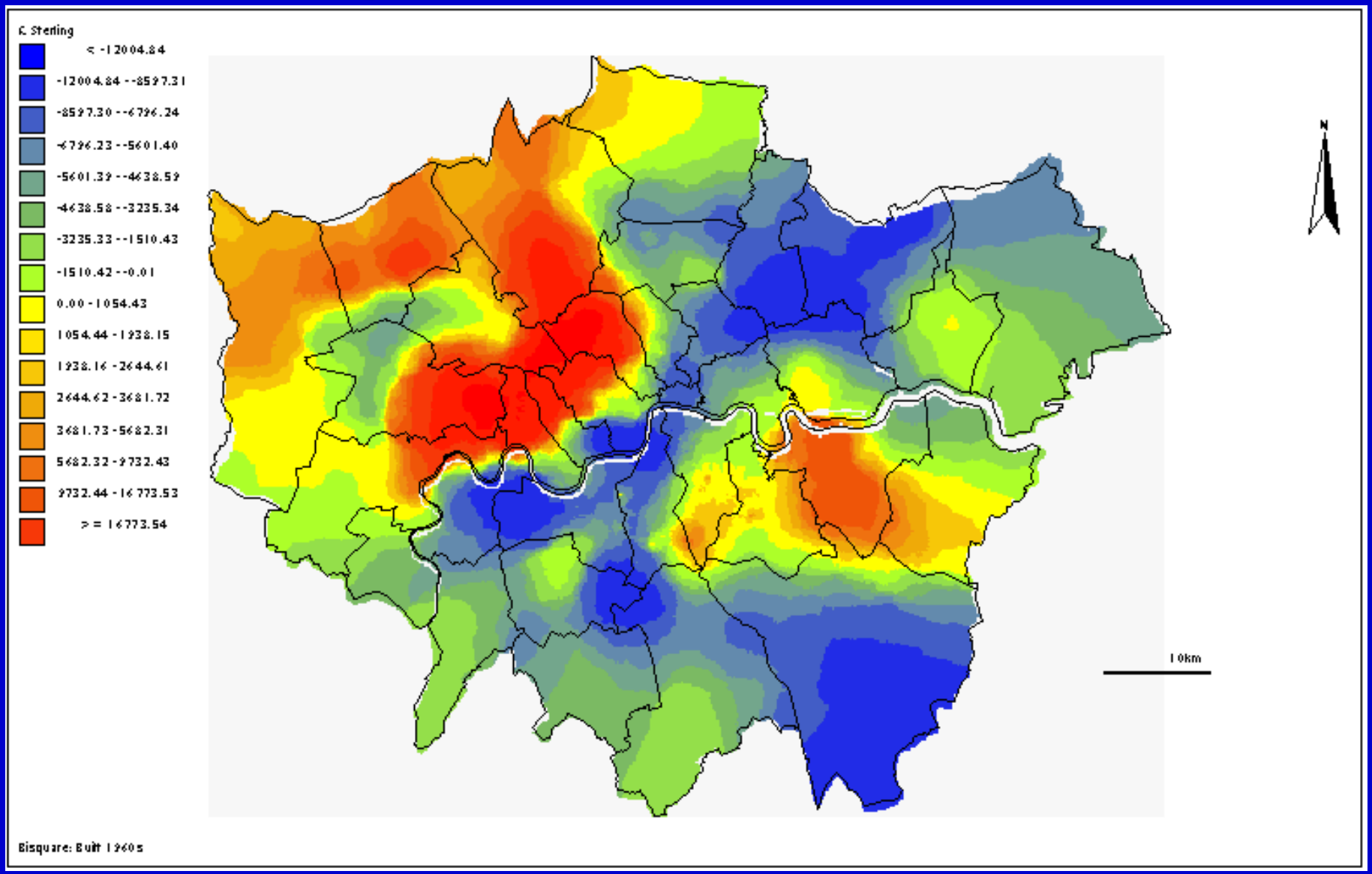
# Value of terraced property £/m² (global estimate = £578)



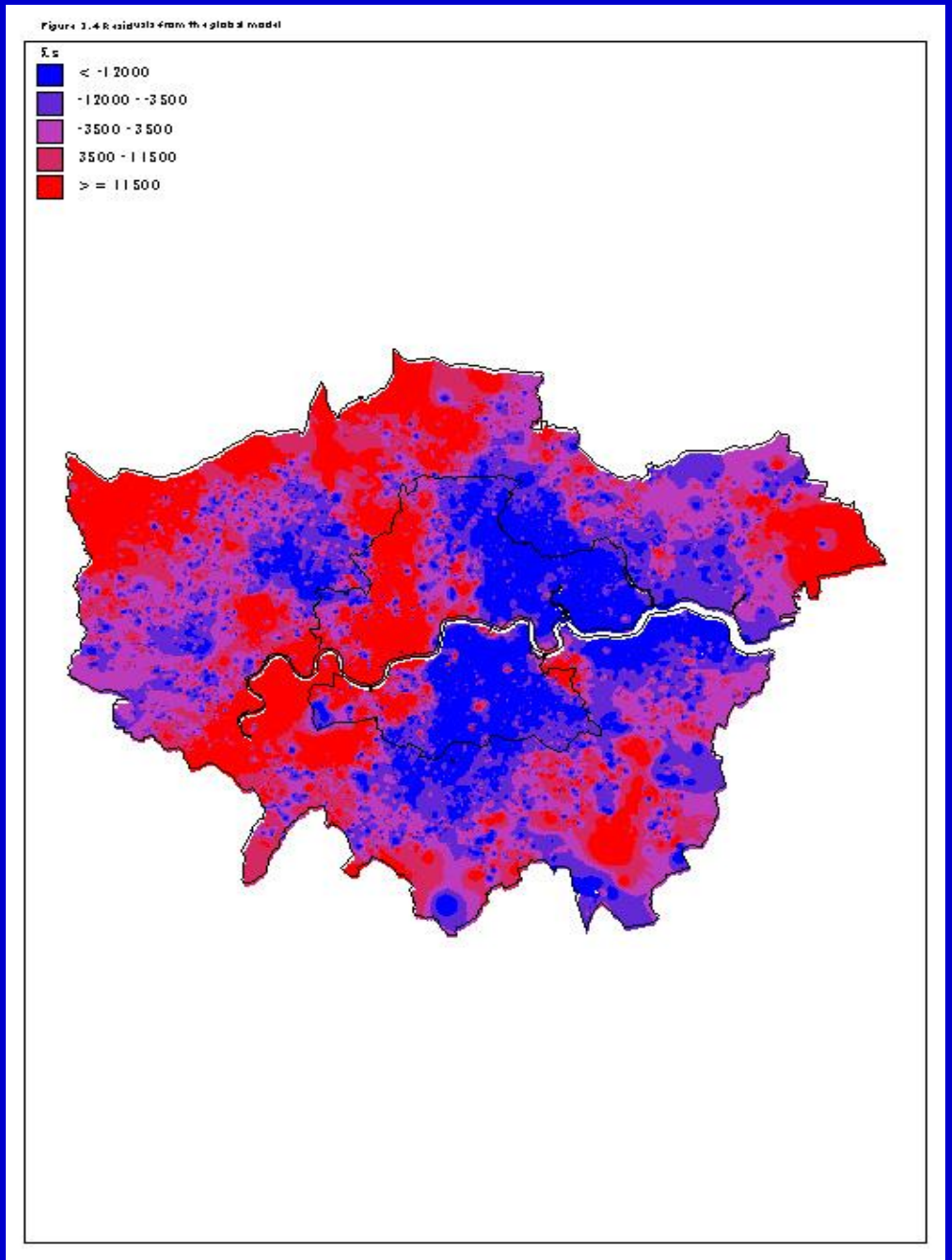
# Pre-1914 housing compared to inter-war (global estimate = £-2,340)



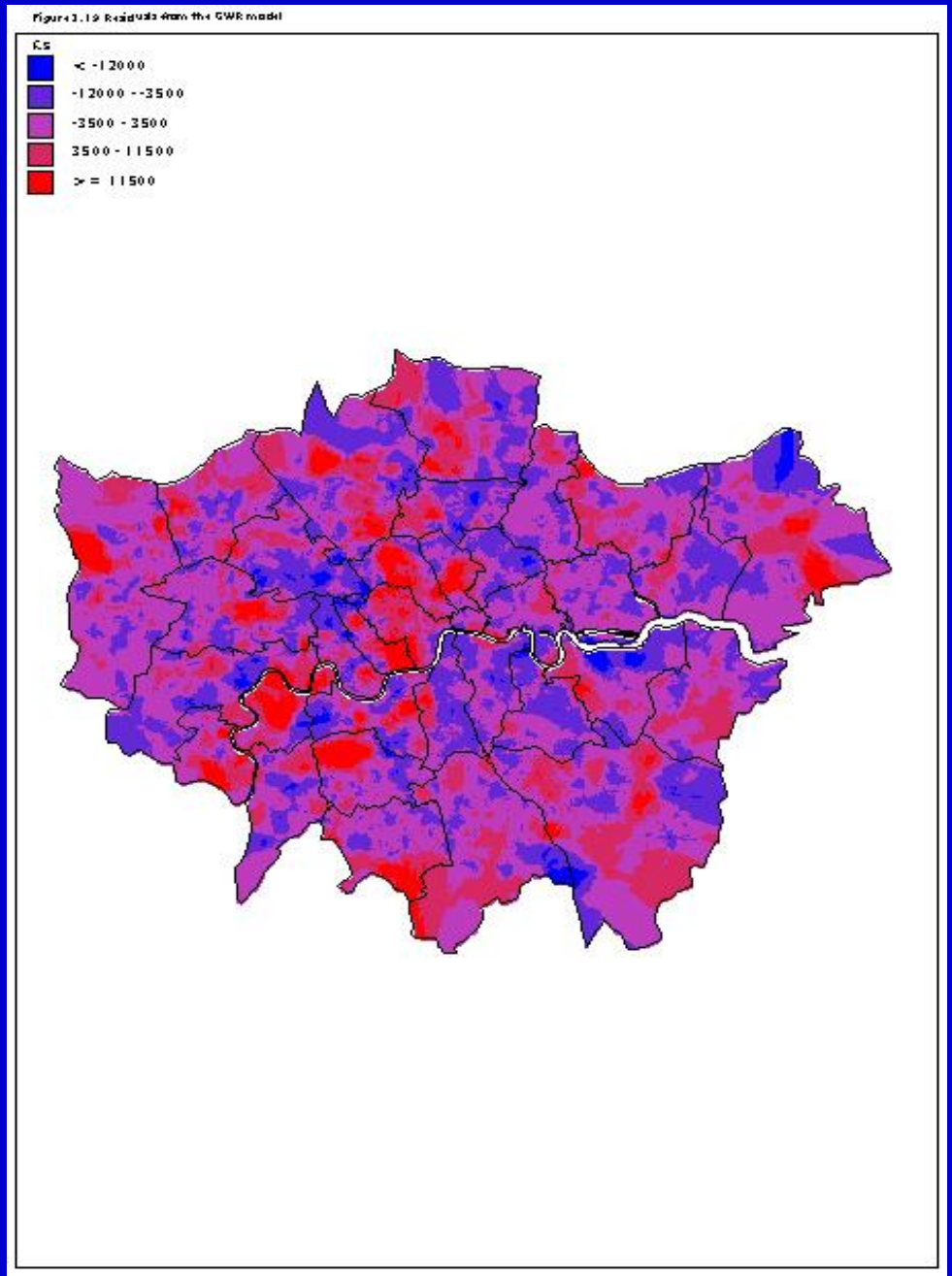
# 1960s housing compared to inter-war (global estimate = £-5,177)



# Residuals from Global Model



# Residuals from GWR Model



# The concept of geographical weighting can be applied to many other models

- GW poisson regression
- GW logistic models
- GW principal components analysis
- GW discriminant analysis
- GW spatial interaction models *(more on this in second lecture…)*
- etc…

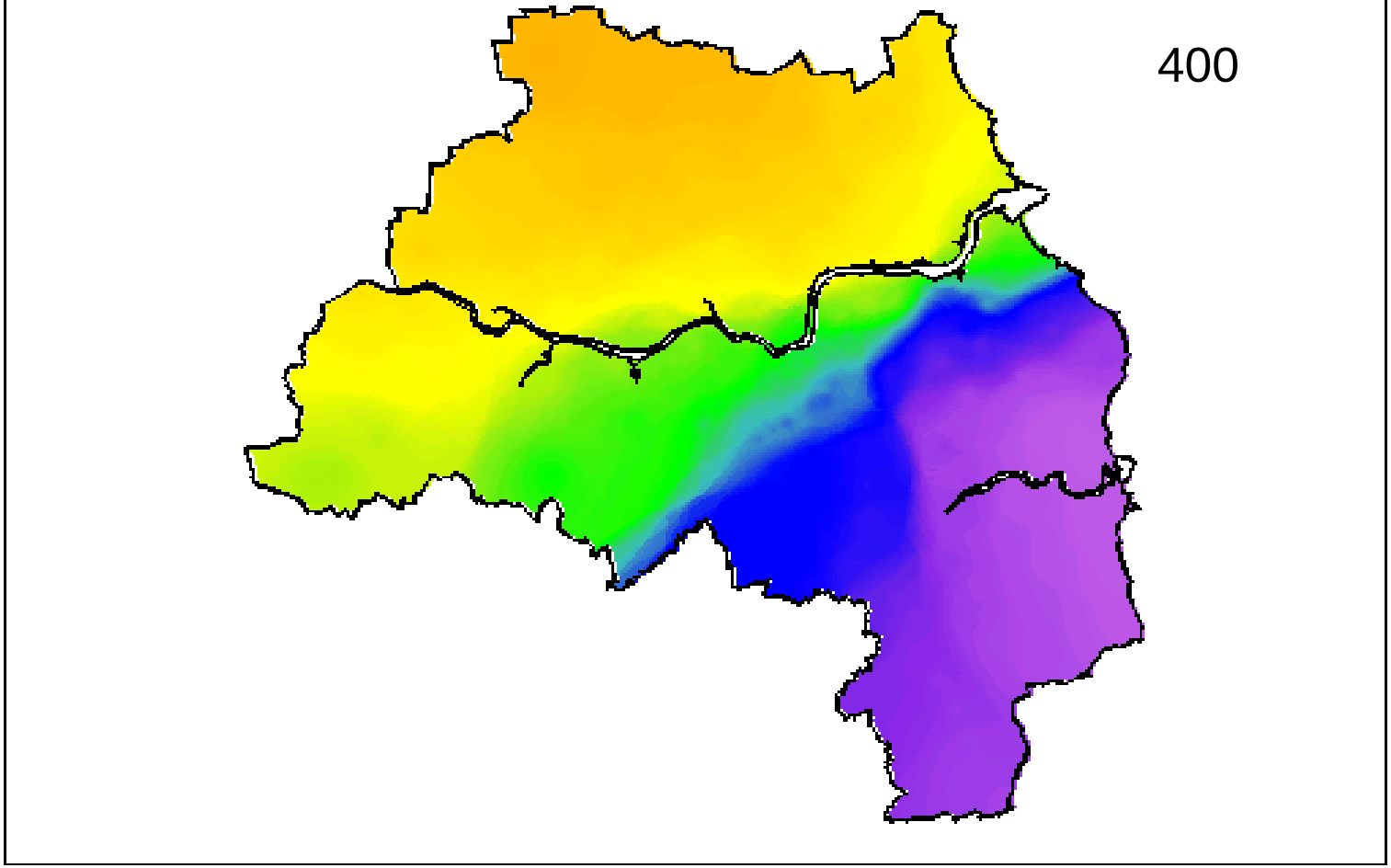
# Can use GWR as a 'Spatial Microscope'

- Instead of determining an optimal bandwidth during the calibration of a GWR model, a bandwidth can be input *a priori.*
- A series of bandwidths can be selected and the resulting parameter surfaces examined at different levels of smoothing
- For example, consider a very simple model of house prices regressed on floor area for 570 houses in Tyne & Wear, North East England.
- Surfaces of the local floorspace parameter are derived for bandwidths corresponding to 400, 350, 300, 250, 200, 150, 100 and 50 NN

DRAW

Pan/Zoom

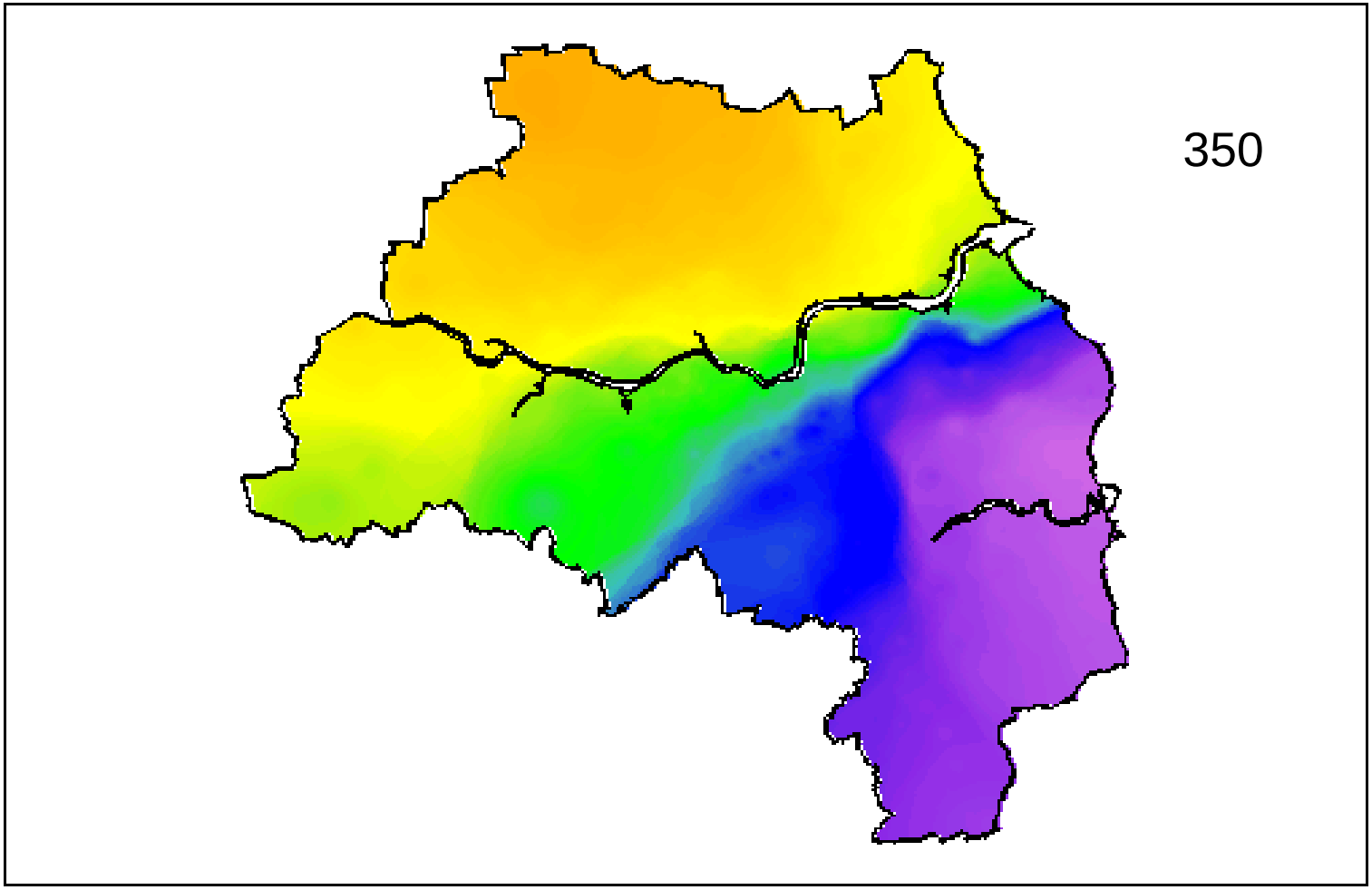
x,y: 8.68605,4.12444
dx,dy: 8.68605,4.12444 dist: 9.61554



DRAW

Pan/Zoom

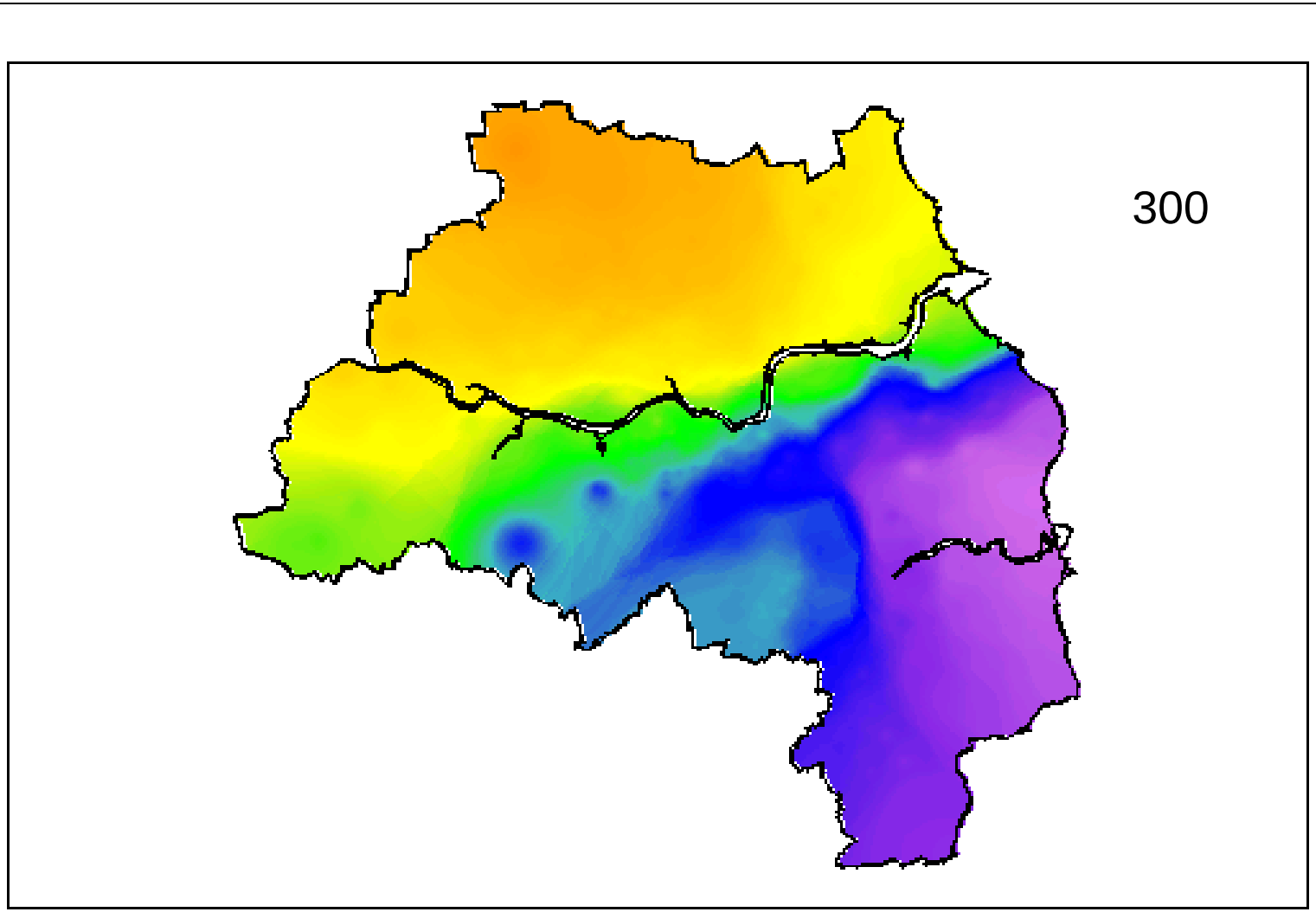
x,y: 8.34816,3.27970
dx,dy: 8.34816,3.27970 dist: 8.96929



DRAW

Pan/Zoom

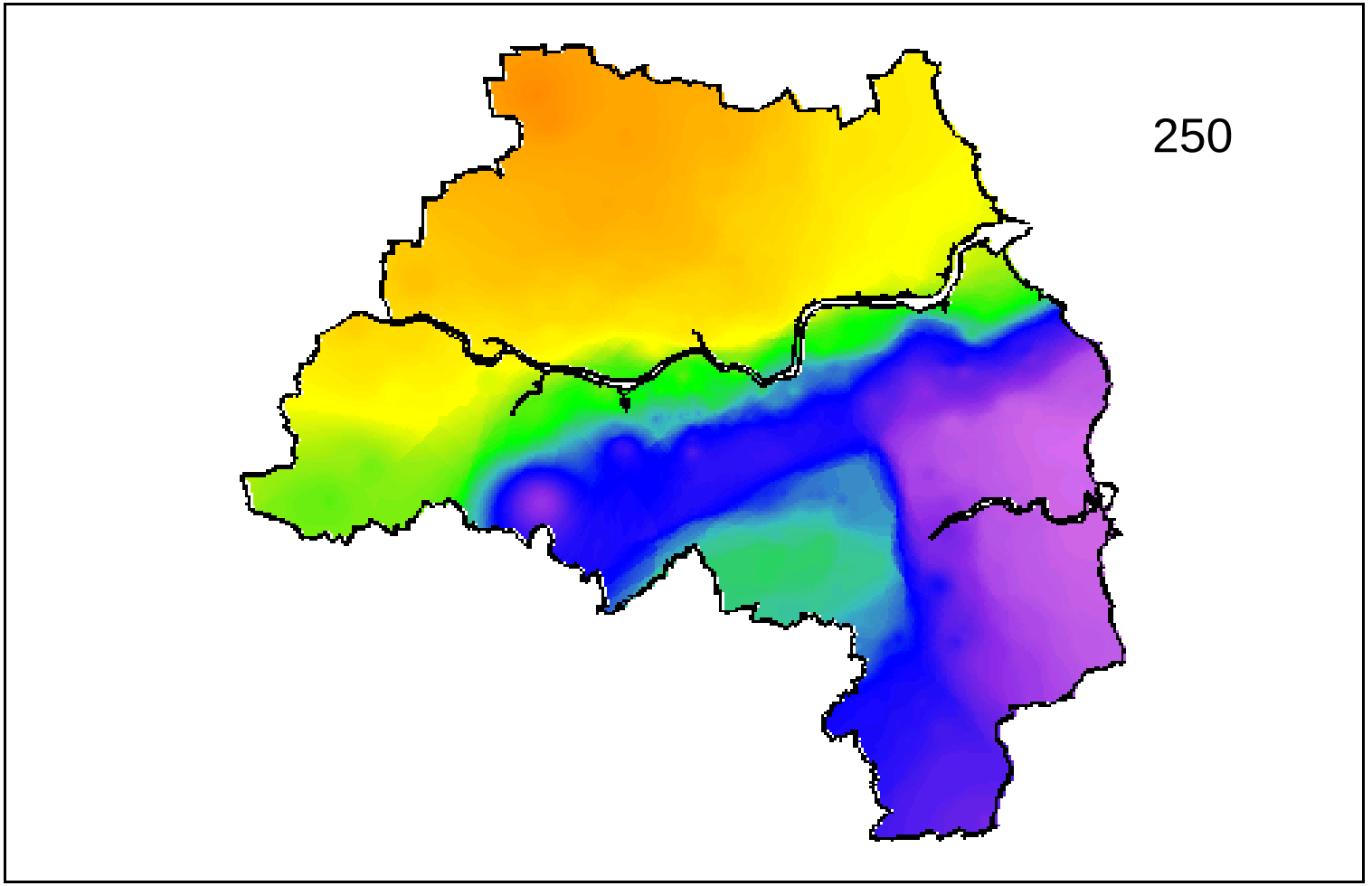
x,y: 8.13697,2.94180
dx,dy: 0.00000,0.00000 dist: 0.00000



DRAW

Pan/Zoom

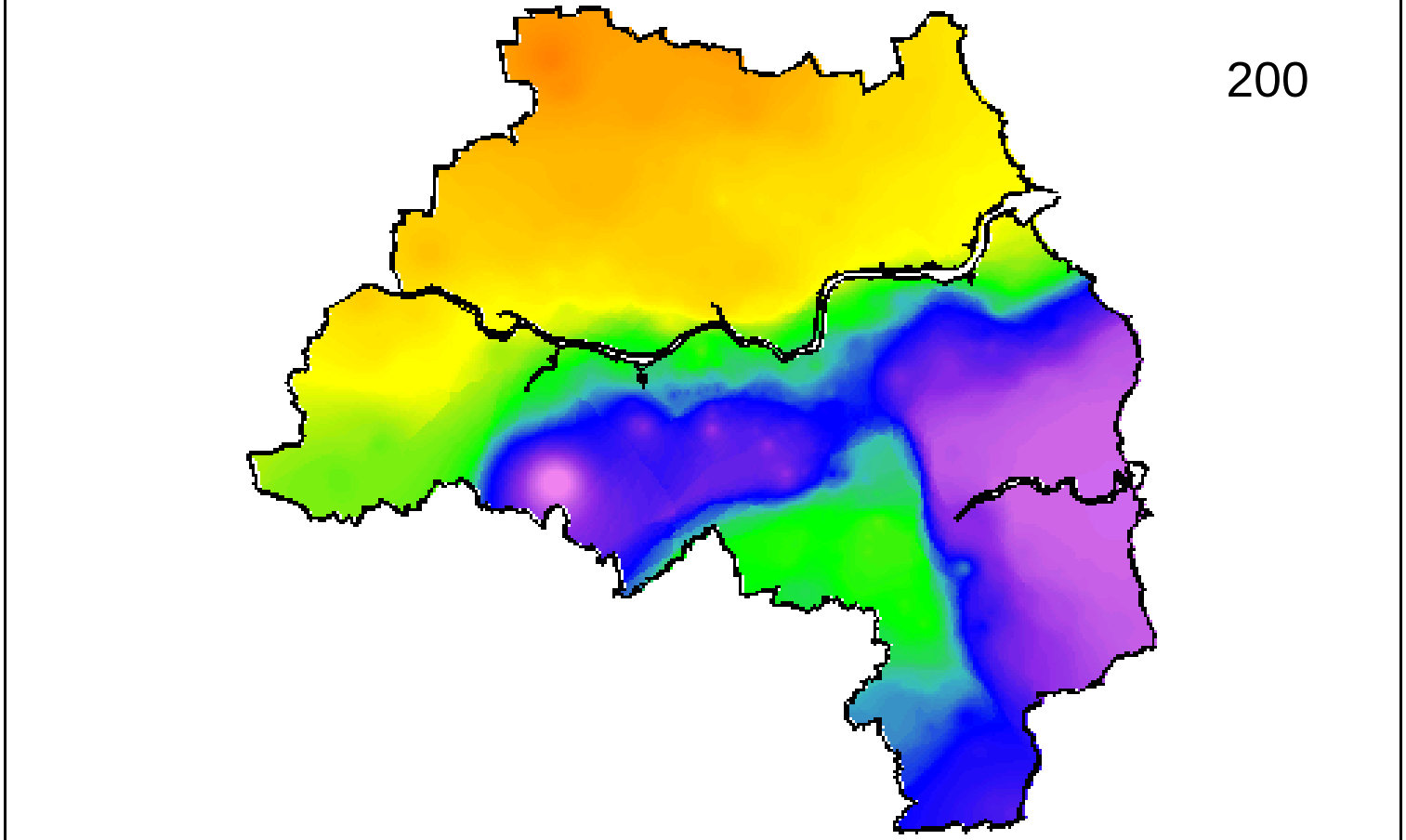
x,y: 7.16551,3.78654
dx,dy: 7.16551,3.78654 dist: 8.10448



DRAW

Pan/Zoom

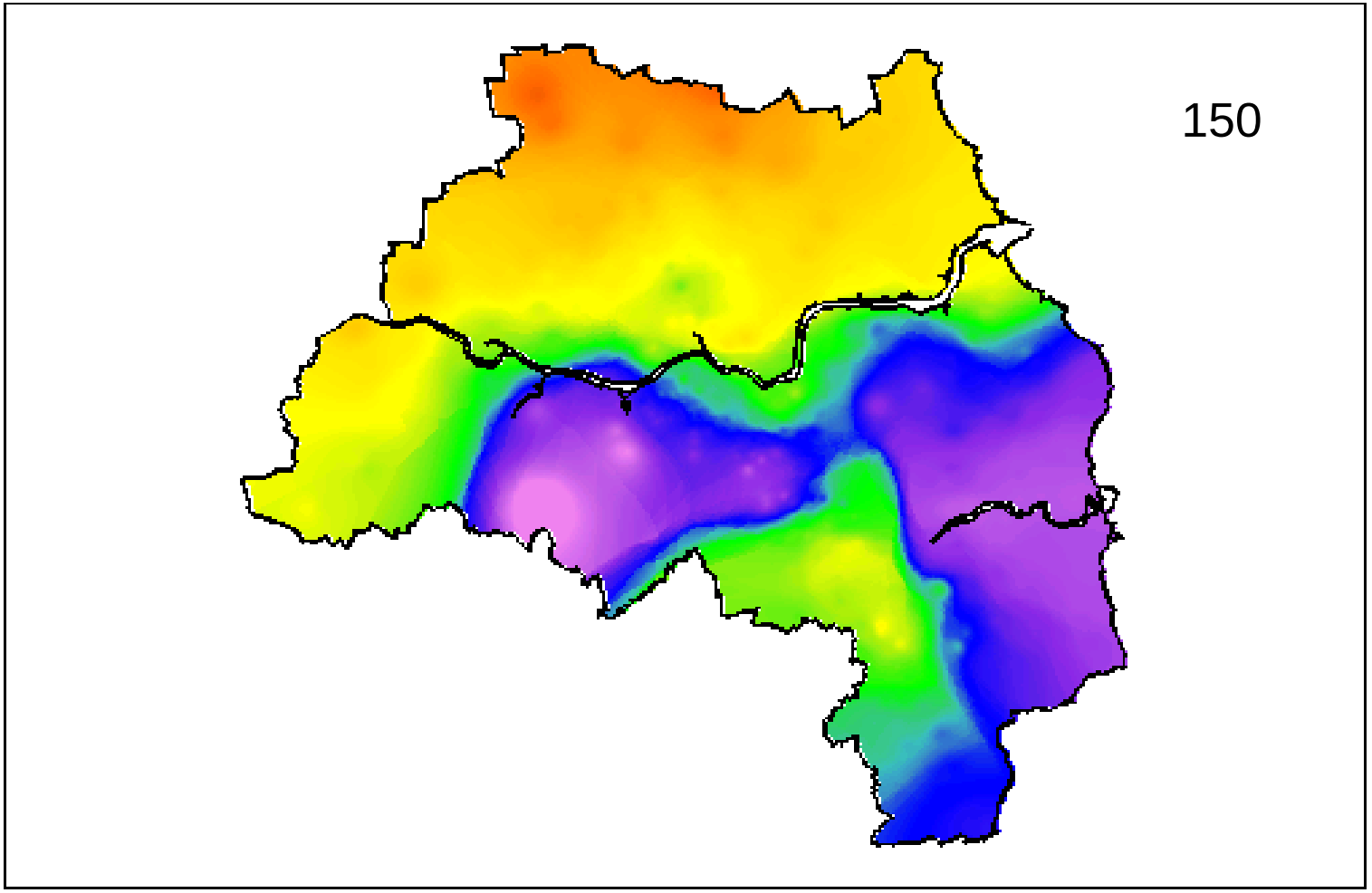
x,y: 10.18547,2.96292
dx,dy: 10.18547,2.96292 dist: 10.60767



DRAW

Pan/Zoom

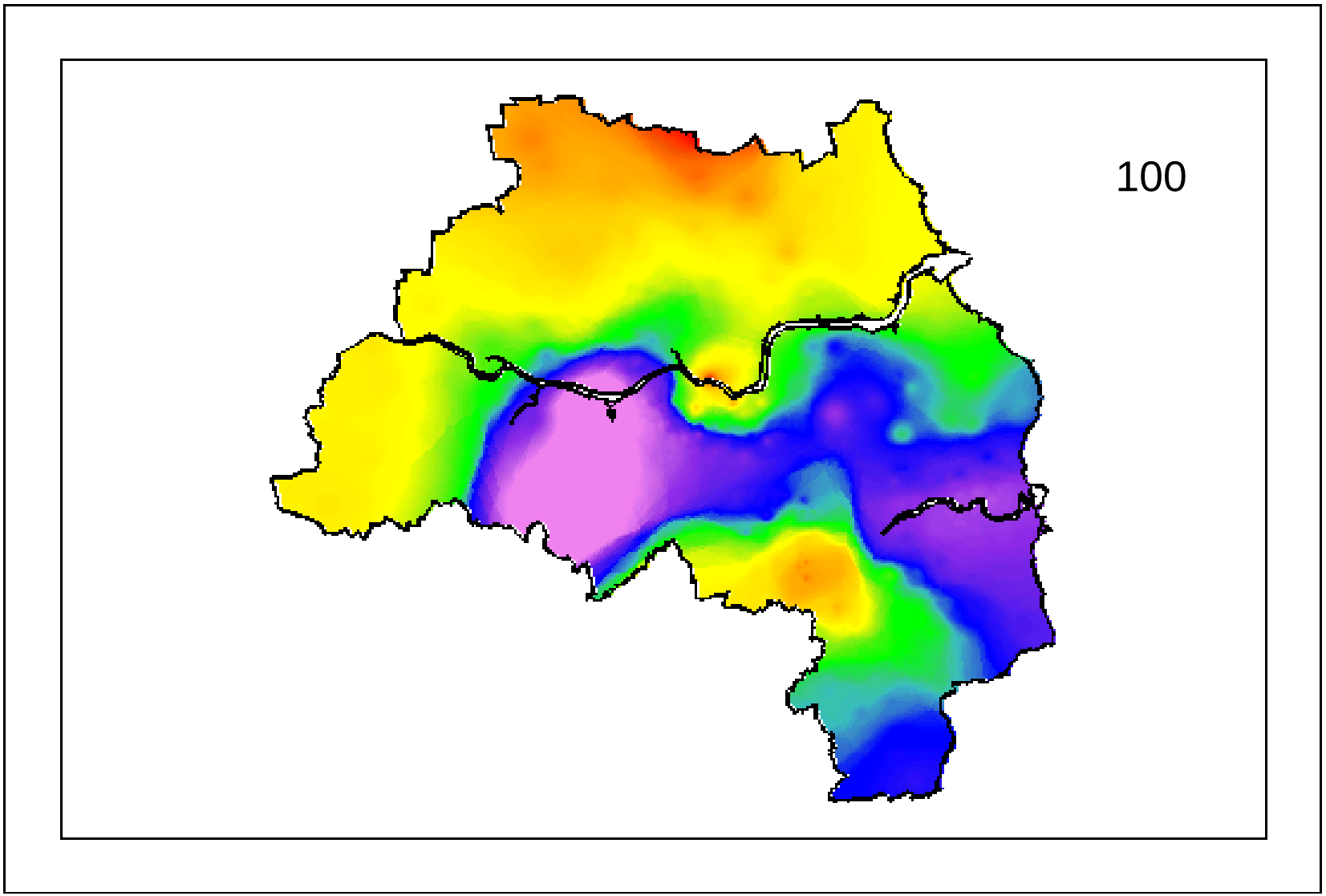
x,y: 6.91209,3.78654
dx,dy: 6.91209,3.78654 dist: 7.88130



DRAW

Pan/Zoom

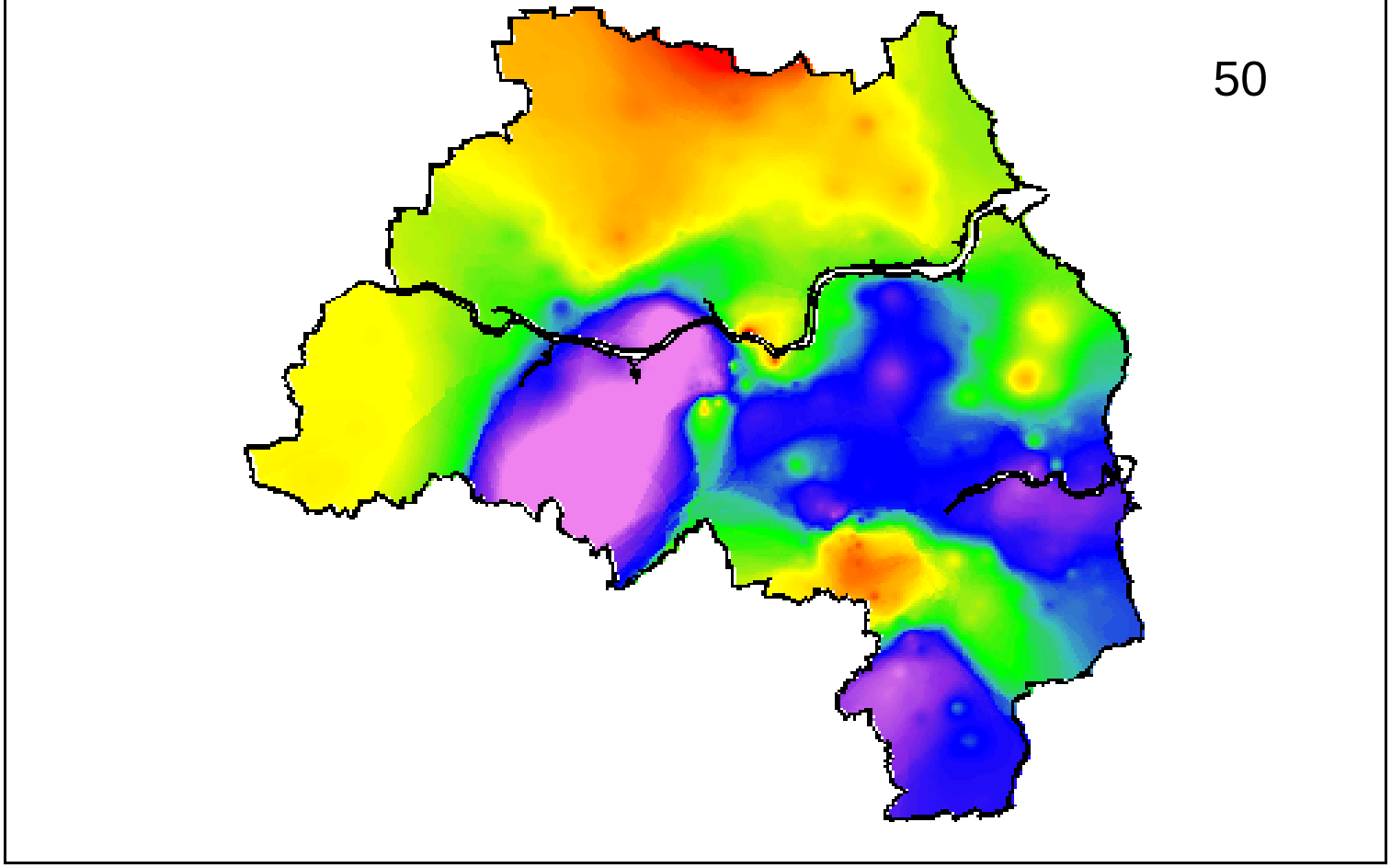
x,y: 3.00516,2.68838
dx,dy: 3.00516,2.68838 dist: 4.03216



DRAW

Pan/Zoom

x,y: 6.51084,4.99030
dx,dy: 6.51084,4.99030 dist: 8.20330



# Summary

**When used properly**, GWR provides a flexible, intuitive, easy-to-calibrate, scalable, extendable, framework that provides local parameter estimates, local standard errors, and local goodness-of-fit statistics, all of which help us think more deeply about our models of the real world.

It also provides information on the spatial scale at which different processes operate and

the concept is applicable beyond regression (e.g. GWPCA; GWDA; SWIM)

## What GWR is:

1. A means of modeling and detecting spatial non-stationarity in associations arising from spatial variations in attitudes, preferences or different administrative, political or other contextual effects that produce different responses to the same stimuli

**This view is in alignment with social theory!**

**GWR is a technique that is neither truly nomothetic nor truly idiographic**

## What GWR is:

2. A ‘data-borrowing’ technique - provides an intermediate level of analysis between observing multiple outcomes of a process at each location (the ideal) and only observing one outcome of a process at each location (reality).

GWR calibrates a model specific to location i by ‘borrowing’ data from nearby locations and weighting these data according to how far from i they are located

# GWR is also…

3. A means of identifying and quantifying the spatial scale at which processes occur via an optimal bandwidth.

4. As a side benefit – removes spatial autocorrelation of error terms resulting from applying a global model to spatially varying processes.

5. A diagnostic tool for detecting model misspecifcation if the processes being modeled are stationary

**This view is more in alignment with positivism**

## What GWR is not:

1. A means of improving a poorly specified  global model

2. A competitor to spatial regression models – you can have geographically weighted SAR and CAR models

3. A solution to any other potential problems inherent in the global model (except possibly spatially autocorrelated errors)

4. An excuse for not thinking about the processes that might have produced your data

## *King (1996: p 161) Political Geography*

***"Geographical tools are essential for displaying areal variation in what we know, but this is nowhere near as powerful as the role of geography in revealing features of data…that we would not otherwise have considered."***

***Thank you***

# 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 $\beta_i$

| | | | | |
|---|---|---|---|---|
| .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 $\beta_i$ 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 | - | - | - | - |

# The BIG question is then…

*When we observe spatially autocorrelated error terms in the regression of spatial data, how much of this is due simply to applying a global model to a non-stationary process or to model misspecification?*

**Quite a lot possibly!**

In order to ‘borrow’ data to calibrate a local model at a location where insufficient data are available, some borrowing rule has to be established. Possibilities include:

1. Borrow data from random sets of locations and try lots of different random sets and use the one optimizing some criterion. Ignores our experiences with spatial data. How robust would this be?

2. Borrow data from similar types of locations. For example, we could use MDS on the X covariates and weight locations by how similar they are to the regression location. Problem – couldn’t use this method to calibrate a model where no data are observed.

3. In accordance with Tobler's Law, borrow data from nearby locations under the premise that nearby locations are more likely to share similar processes than locations farther apart and weight the data according to how far away a location is from the regression point. Minimizes bias introduced to the calibration from using data which are the product of different processes.

The latter definition is based on the reasonable geographic assumption that if processes are spatially varying, then this variation is likely to exhibit spatial dependence.

**Rule 1:** Report and map local estimates divided by their local standard errors to produce local t, z or Wald statistics.

**Rule 2:** Adjust critical values of t or z to account for multiple comparisons. In GWR, this is done by first computing the Equivalent Number of Parameters (ENP) by 2Tr(**s**) – Tr(**s's**) and then using the following correction:

$$\lambda = \alpha / (ENP/P_g)$$

where $P_g$ is the number of parameters in the global model. Critical value of t corresponds to λ and not α

*Da Silva and Fotheringham* ***Geographical Analysis*** *2015*

Local parameter estimates will vary if different contextual effects produce different responses to the same stimuli, but also because of sampling variation

So we need to be careful to account for the latter to see if there is any extra variation.

We can do this through judicious significance testing – with the primary objective of not identifying false positives.

This involves following **three** rules:

**Rule 3:** Also test the significance of the spatial variability of each **set** of local estimates – numerous tests for this

So, don't read anything into or report

**(i) individual** local estimates that are not significant

**(ii) surfaces** of local estimates that do not exhibit any significant variation

# Obtaining local coefficients

Collect all model terms pertaining to each explanatory variable

$$local = \ global\ coefficient + \sum(local\ modifications)$$

$$X_p^{local} = \beta_p + \sum(\beta_{pE_k} * E_k)$$

$$intercept^{local} = intercept + \sum(\beta_{E_k} * E_k)$$

# Spatial Filter Specification

In a system with 10 covariates and 2500 locations

**10** **2,500** **250,000**

$$Y = \beta_0 1 + \sum_{p=1}^{P} X_p \cdot \beta_p + \sum_{k=1}^{K} E_k \beta_{E_k} + \sum_{p=1}^{P} \sum_{k=1}^{K} X_p \cdot E_k \beta_{pE_k} + \varepsilon$$

intercept

P global variables

K spatial filter eigenvectors

K*P potential interaction terms

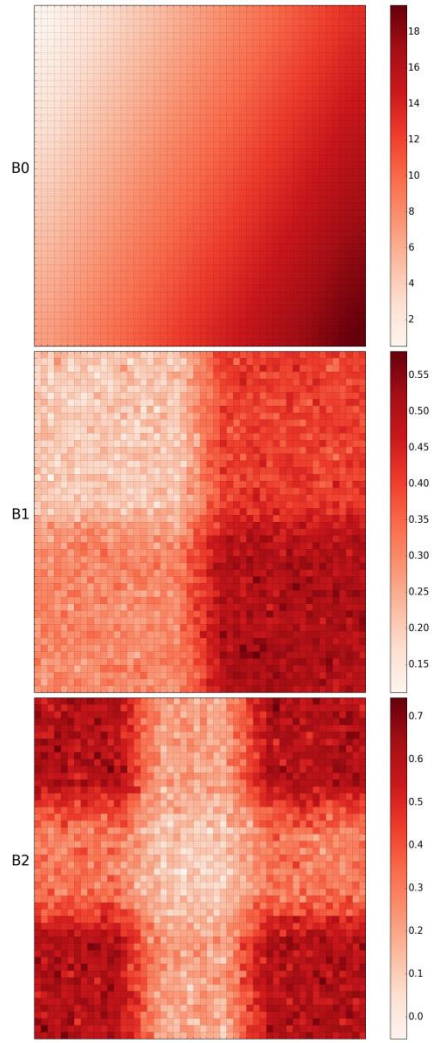
# Experiment

$$Y_i = B_{0i} + B_{1i}X_{1i} + B_{2i}X_{2i} + \varepsilon$$

$\beta_0$

$\beta_1$

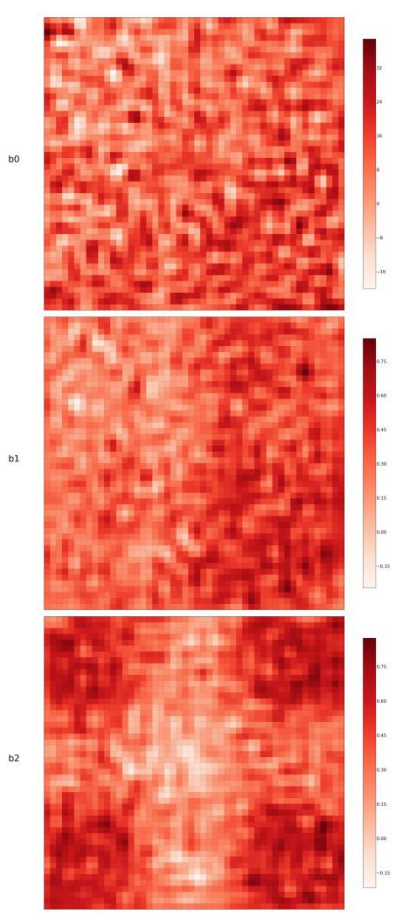
$\beta_2$



Spatial Filtering

Known

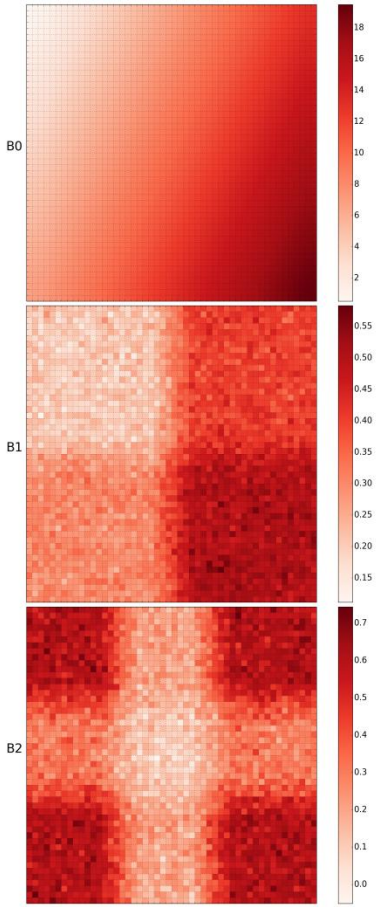
GWR

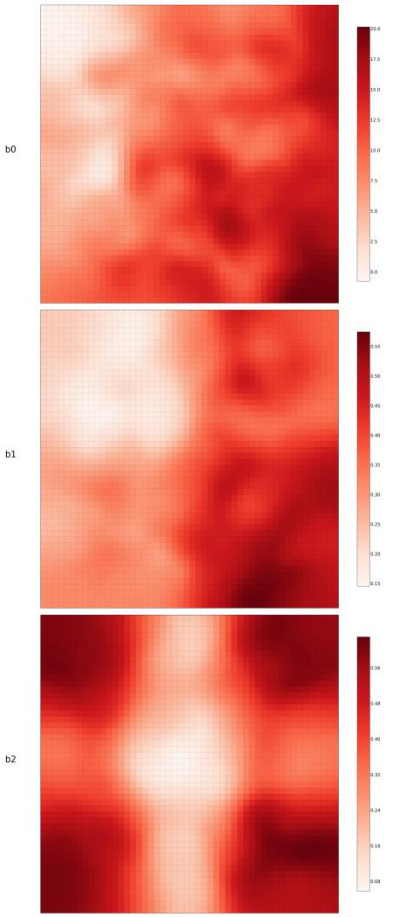



$\beta_0$

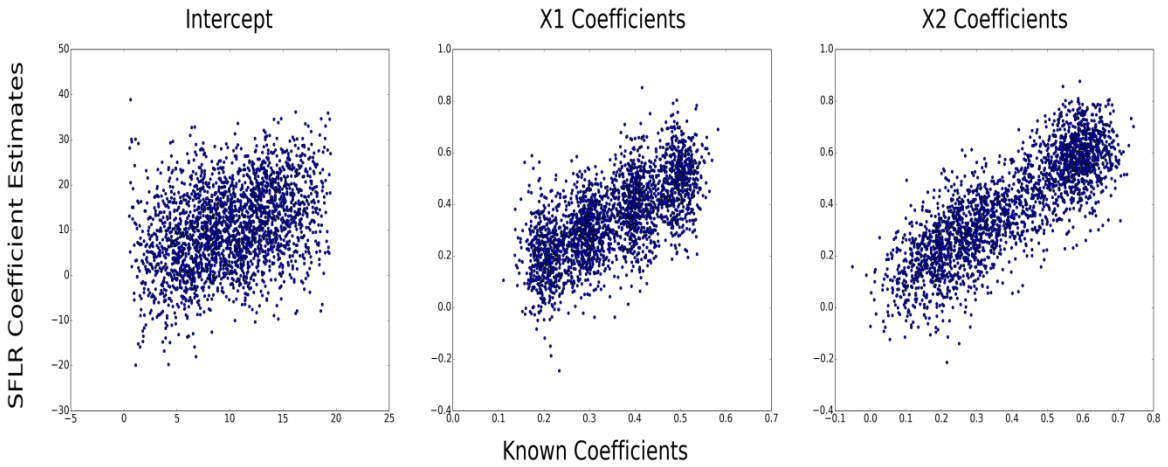
$\beta_1$

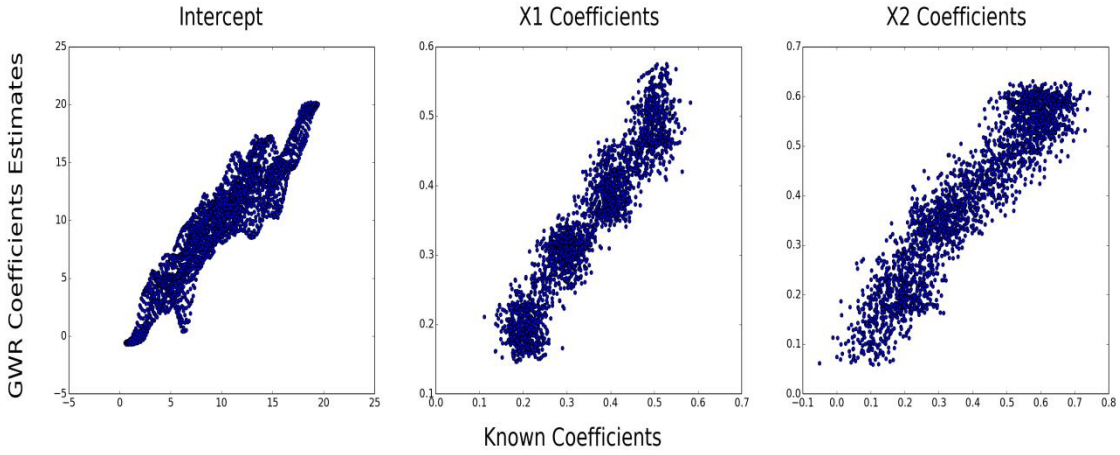
$\beta_2$



Spatial Filtering




GWR