Regression on a Graph

Andrew Smith

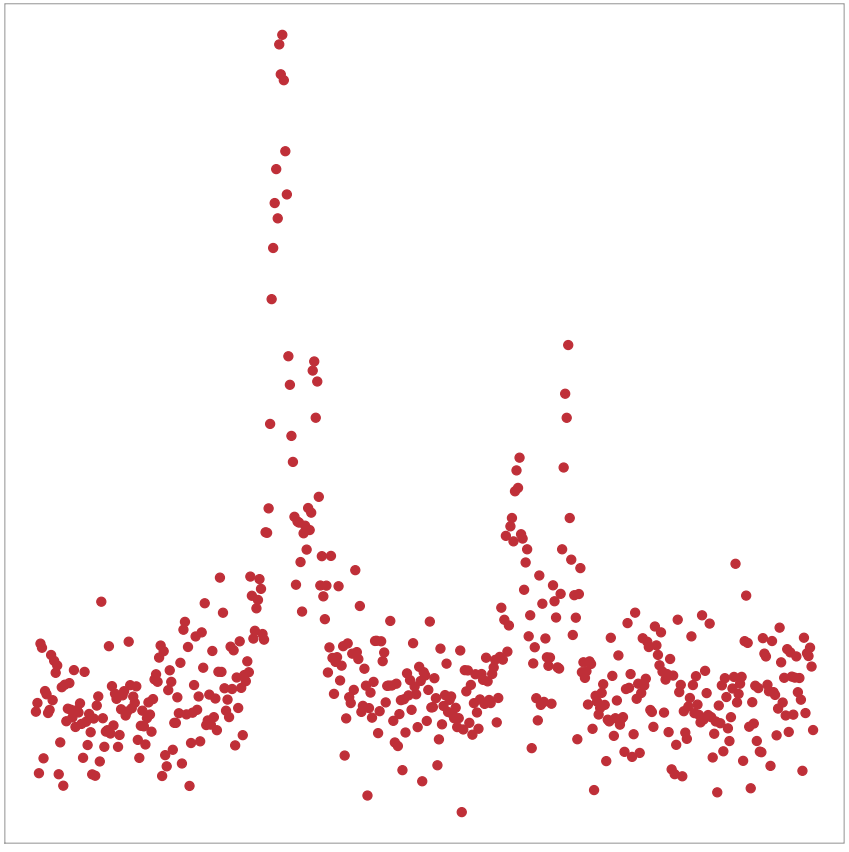
www.maths.bris.ac.uk/∼as1637

MINGLE, 7 October 2009

Regression on a Graph

1. Regression
   - Example: Scatterplot Smoothing
   - Example: Image Analysis
   - Example: UK House Prices
2. Graphs
3. Regression on a Graph
   - Signal + Noise Model
   - Graphs in Statistics
   - Regression Tradeoff

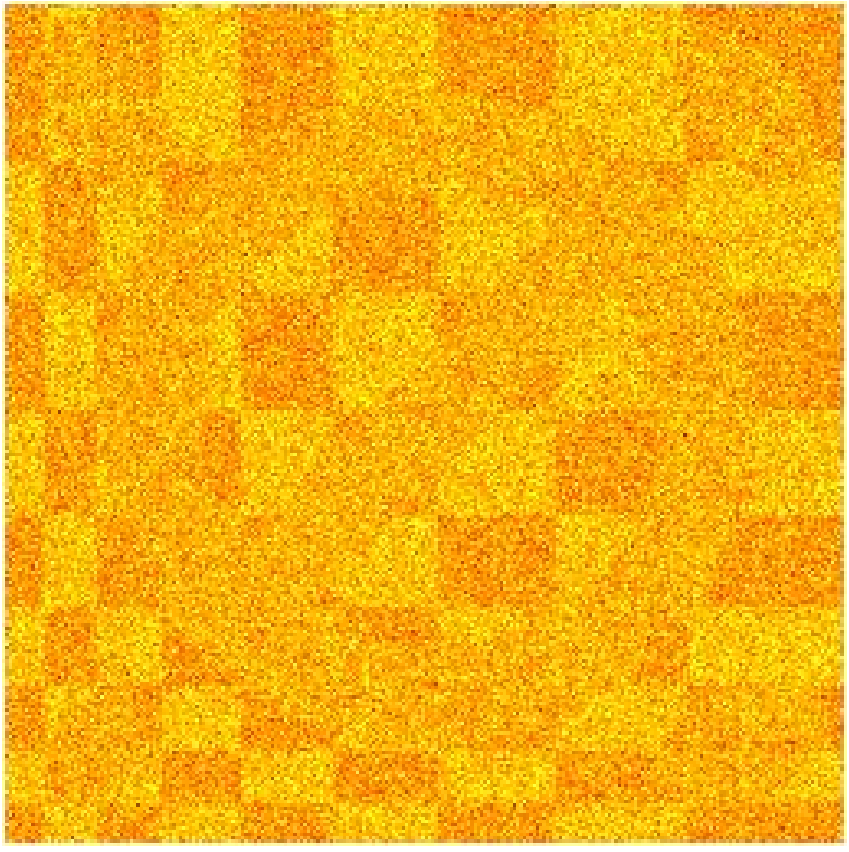
Regression: Scatterplot Smoothing



Davies and Kovac, 2001, Ann. Stat. 29, p1–65.

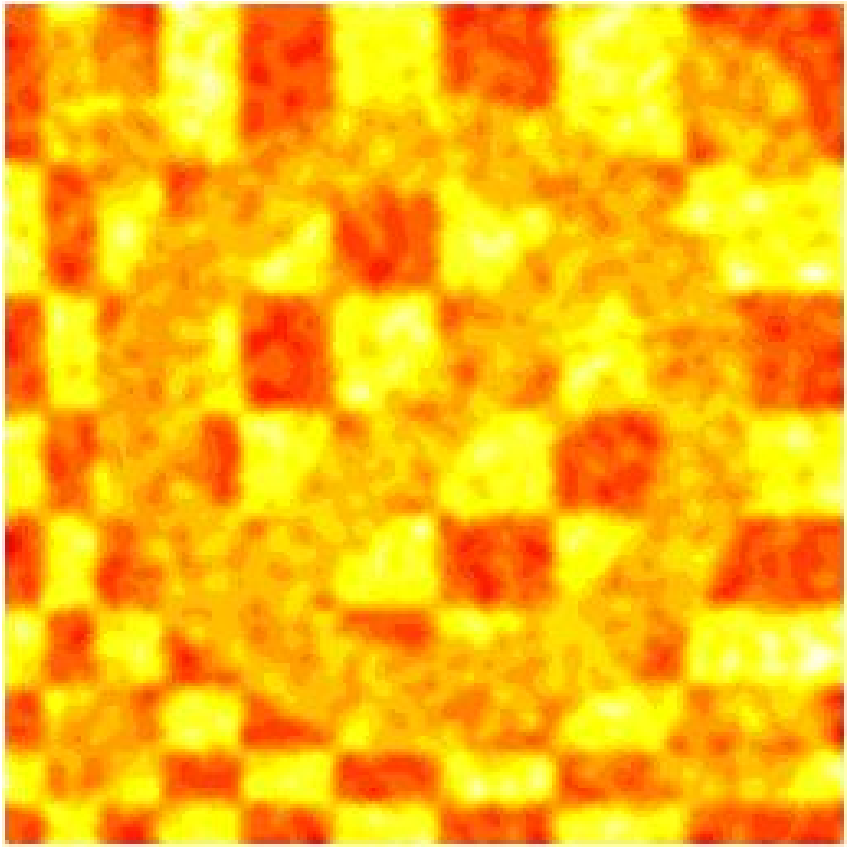
Regression: Image Analysis

Regression: Image Analysis

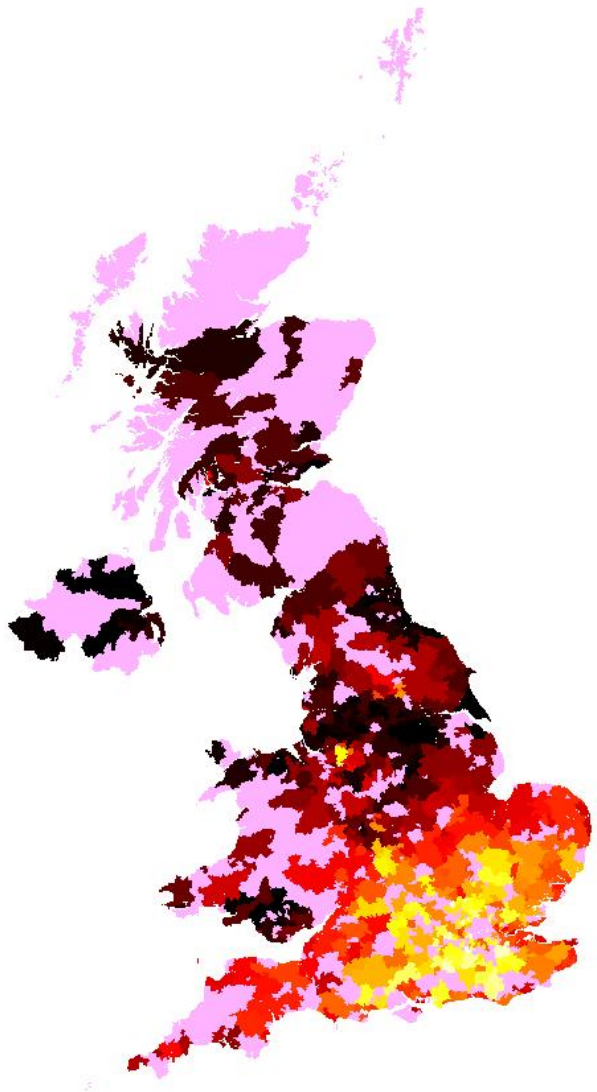


Polzehl and Spokoiny, 2000, JRSSB 62, p355–54.

Regression: Kernel Smoothing

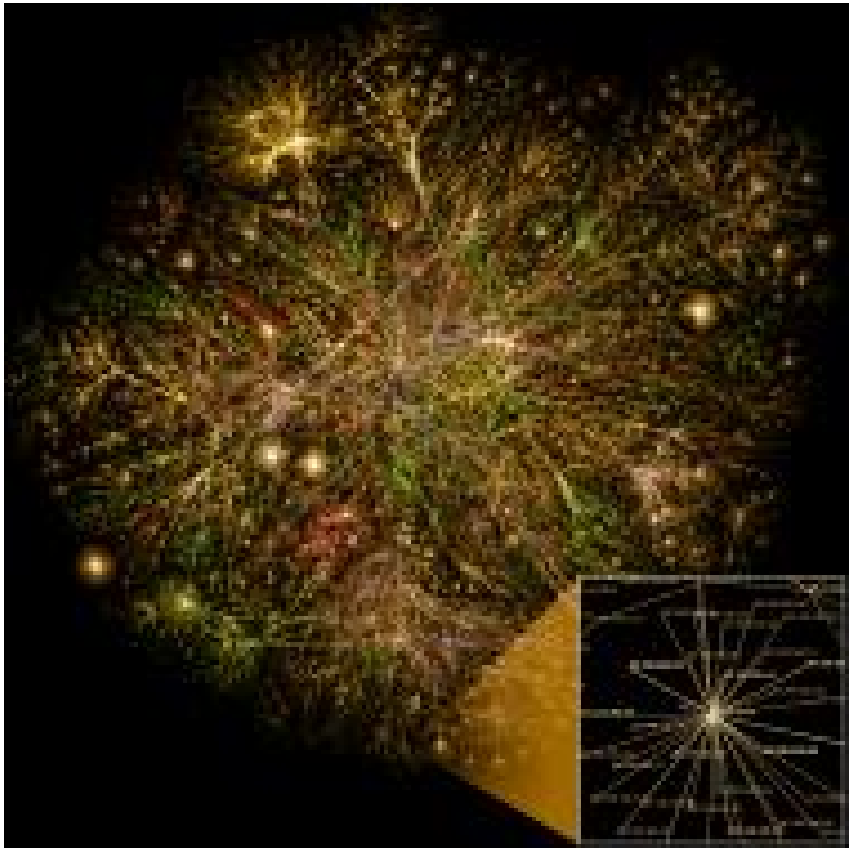


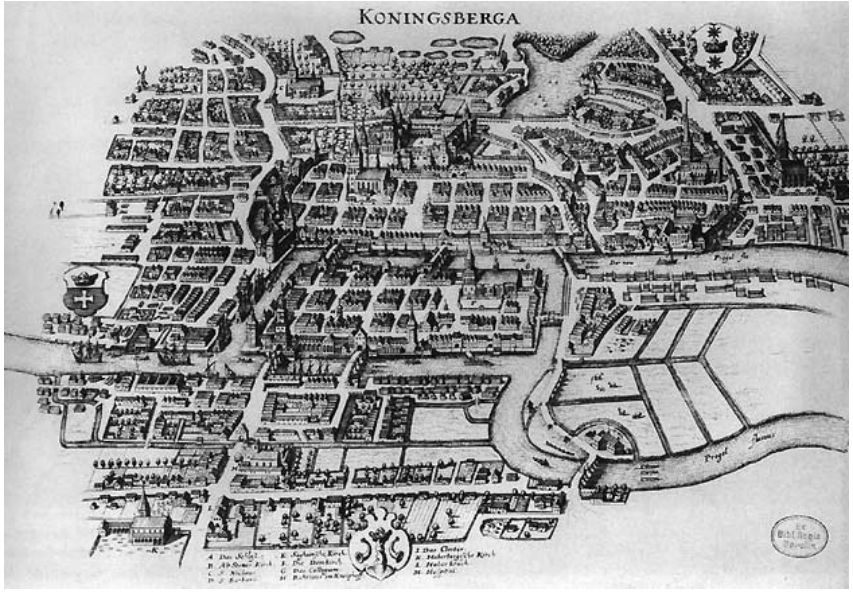
Regression: UK House Prices



Source: Halifax

Graphs





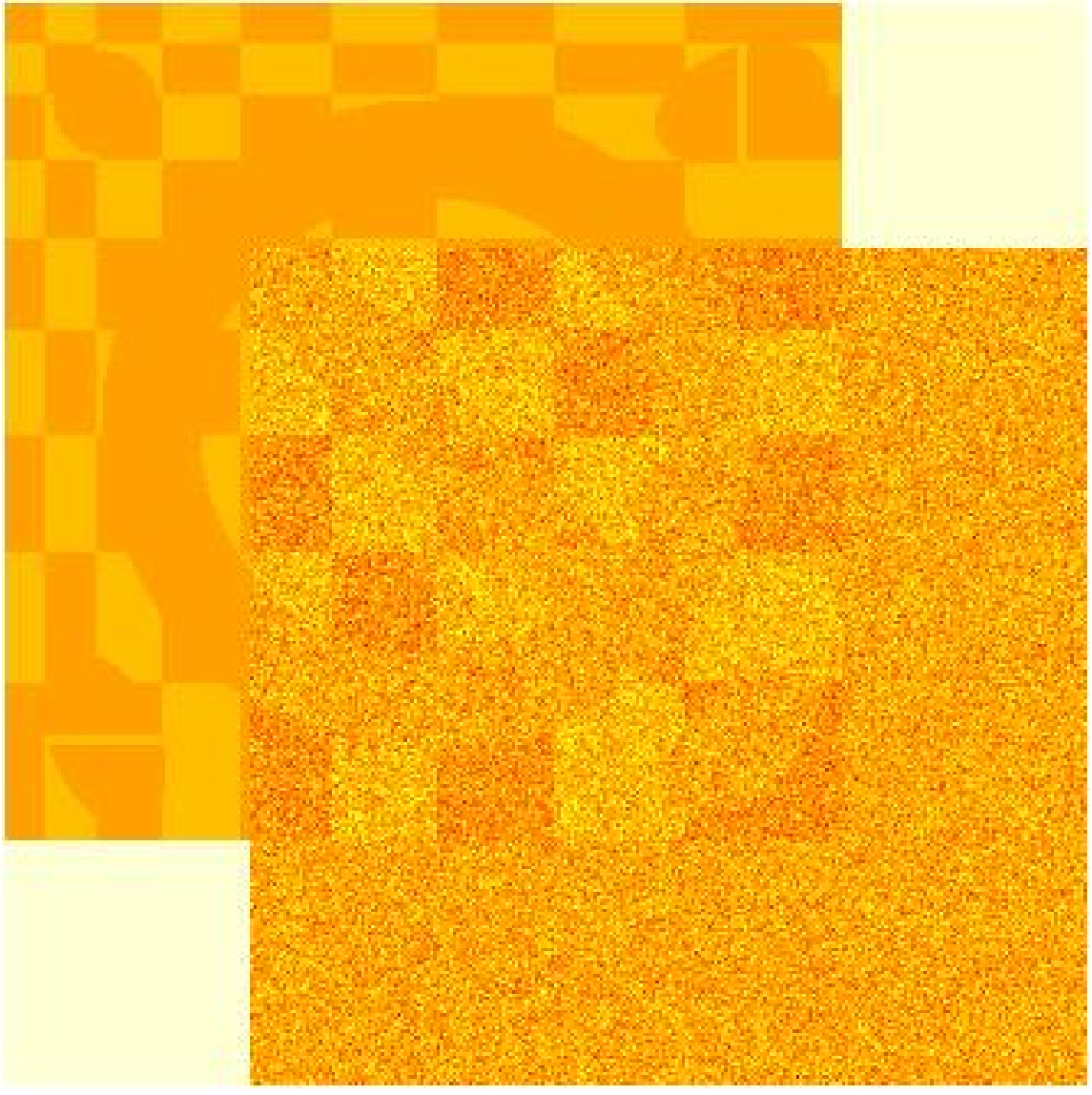
$(\mathcal{V}, \mathcal{E})$

Graphs in Statistics

Graphical structures can be found in . . .

- Scatterplot smoothing
- Image analysis
- Disease risk mapping
- Discrete spatial variation
- Longitudinal data
- . . .

Data = Signal + Noise



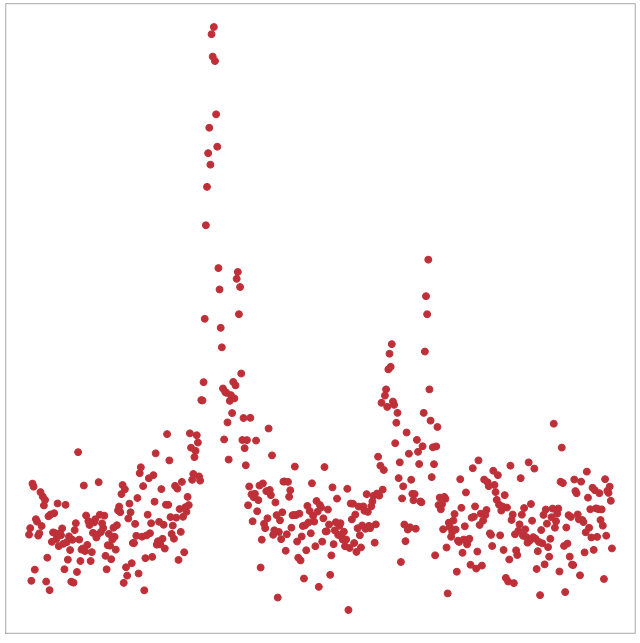
Model for Regression on a Graph

Suppose the observations come from a graph $(\mathcal{V}, \mathcal{E})$.
Given response observations y we estimate the signal function f that best approximates the data, according to

$$\begin{array}{rcccl} \text{Data} & = & \text{Signal} & + & \text{Noise} \\ y_i & = & f_i & + & \sigma z_i, \quad i \in \mathcal{V} \end{array}$$

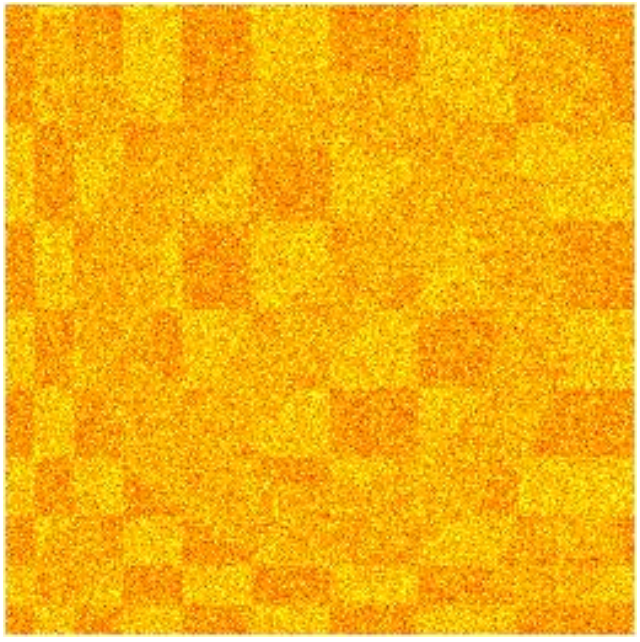
- An observation at every vertex
- Edges tell us which observations are close together

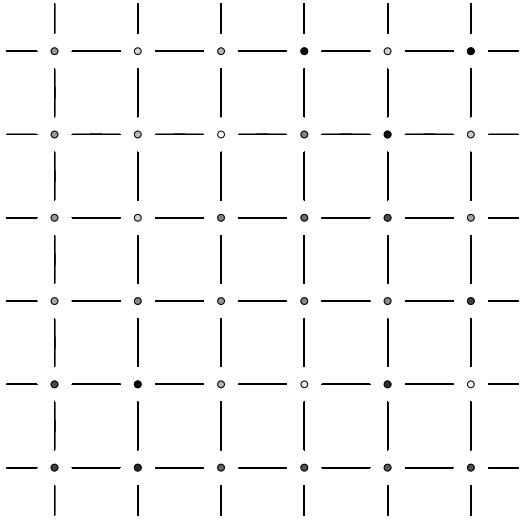
Scatterplot Smoothing Graph



$$\mathcal{V} = \{1, \ldots, n\} \quad \mathcal{E} = \{\{1,2\}, \{2,3\}, \ldots, \{n-1, n\}\}$$

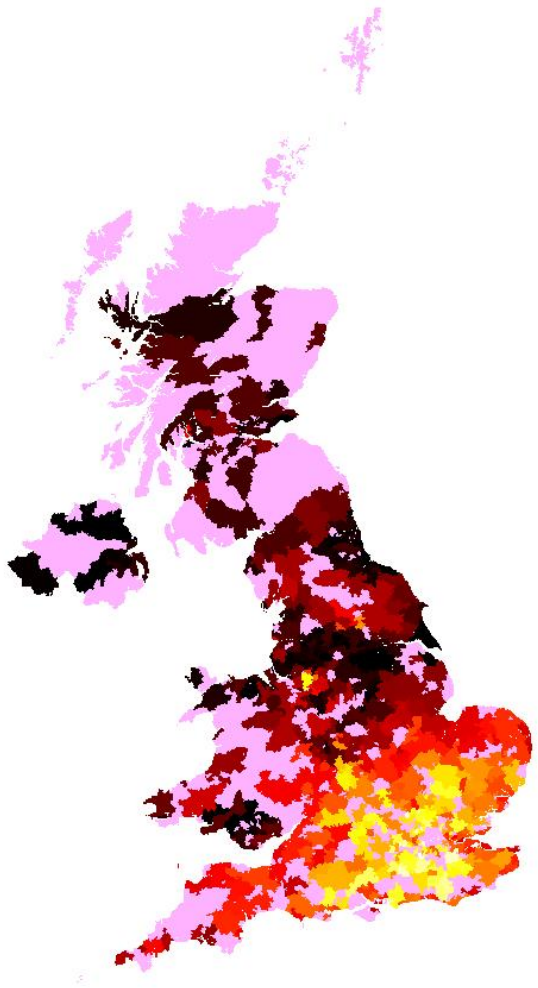
Image Analysis Graph





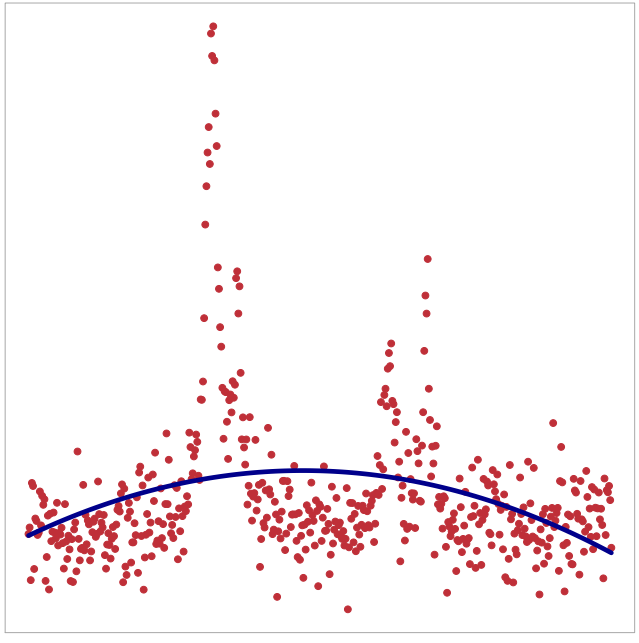
$\mathcal{V} =$ pixels $\quad \mathcal{E} =$ borders between pixels

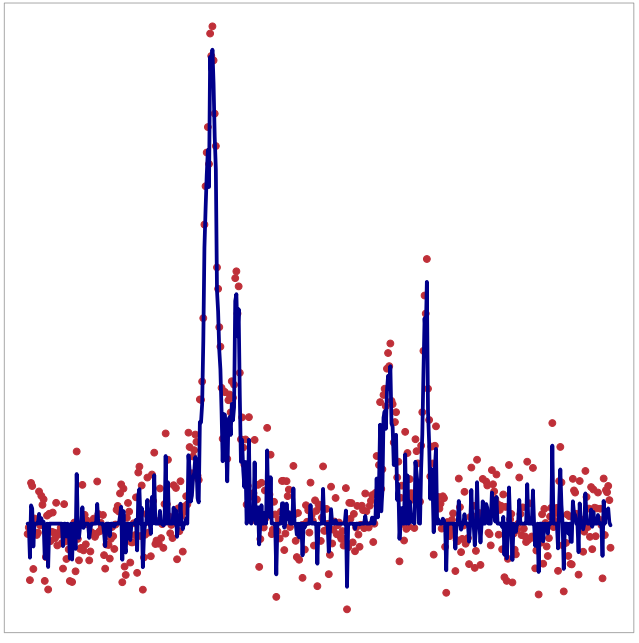
UK House Price Graph



- $\mathcal{V} =$ post towns
- $\mathcal{E} =$ neighbouring towns

Regression Tradeoff





Too far from data $\Leftrightarrow$ Too rough

Regression Tradeoff

Too far from data $\Leftrightarrow$ Too rough

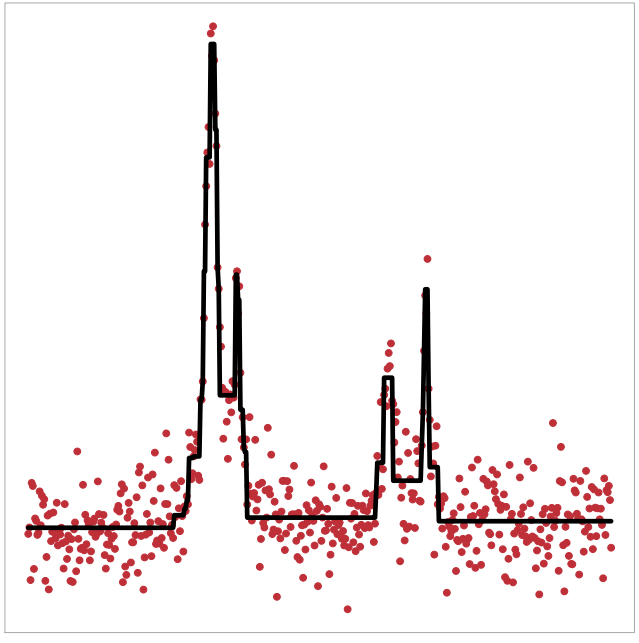
- Measure distance from data at the vertices
- Measure roughness at the edges of the graph

Scatterplot Smoothing

- Measure distance from data at the vertices
- Measure roughness at the edges of the graph

Minimise

$$\text{Distance} + \lambda \times \text{Roughness}$$



- Might have to use a new algorithm

Image Analysis

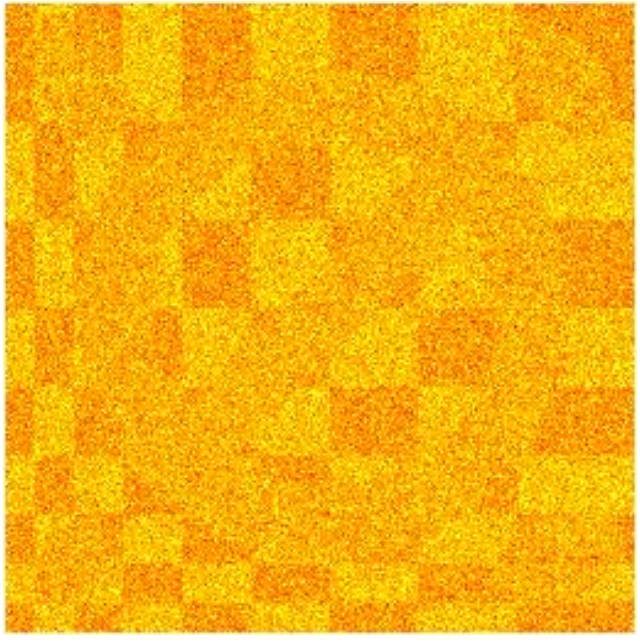
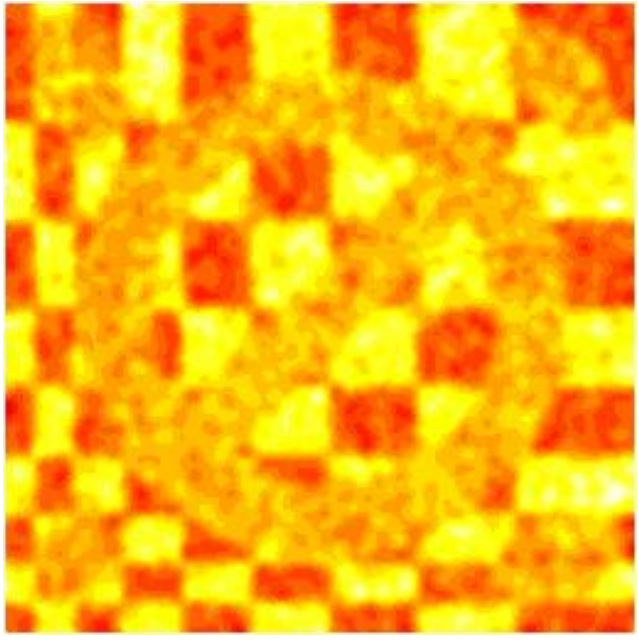
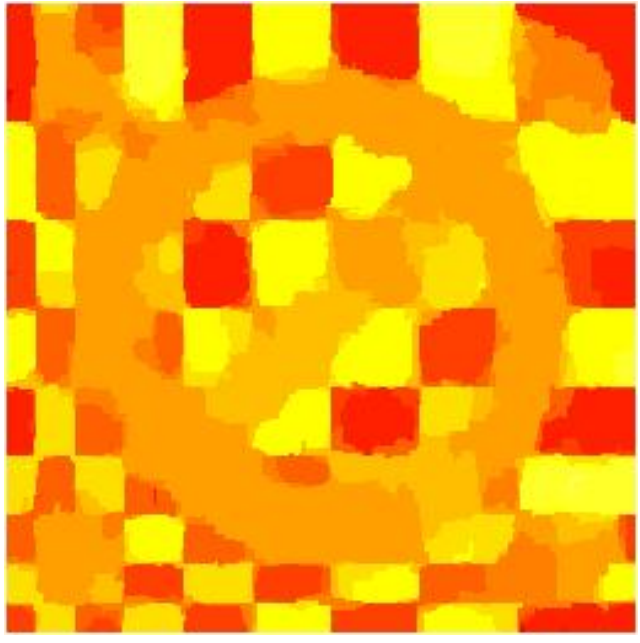


Image Analysis



Kernel smoothing



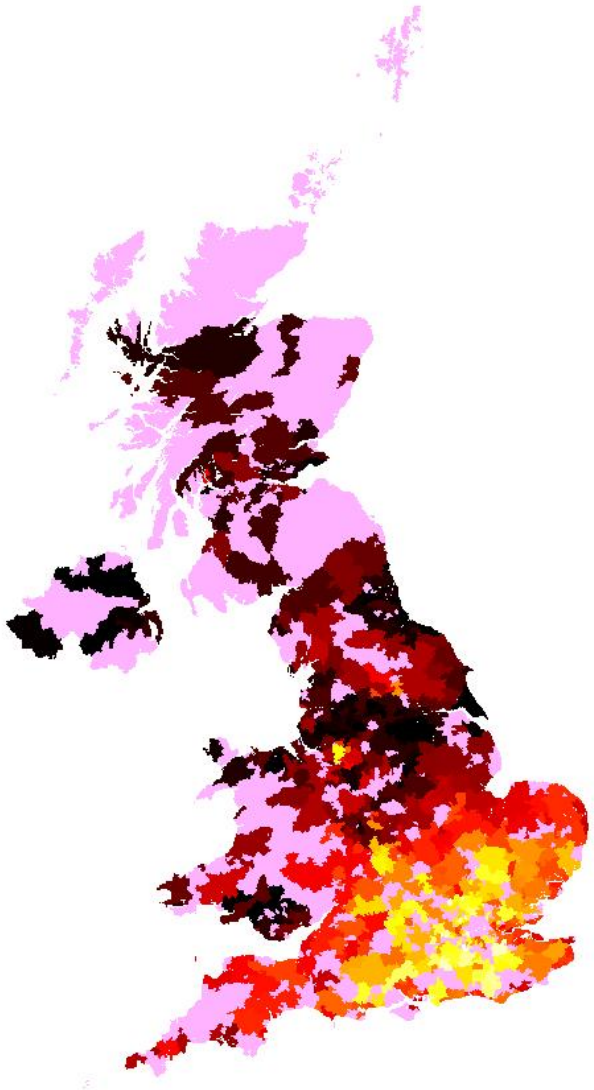
Penalised regression

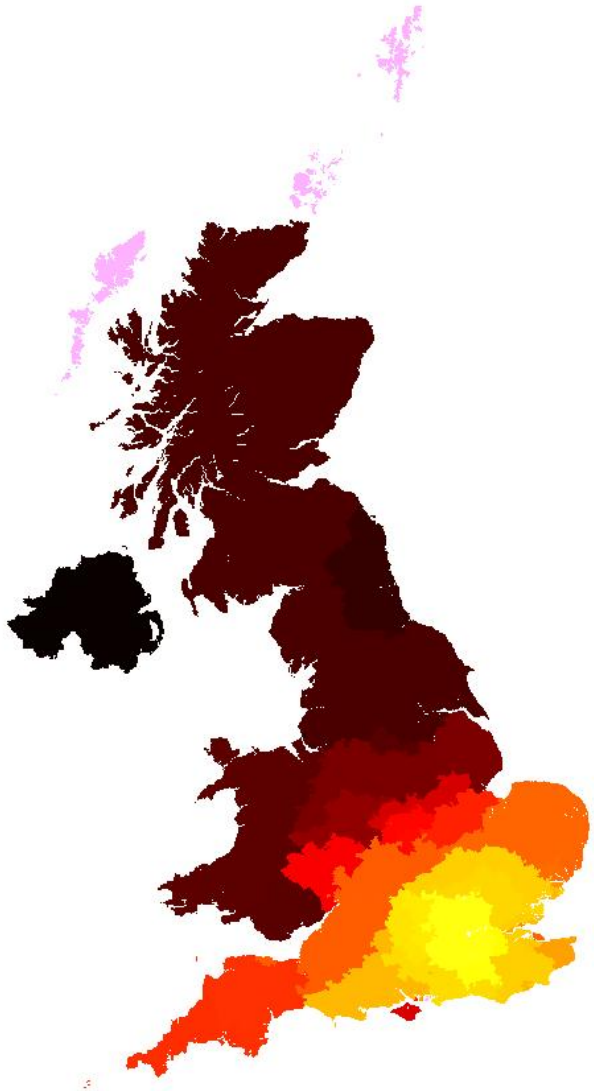
Image Analysis

Kernel smoothing

Penalised regression

UK House Prices





Summary

- Many problems in regression have a graphical interpretation.
- Measure distance from data and roughness on the graph. This is called penalised regression.
- We have developed a fast algorithm for producing these fits.