

Geographical Spatial Analysis in Personal Lines Territorial Ratemaking

2004 CAS Seminar on
Ratemaking

Duncan Anderson MA FIA

Watson Wyatt LLP



WWW.WATSONWYATT.COM/INSURANCE



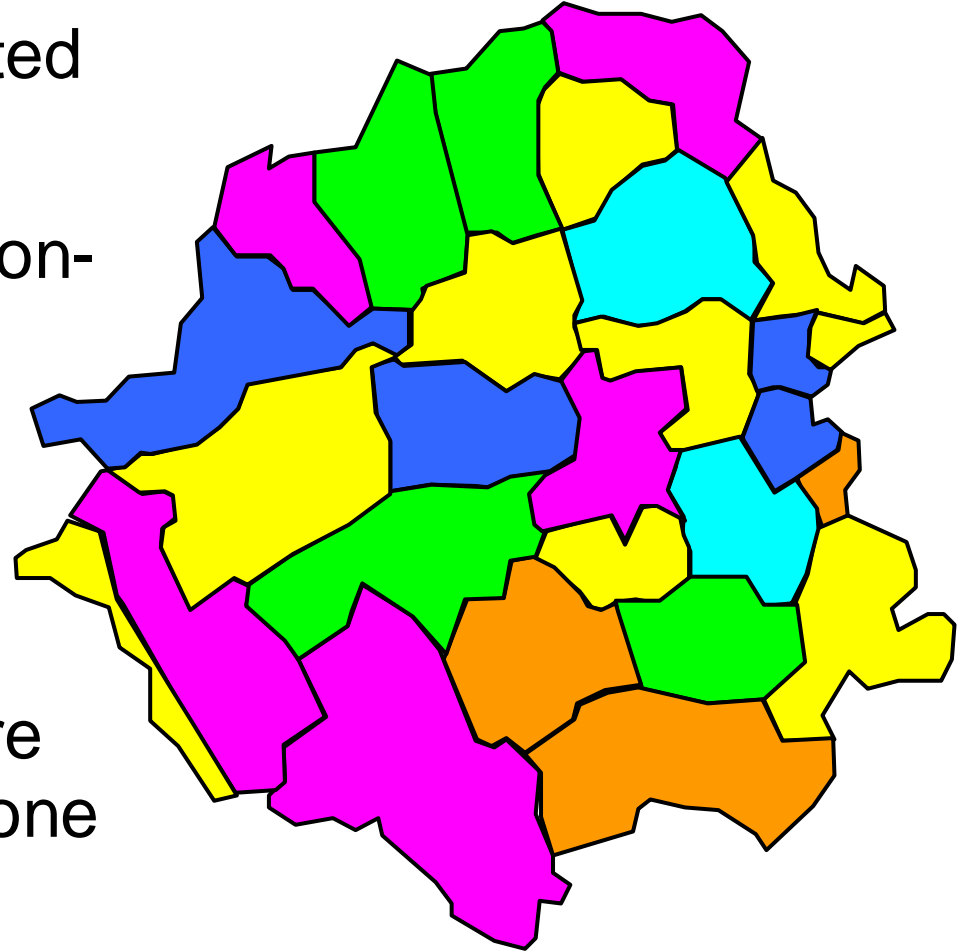


Geographical zoning

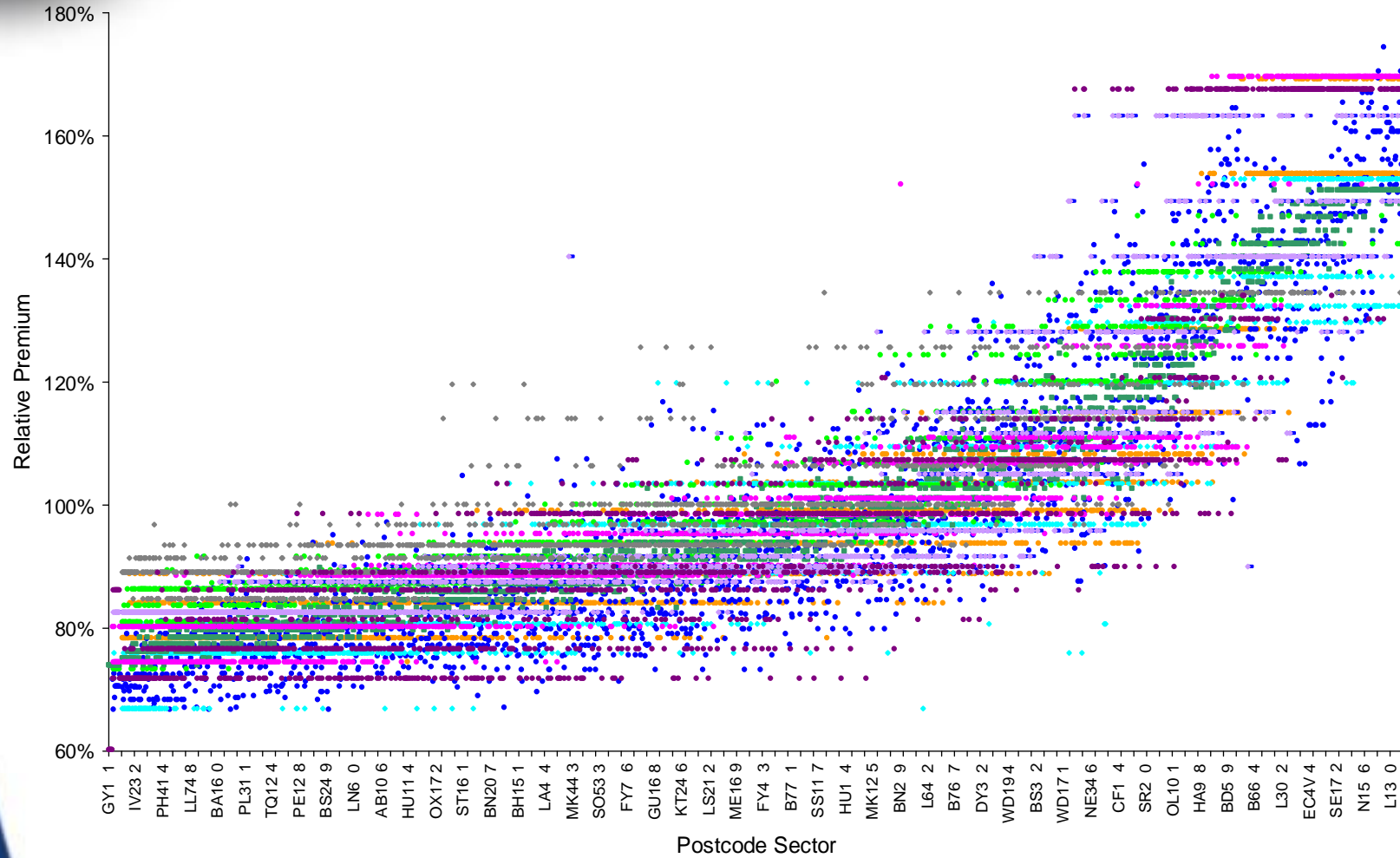
- Area is one of the main drivers of cost
- Many markets show considerable variety between insurers
- One insurer will have limited exposure in any one narrowly-defined area (eg zip code)
- Agenda
 - quick look at example of market differences
 - discussion of a method

Geographical "zones"

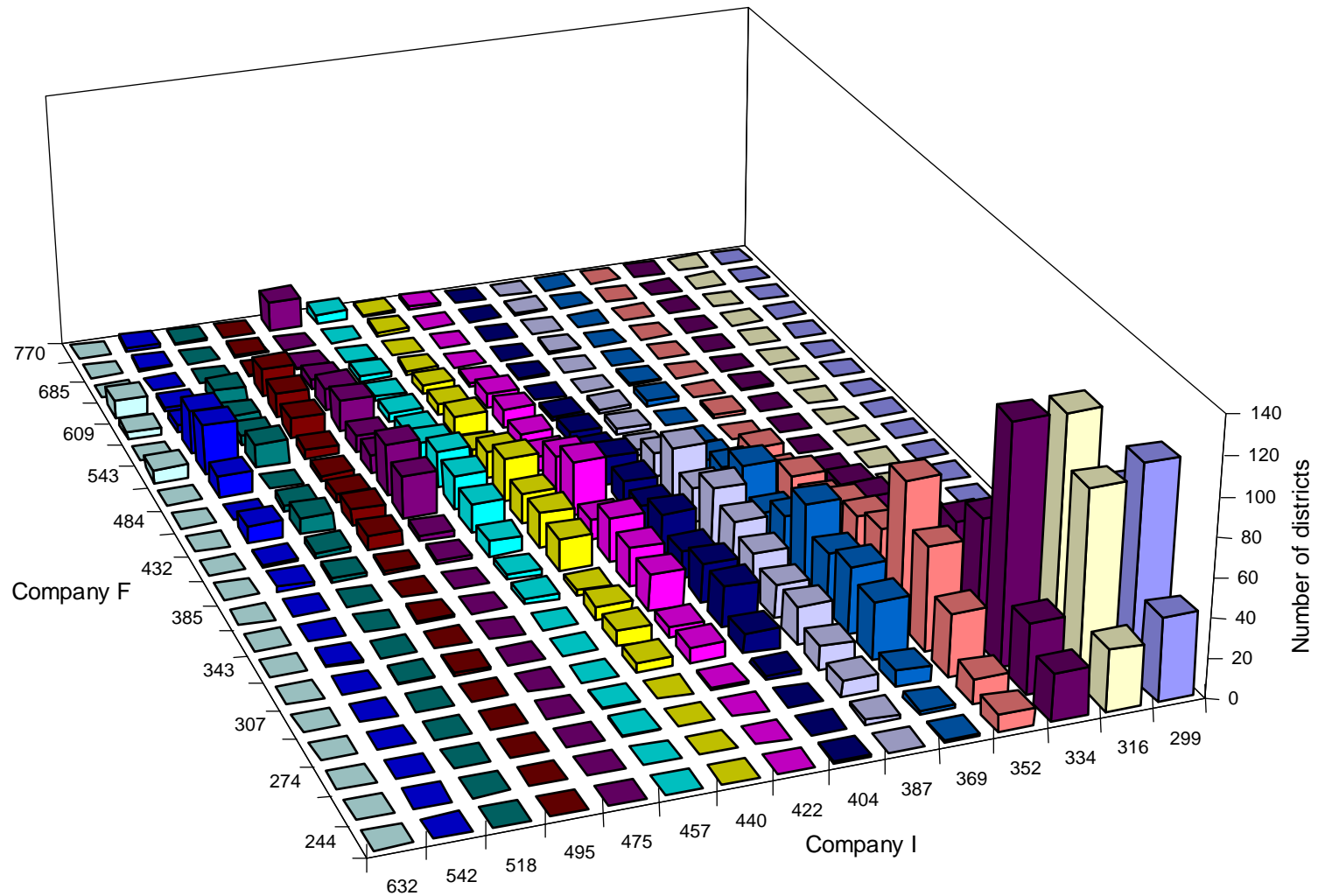
- Each zip code allocated to a zone
- Zones may contain non-contiguous zip codes
- Each zone given a premium loading
- In UK auto & homeowners there are typically around 20 zone categories, although sometimes 100+



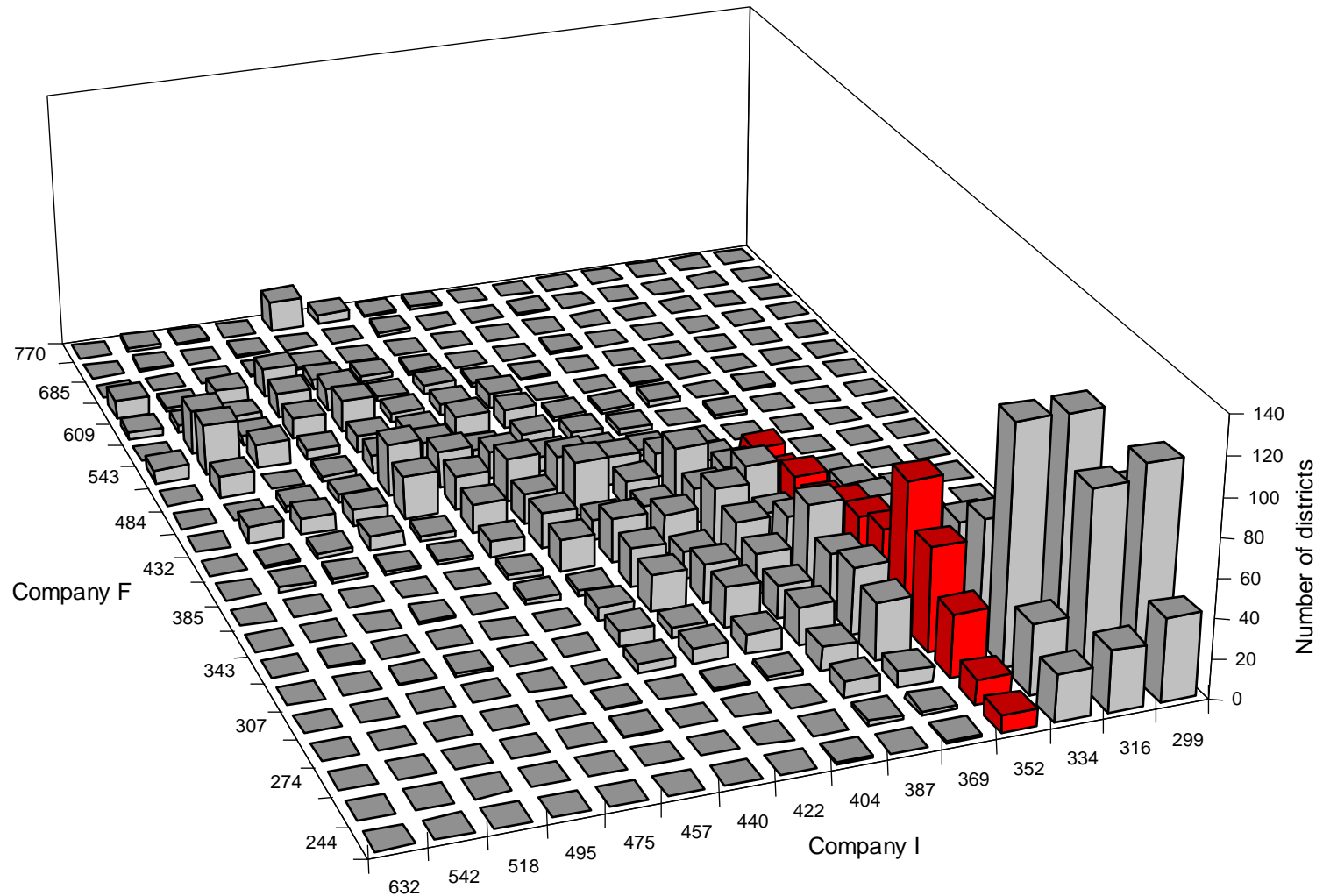
UK auto postcode loadings



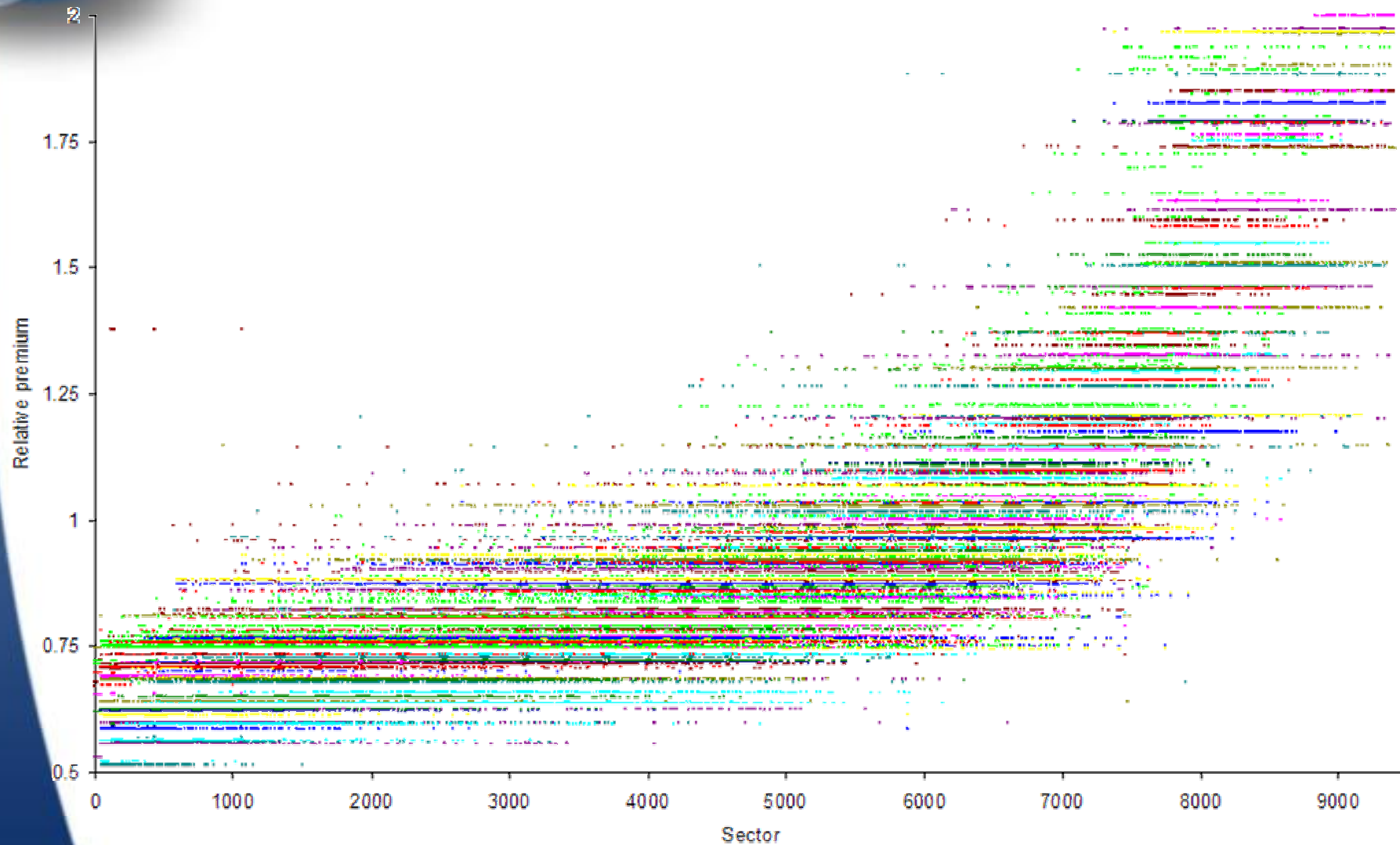
Comparison of UK auto categorizations



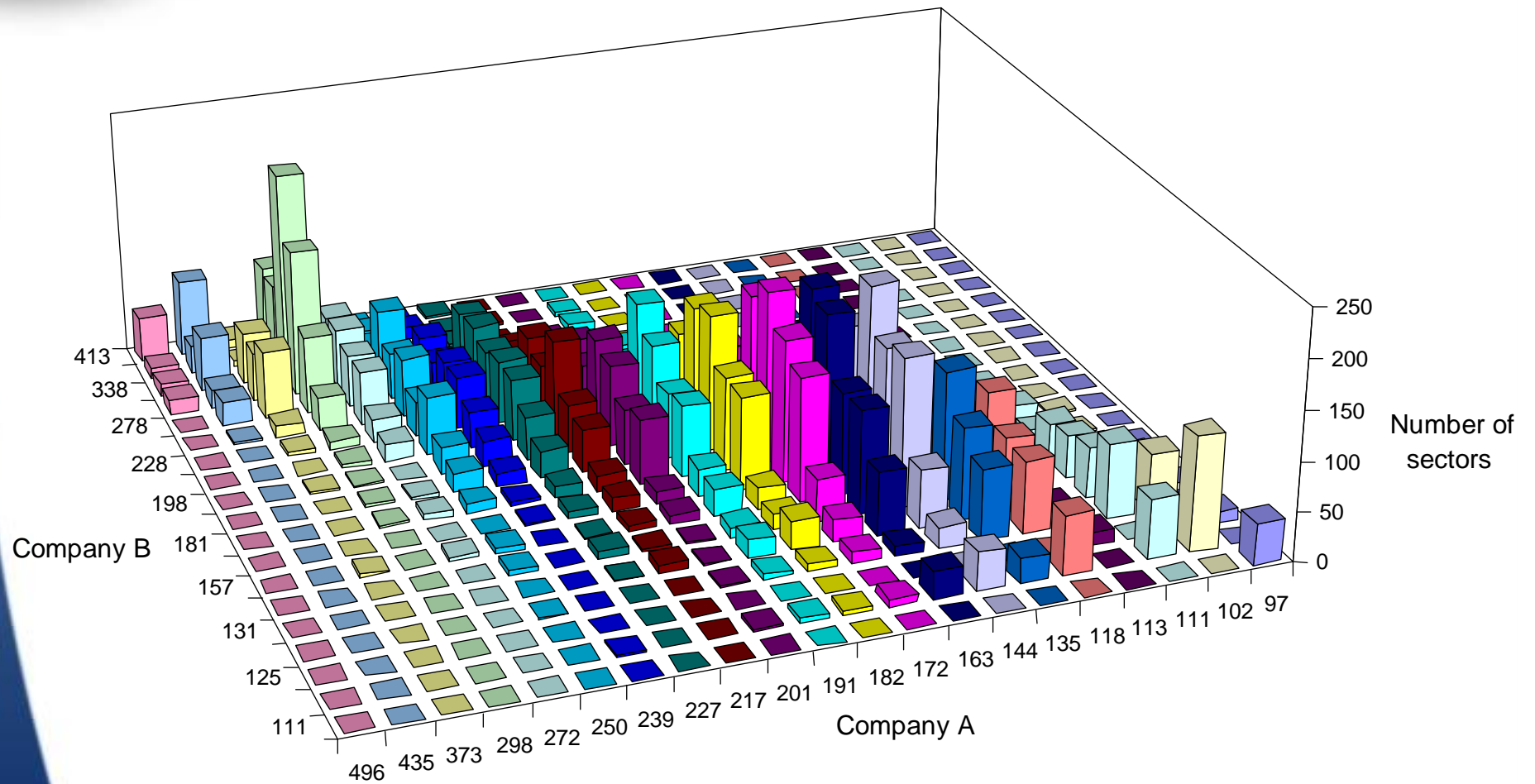
Comparison of UK auto categorizations



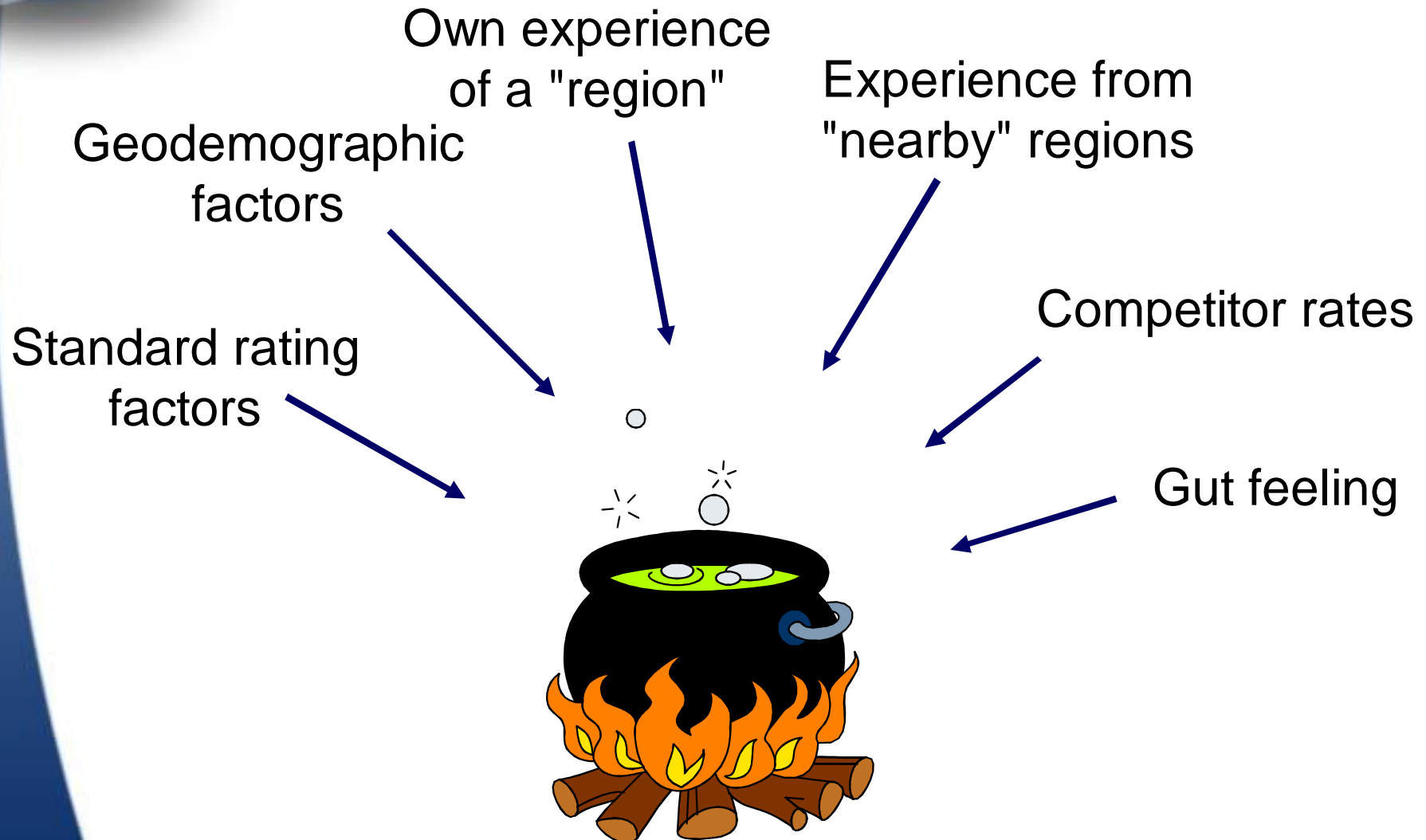
UK homeowners (contents) postcode loadings



Comparison of UK homeowners contents zones

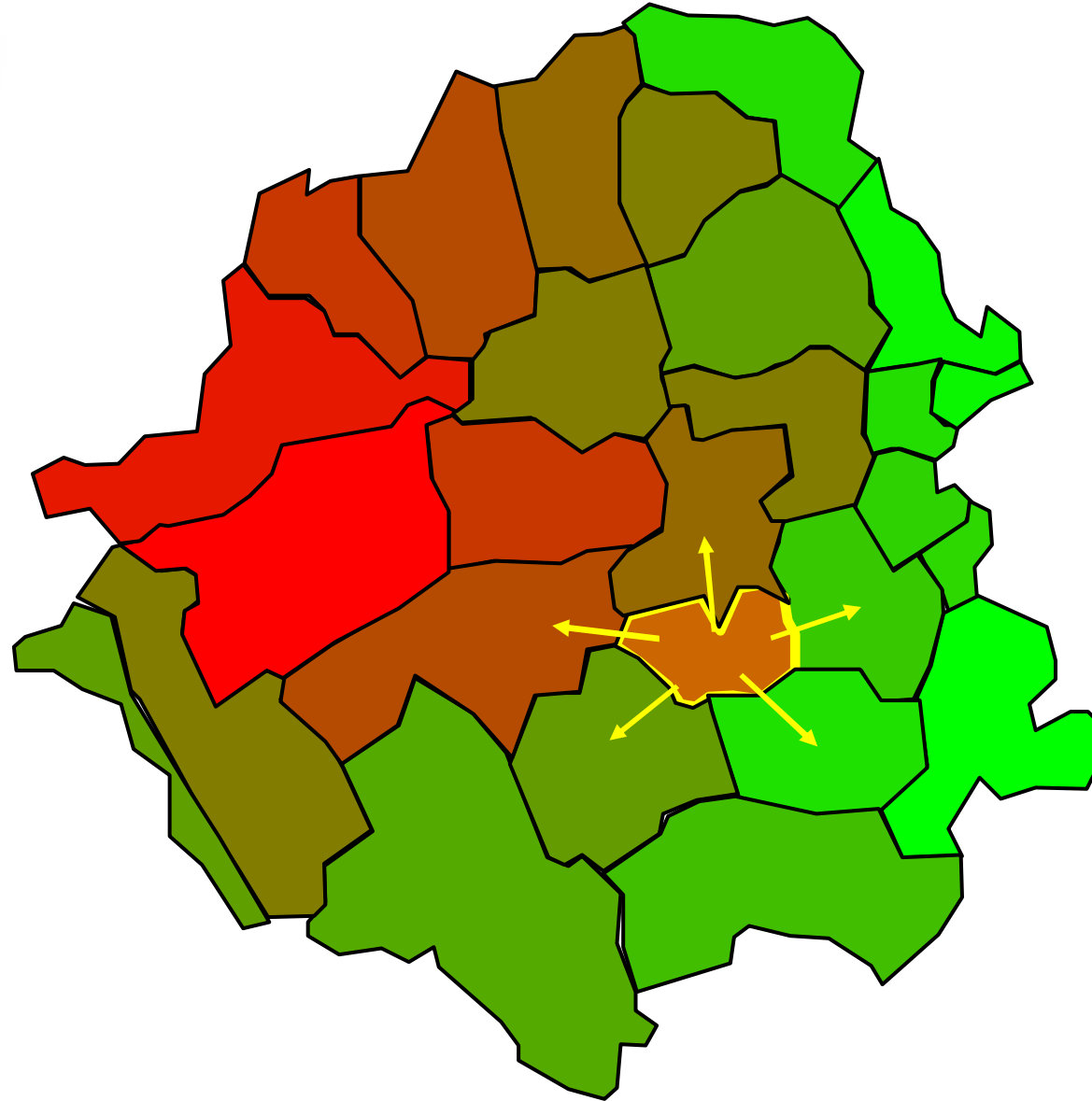


Ingredients for a solution





Proximity



The general approach

- Select which element of experience to model

BI Freq x Amt = Cost 1

PD Freq x Amt = Cost 2

MED Freq x Amt = Cost 3

COL Freq x Amt = Cost 4

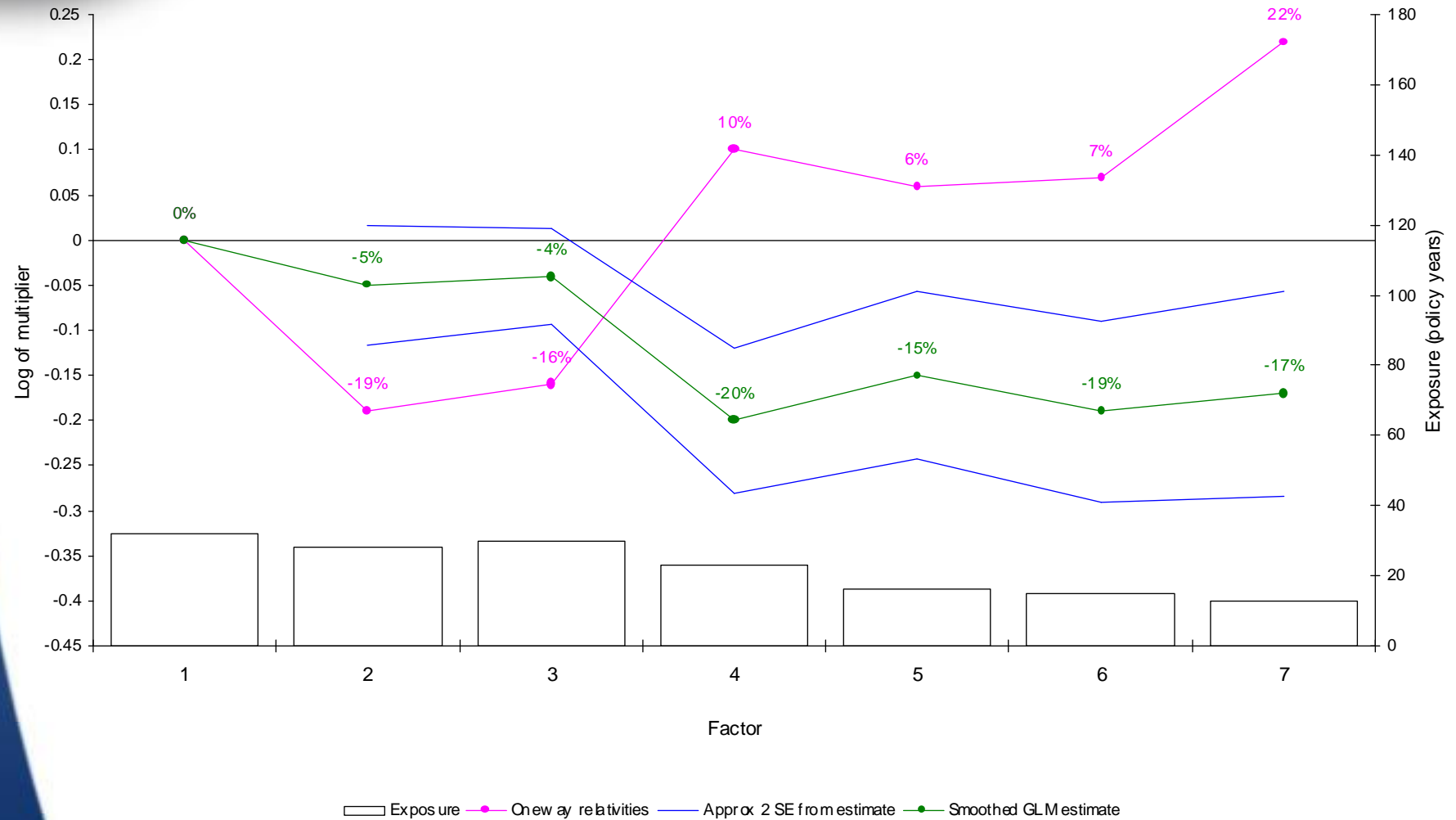
OTC Freq x Amt = Cost 5



The general approach

- Do not wish to attribute to any region experience which can be explained by other rating factors
- Standardize for other factors by fitting a GLM (excluding current zones)

Generalized linear models





The general approach

- Do not wish to attribute to any region experience which can be explained by other rating factors
- Standardize for other factors by fitting a GLM (excluding current zones)
- Consider "residual" risk by "region"
- Seek to make this residual risk more predictive
- Then categorize into zones to derive appropriate loadings



A model form

$$r_i^* = Z \cdot r_i + (1 - Z) \cdot \text{neighboring experience}$$

where

r_i^* = smoothed residual risk

r_i = unsmoothed residual risk

Z = credibility function



A model form

$$r_i^* = Z \cdot r_i + (1 - Z) \cdot \text{neighboring experience}$$

where

r_i^* = smoothed residual risk

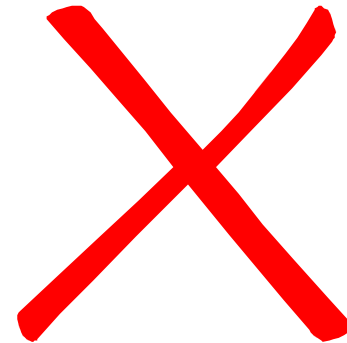
r_i = unsmoothed residual risk

Z = credibility function



What is the "residual"?

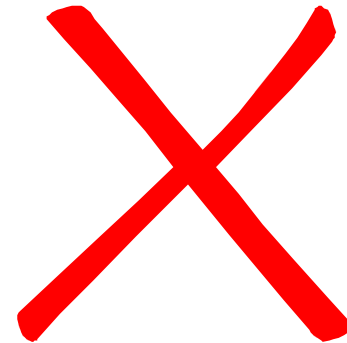
- Ideally measure "residual" by fitting region as a GLM factor with thousands of levels
- One region with no claims will cause GLM not to converge
- If model does converge, definition of region probably too large for these techniques





A / E ?

- Theoretically aligned with multiplicative GLM
- If there are no claims ($A=0$), the residual is zero regardless of the value of E , thus losing information
- If there is so much data that A is never 0, you can probably fit a GLM using region anyway





A - E

- Works for $A=0$
- Not aligned with multiplicative GLM
- Simple, robust and easy to calculate
- Eg for claim frequency consider

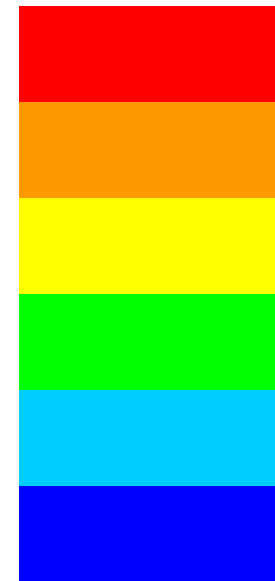
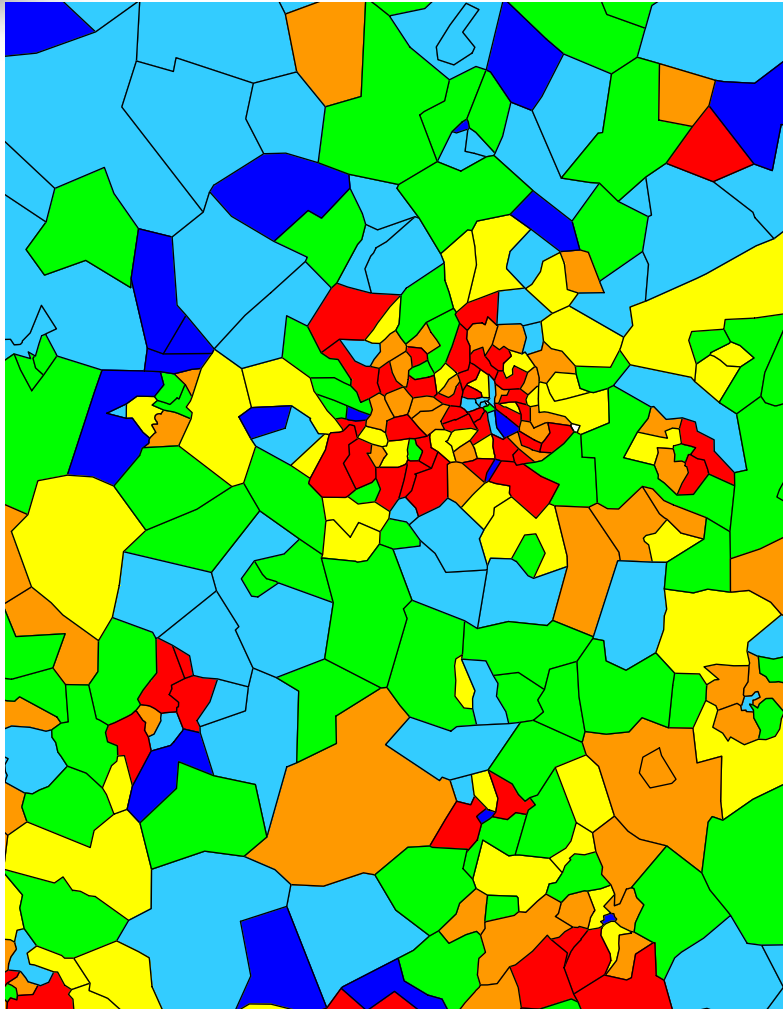
$(\text{Actual number} - \text{Expected number}) / \text{Exposure}$





Example residual risk

UK homeowners contents theft frequency



High residual



Low (negative)
residual



A model form

$$r_i^* = Z \cdot r_i + (1 - Z) \cdot \text{neighboring experience}$$

where

r_i^* = smoothed residual risk

r_i = unsmoothed residual risk

Z = credibility function

$$Z(e_i) = \left\{ \frac{e_i}{e_i + a} \right\}^m, \quad e_i = \text{exposure in region } i$$



A model form

$$r_i^* = Z \cdot r_i + (1 - Z) \cdot \text{neighboring experience}$$

where

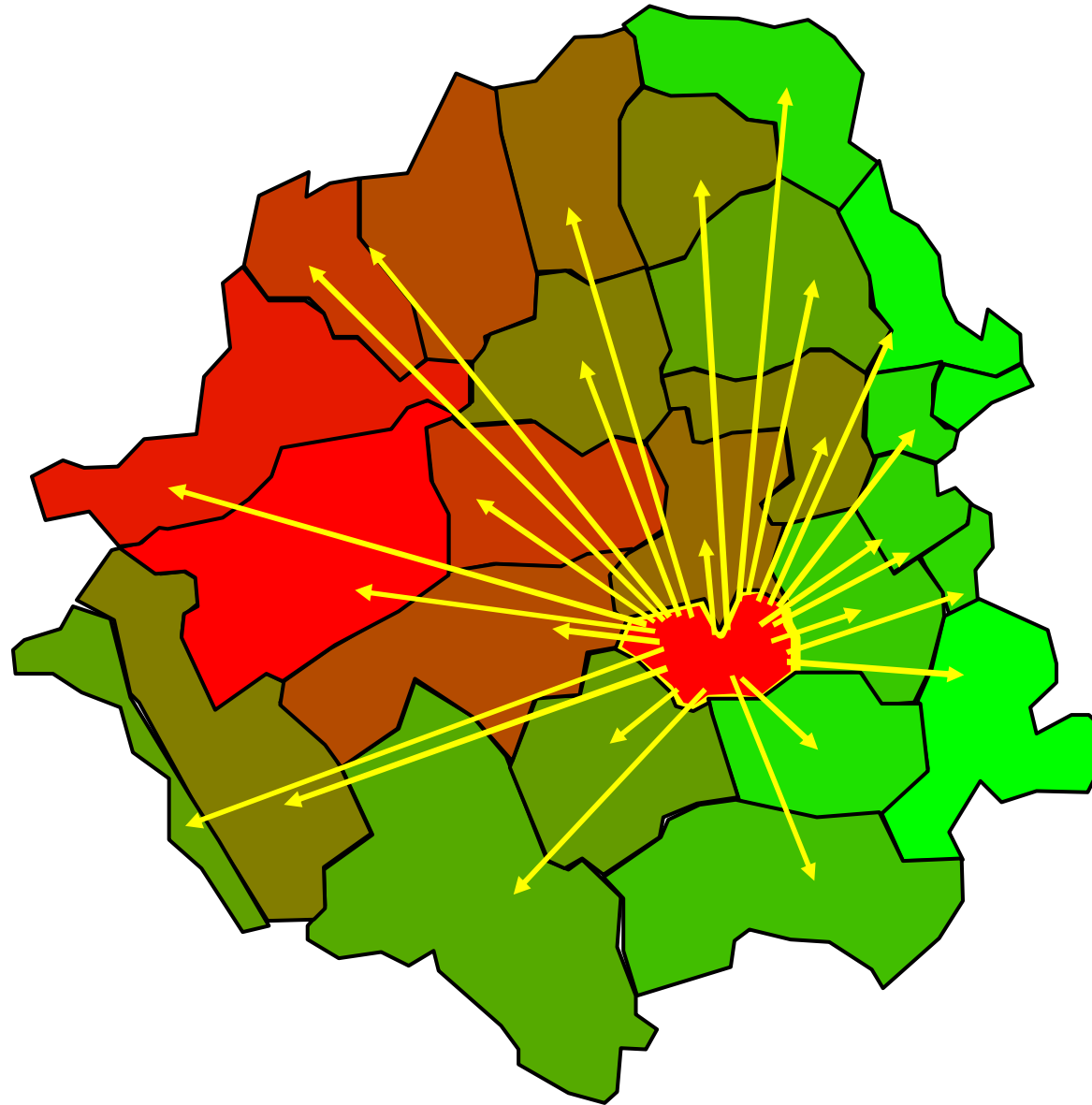
r_i^* = smoothed residual risk

r_i = unsmoothed residual risk

Z = credibility function

$$Z(e_i) = \left\{ \frac{e_i}{e_i + a} \right\}^m, \quad e_i = \text{exposure in region } i$$

Definitions of "neighboring"





Model

$$r_i^* = Z(e_i) \cdot r_i + (1 - Z(e_i)) \frac{\sum_j e_j \cdot r_j \cdot f(d_{ij})}{\sum_j e_j \cdot f(d_{ij})}$$

where

r_i^* = smoothed residual r_i = unsmoothed residual

$Z(e_i) = \{ e_i / (e_i + a) \}^m$ e_i = exposure in region i

$$d_{ij} = \{ (x_i - x_j)^2 + (y_i - y_j)^2 \}^{1/2}$$

$f(d_{ij}) = 1/d_{ij}^n$ or $1/(d_{ij}^n + b^n)$ or $\exp(-n \cdot d_{ij})$ etc



Parameters

$$r_i^* = Z(e_i) \cdot r_i + (1 - Z(e_i)) \frac{\sum_j e_j \cdot r_j \cdot f(d_{ij})}{\sum_j e_j \cdot f(d_{ij})}$$

where

r_i^* = smoothed residual r_i = unsmoothed residual

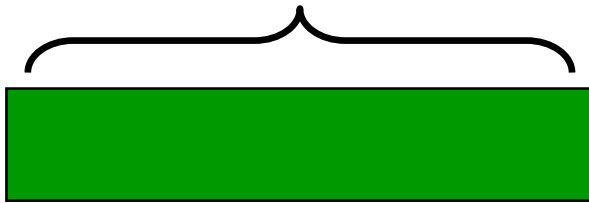
$Z(e_i) = \{ e_i / (e_i + a) \}^m$ e_i = exposure in region i

$$d_{ij} = \{ (x_i - x_j)^2 + (y_i - y_j)^2 \}^{1/2}$$

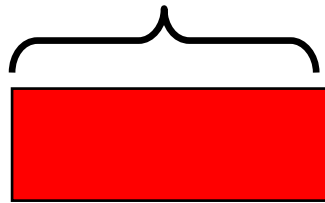
$f(d_{ij}) = 1/d_{ij}^n$ or $1/(d_{ij}^n + b^n)$ or $\exp(-n \cdot d_{ij})$ etc

Finding the parameters

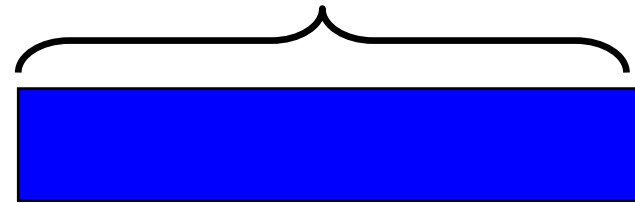
Calculate residuals



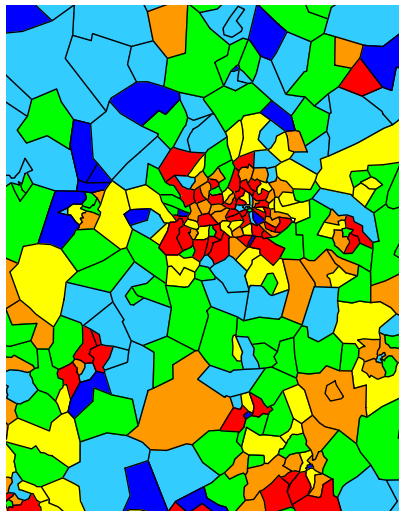
Seek parameters which minimize error



Save for determining zoning relativities

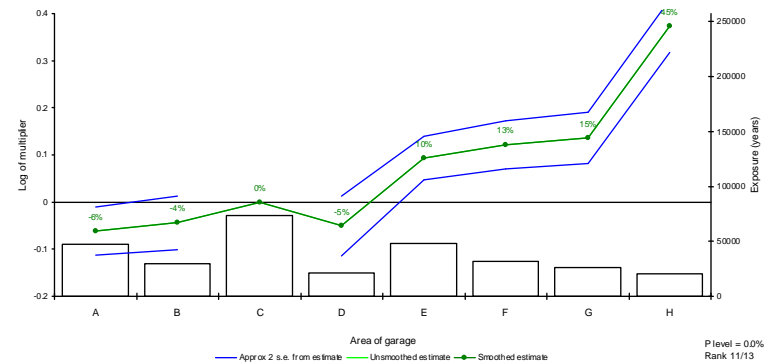


a, m, n, b



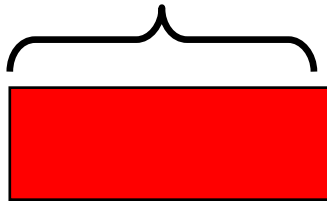
Example job

Run 2 Model 3 - All claim types, all factors, N&A - Third party material damage, Numbers



Finding the parameters

Seek parameters
which minimize
error



a, m, n, b

Error =

$$\sum (r_i^* - r_i)^2 * e_i$$

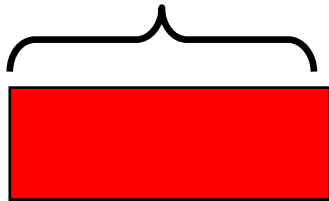
or

$$\sum \ln \{ 1 + (r_i^* - r_i)^2 \} * e_i$$

etc

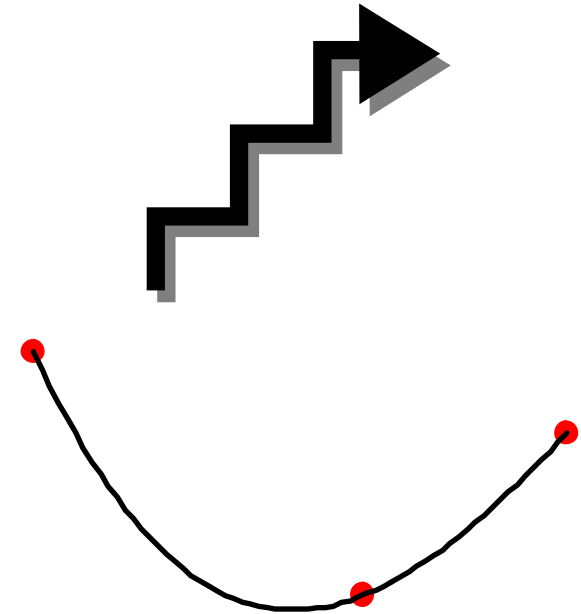
Finding the parameters

Seek parameters
which minimize
error



a, m, n, b

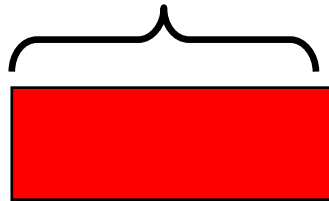
- Simple search
- Golden search
- Newton-Raphson



$$x' = x - \frac{f'(x)}{f''(x)}$$

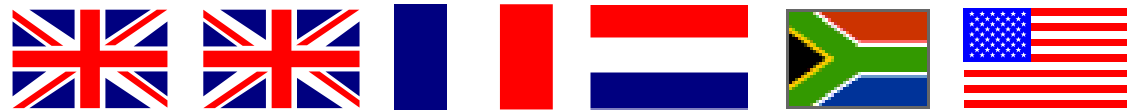
Finding the parameters

Seek parameters
which minimize
error



a, m, n, b

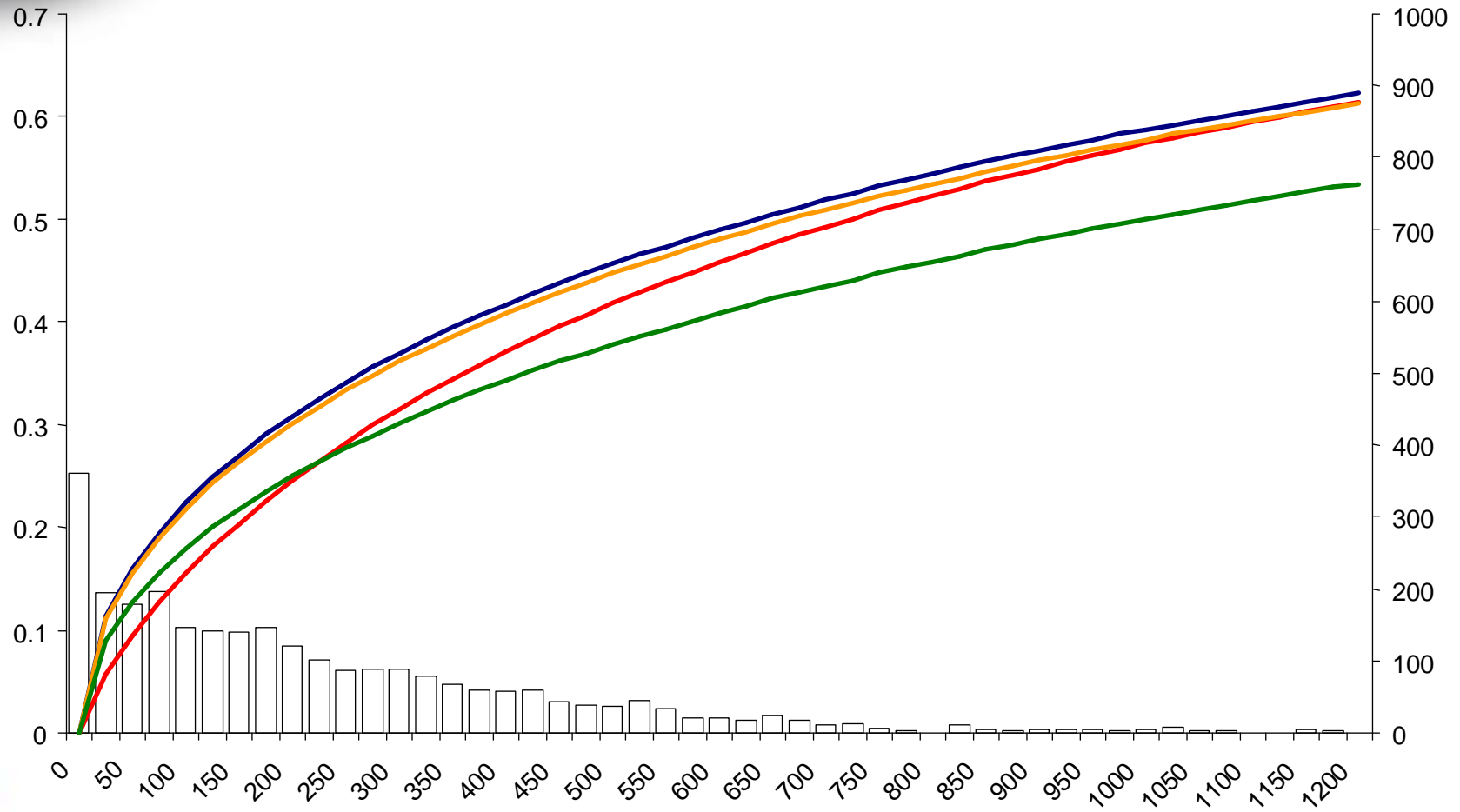
Real examples with $f(d_{ij}) = 1/d_{ij}^n$



n	1.9	2.2	2.0	1.8	2.2	Similar
a	1100	1150	2500	1800	1000	Similar
m	0.75	0.75	0.5	0.47	0.75	Similar

Credibility curves

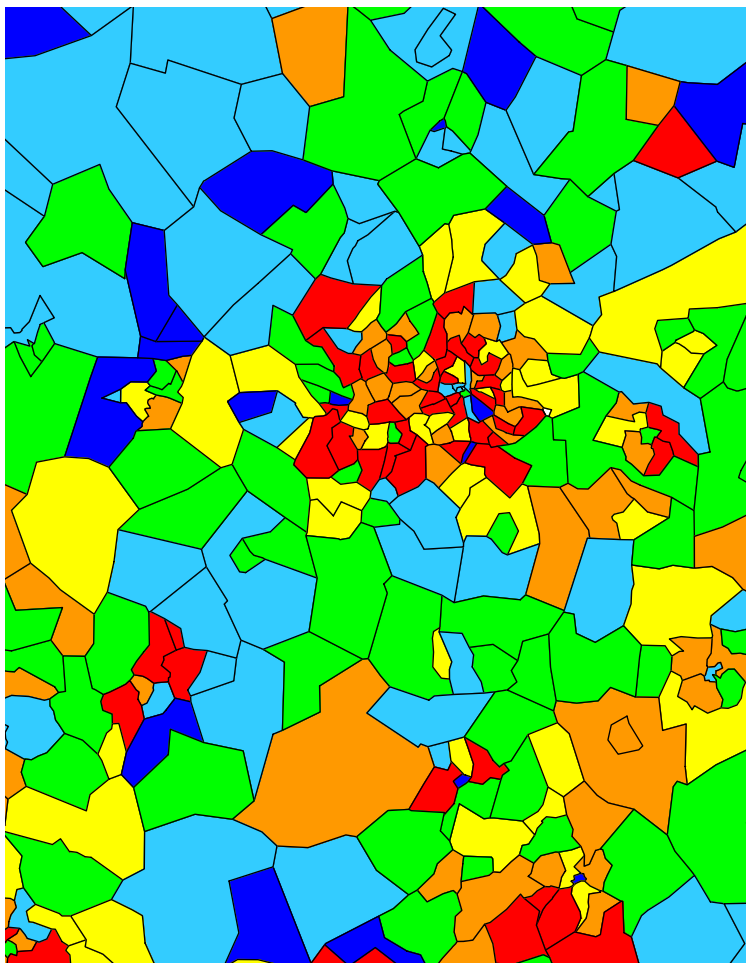
Some UK examples



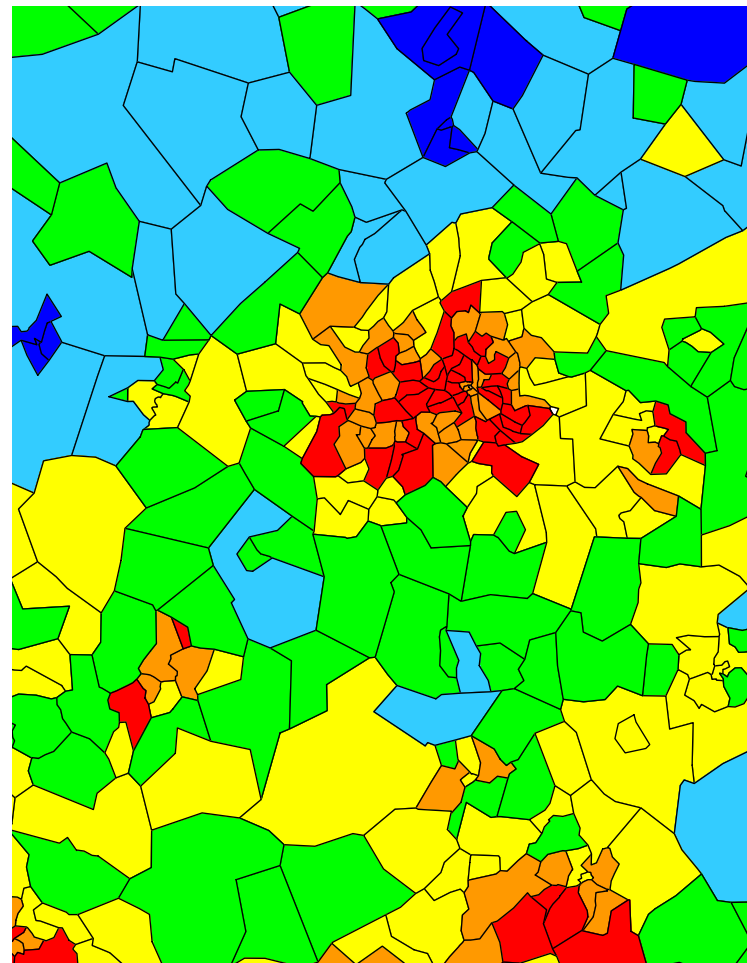


Example results

Unsmoothed residuals

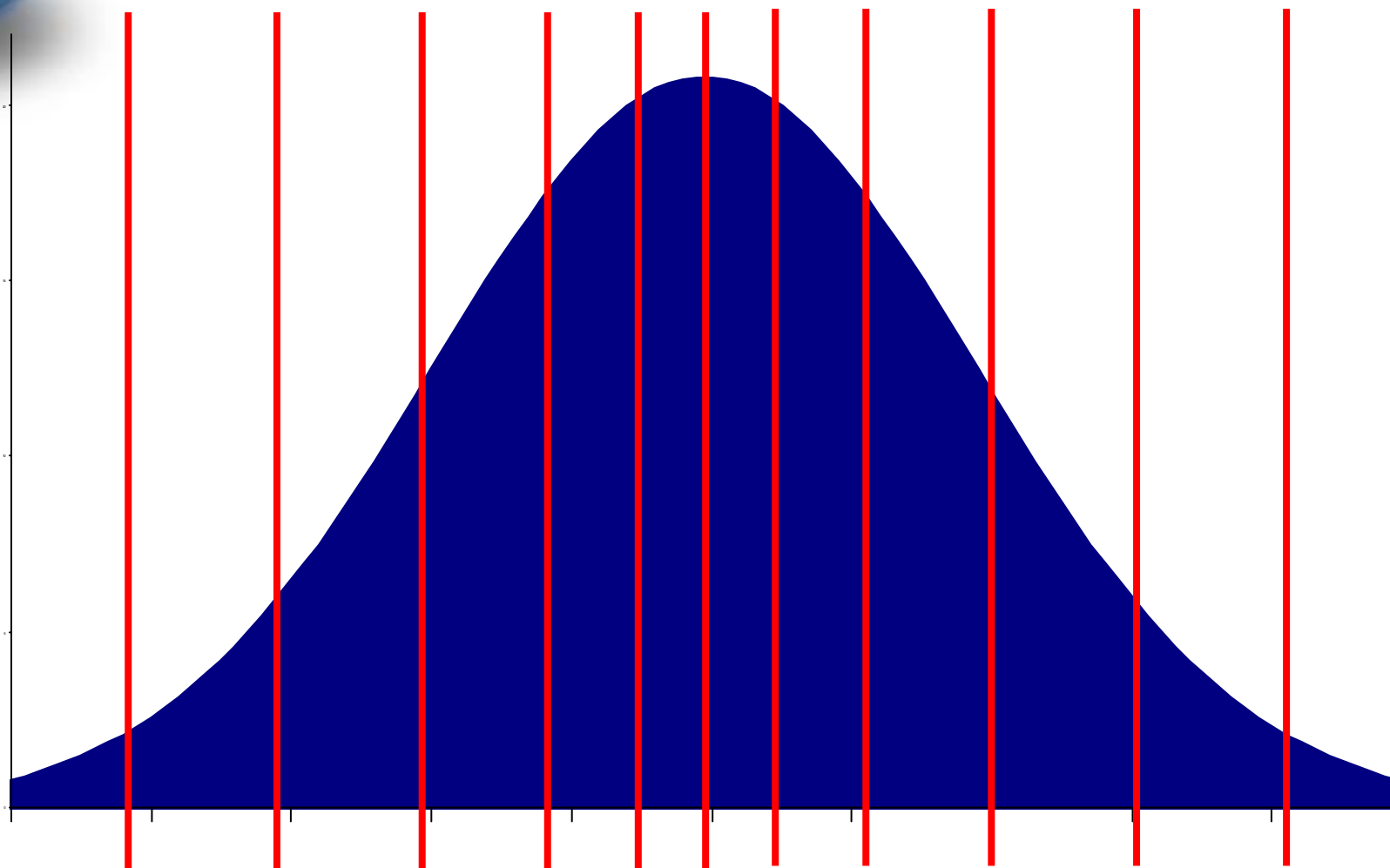


Smoothed residuals



Creating zones

Number of regions

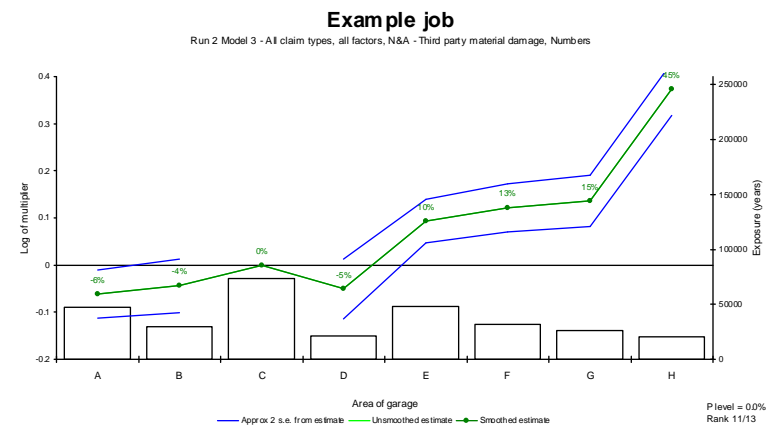
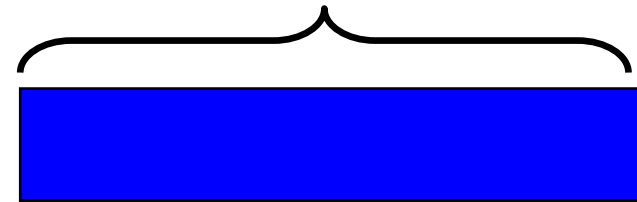


Smoothed residual

Finding the parameters

- Remodeling is necessary to "expand" the "squashed" smoothed residuals
- Fresh data required to avoid self-fulfilling prophecies

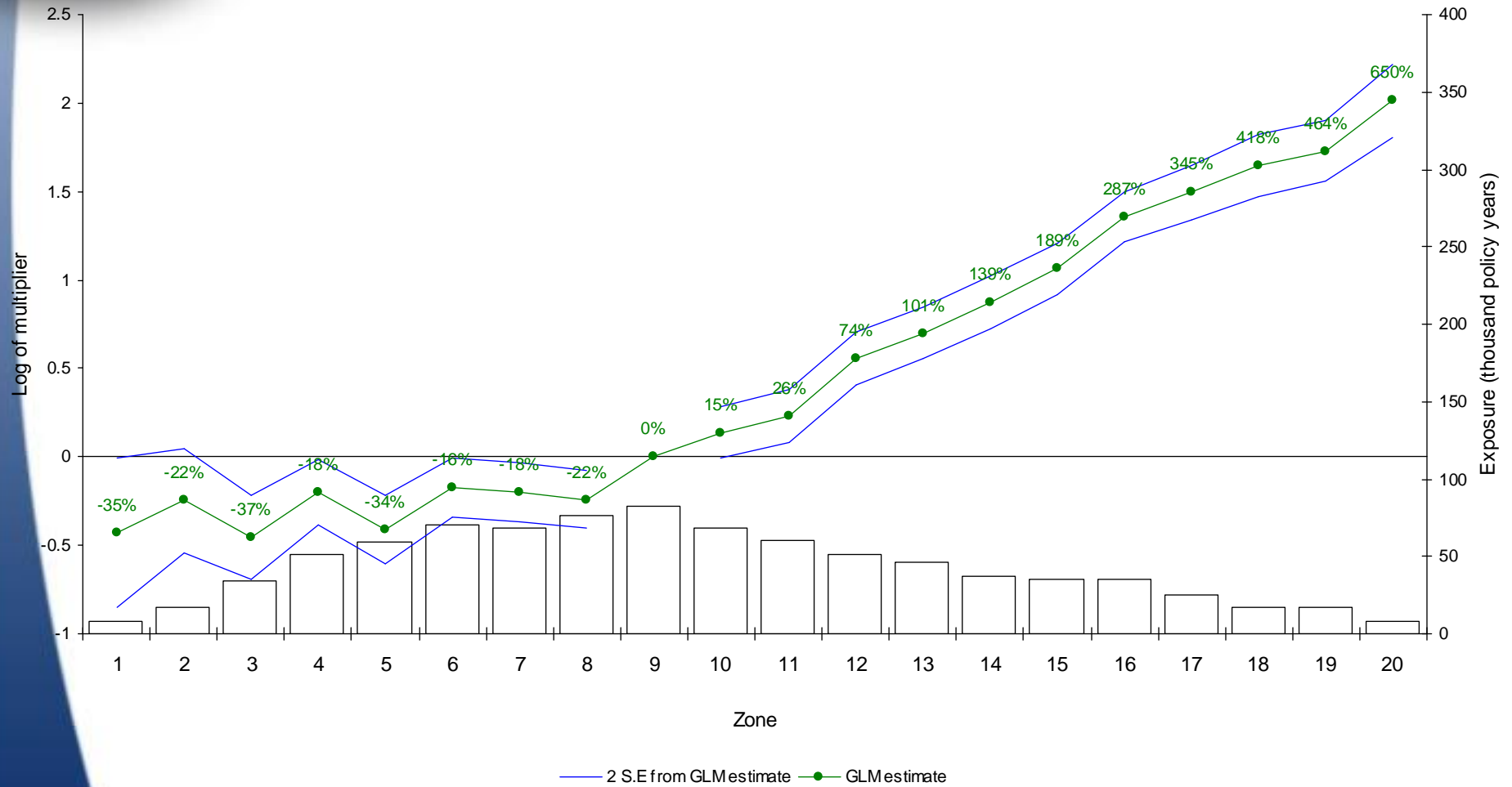
Save for determining zoning relativities



Finding the parameters

Effect of smoothed residual zone on fresh data

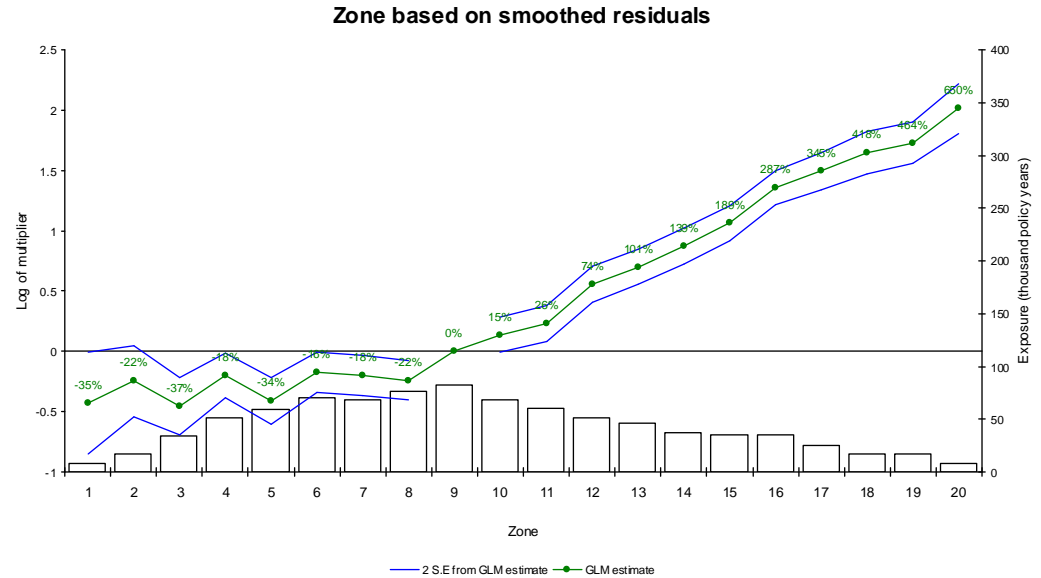
Zone based on smoothed residuals



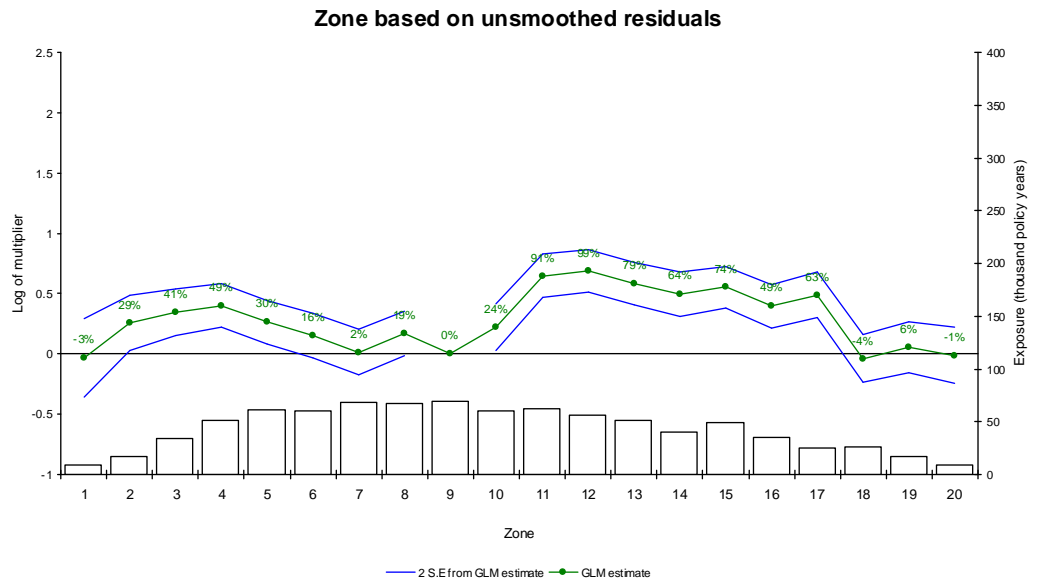
Finding the parameters

Effect of smoothed vs unsmoothed residual zone

Zone based on smoothed residuals

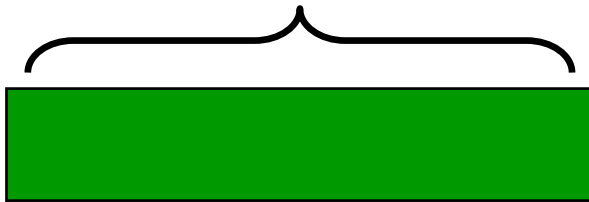


Zone based on unsmoothed residuals

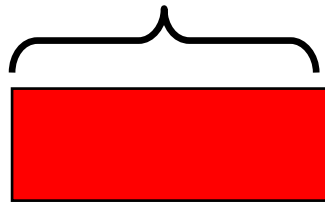


Making use of all the data

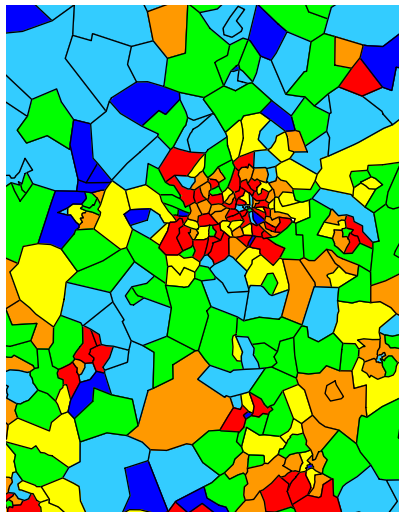
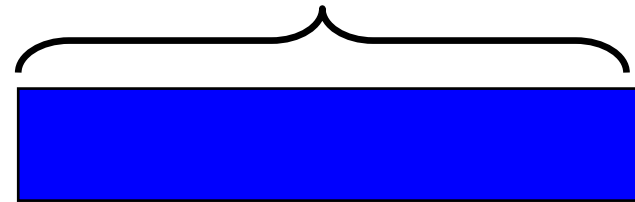
Calculate residuals



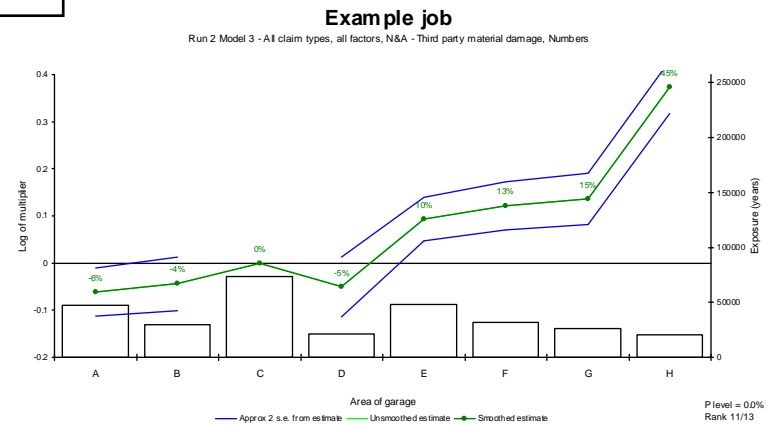
Seek parameters which minimize error



Save for determining zoning relativities



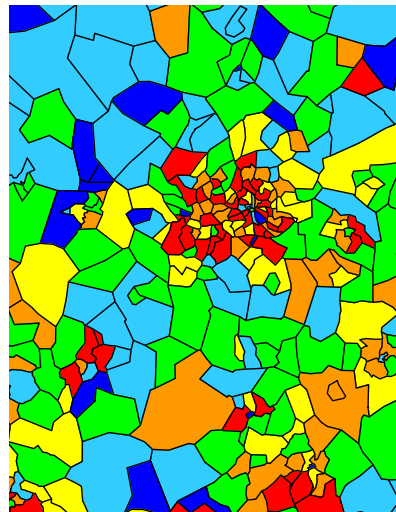
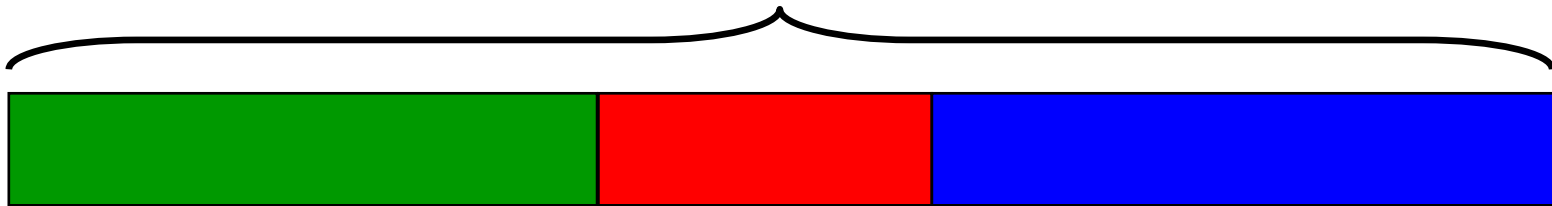
a, m, n, b



Freeze

Making use of all the data

Calculate unsmoothed residuals on all the data, smooth using **frozen** parameters, categorize smoothed residuals into zones in same way, assume **frozen** parameter estimates hold for each zone and set those effects as an offset in the main GLM on the same data

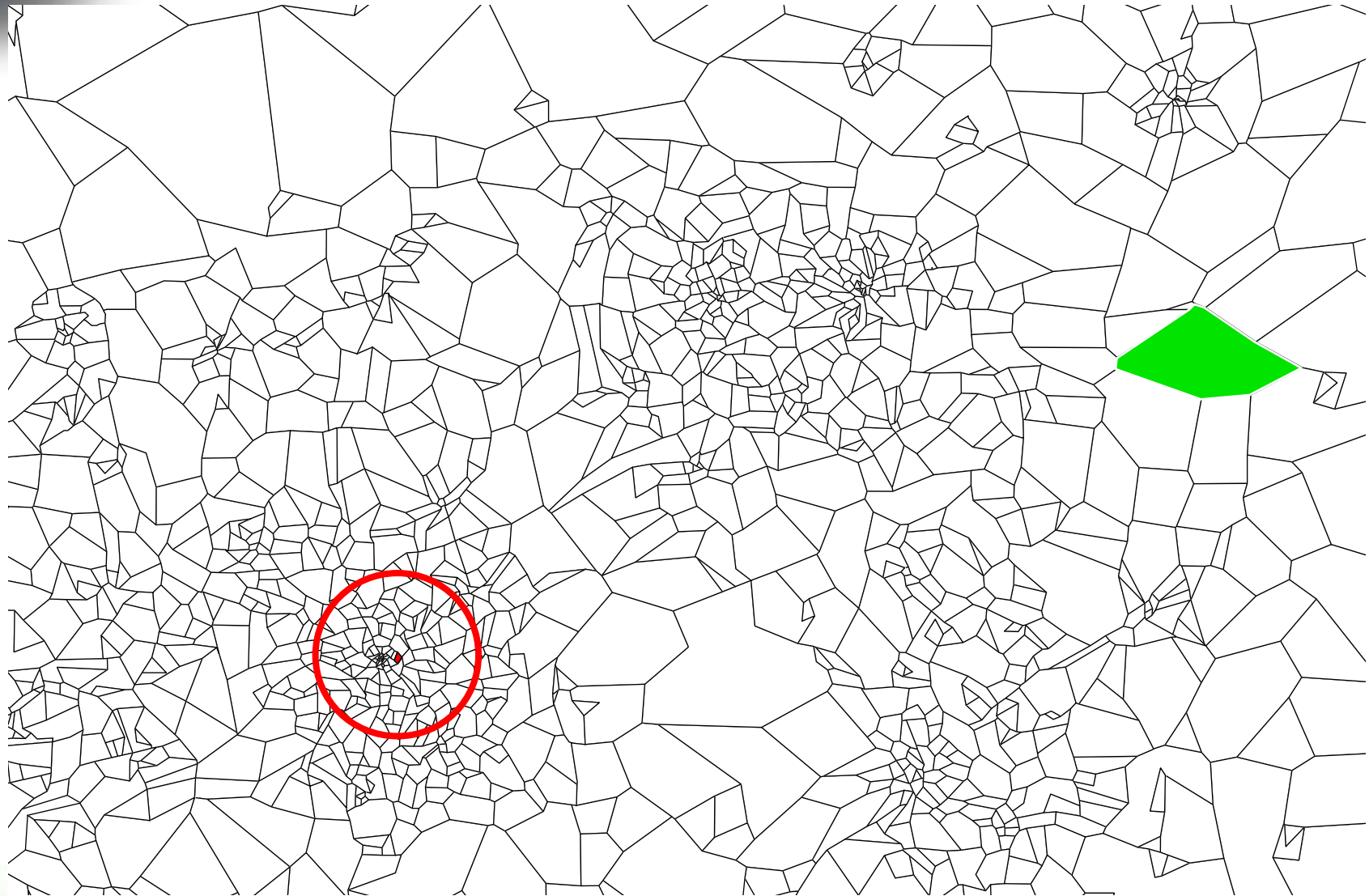




More details...

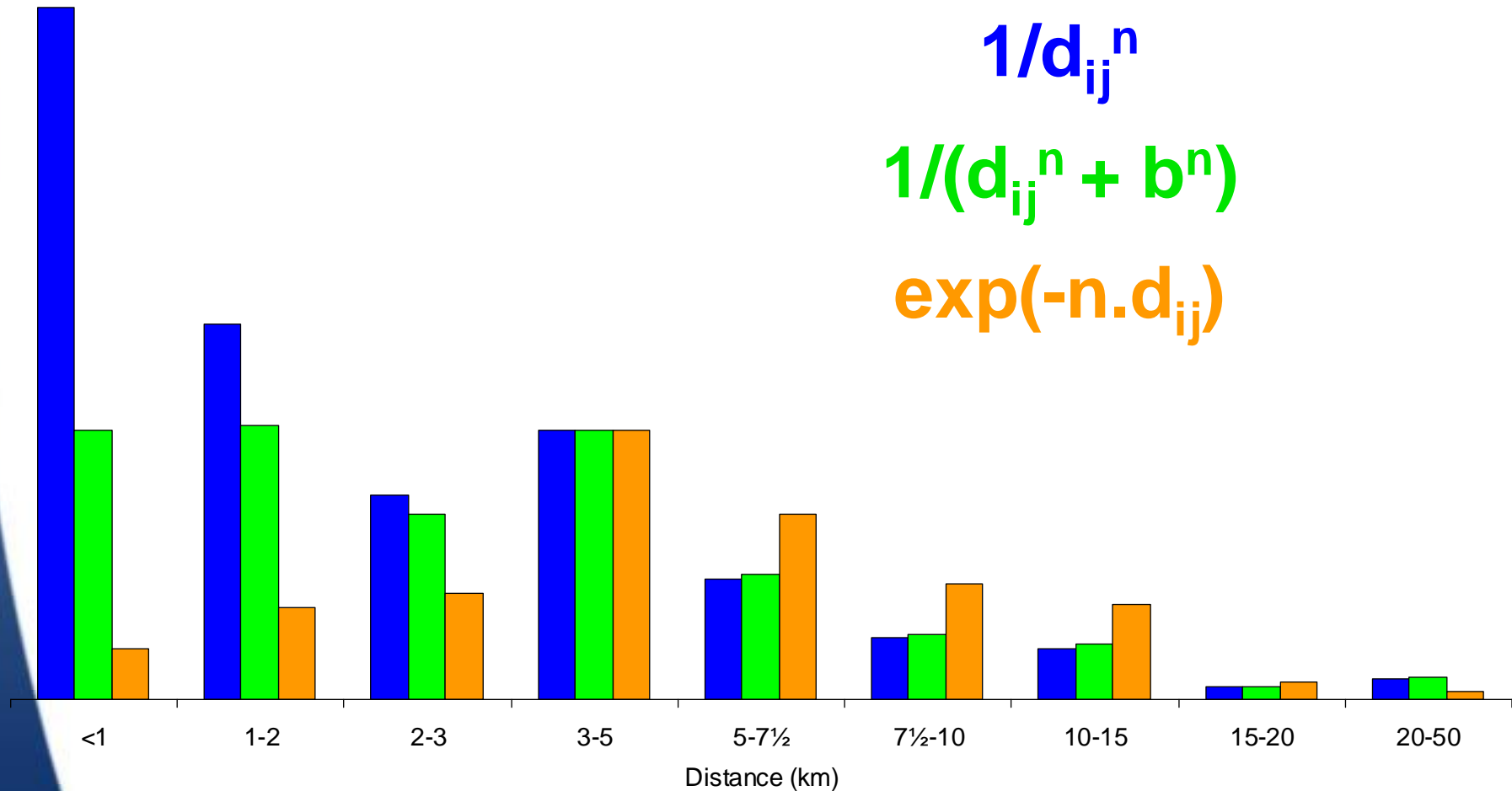
- Different
 - weighting functions
 - metrics
 - splits of the data
- Performance
- What to do when there is
 - no boundary data
 - no zip codes
- Geodemographic factors
- The competitive situation

Different weighting functions



Different weighting functions

Influence of neighbors in total - **urban** area



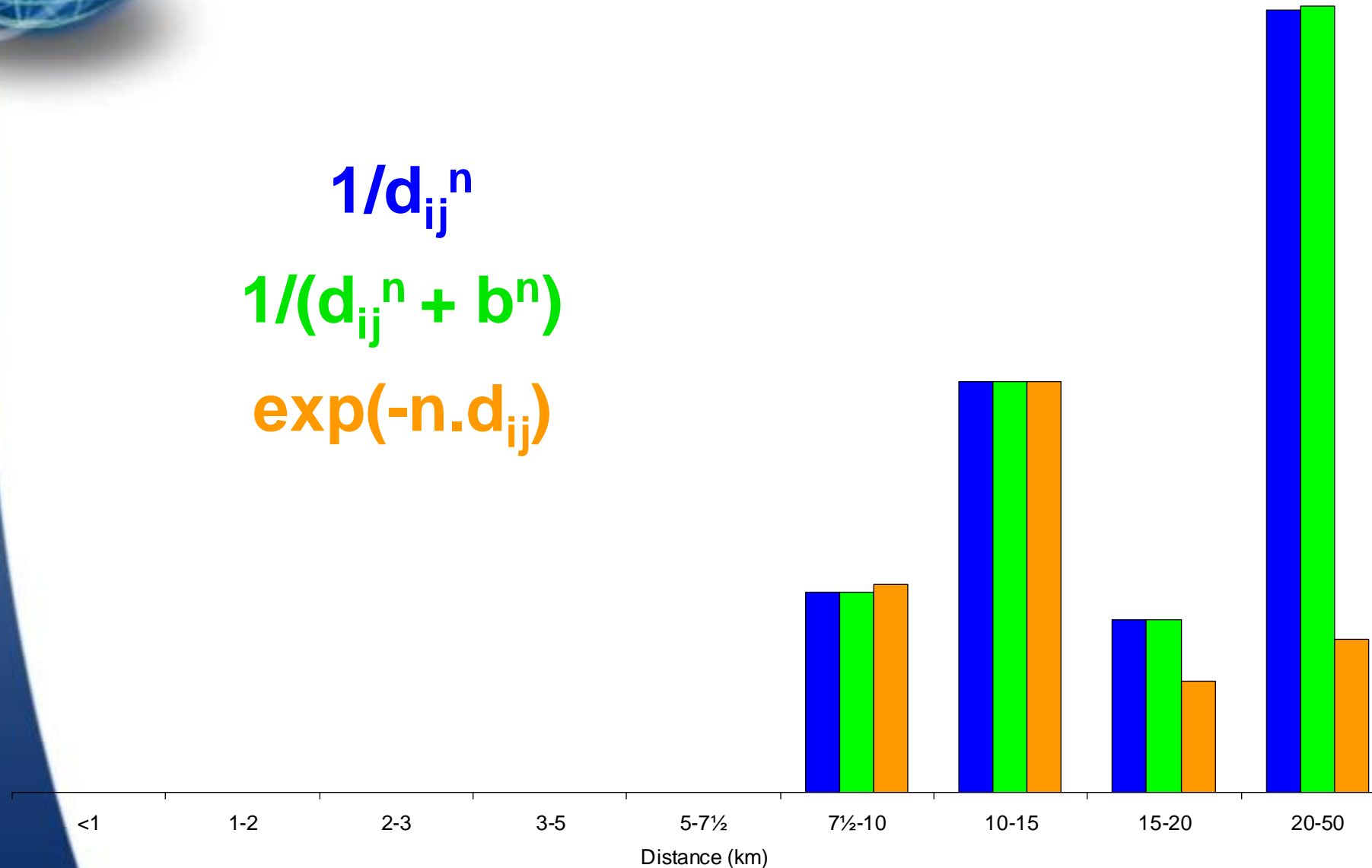
Different weighting functions

Influence of neighbors in total - rural area

$$1/d_{ij}^n$$

$$1/(d_{ij}^n + b^n)$$

$$\exp(-n \cdot d_{ij})$$





Different metrics

$$r_i^* = Z(e_i) \cdot r_i + (1 - Z(e_i)) \frac{\sum_j e_j \cdot r_j \cdot f(d_{ij})}{\sum_j e_j \cdot f(d_{ij})}$$

where

r_i^* = smoothed residual r_i = unsmoothed residual

$Z(e_i) = \{ e_i / (e_i + a) \}^m$ e_i = exposure in region i

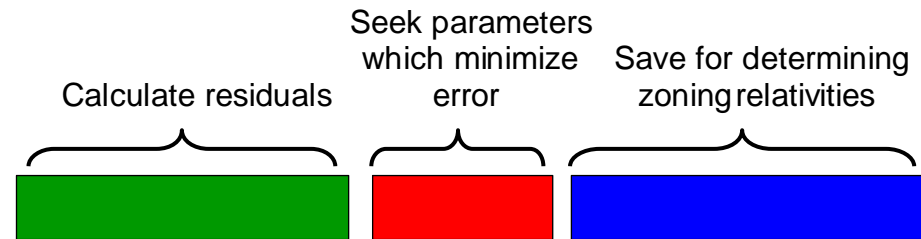
$$d_{ij} = \{ (x_i - x_j)^2 + (y_i - y_j)^2 + (s \cdot q_i - s \cdot q_j)^2 \}^{1/2}$$

$f(d_{ij}) = 1/d_{ij}^n$ or $1/(d_{ij}^n + b^n)$ or $\exp(-n \cdot d_{ij})$ etc

Splitting the data

- Generally a random split is best, otherwise policy characteristics distort results
- Weather related perils are a notable exception - here a time split may be more appropriate

(not appropriate to model large weather events with this method)



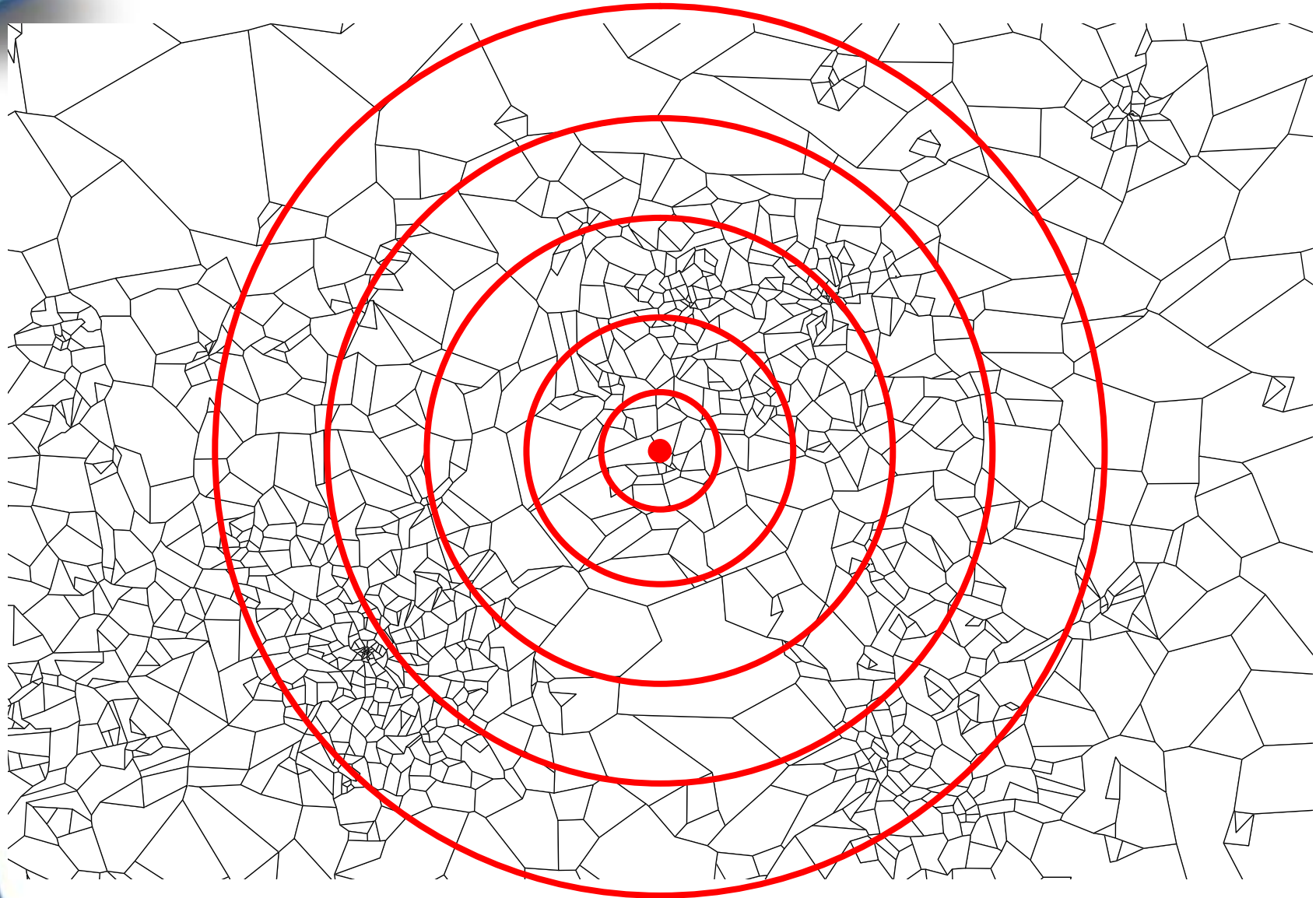


More details...

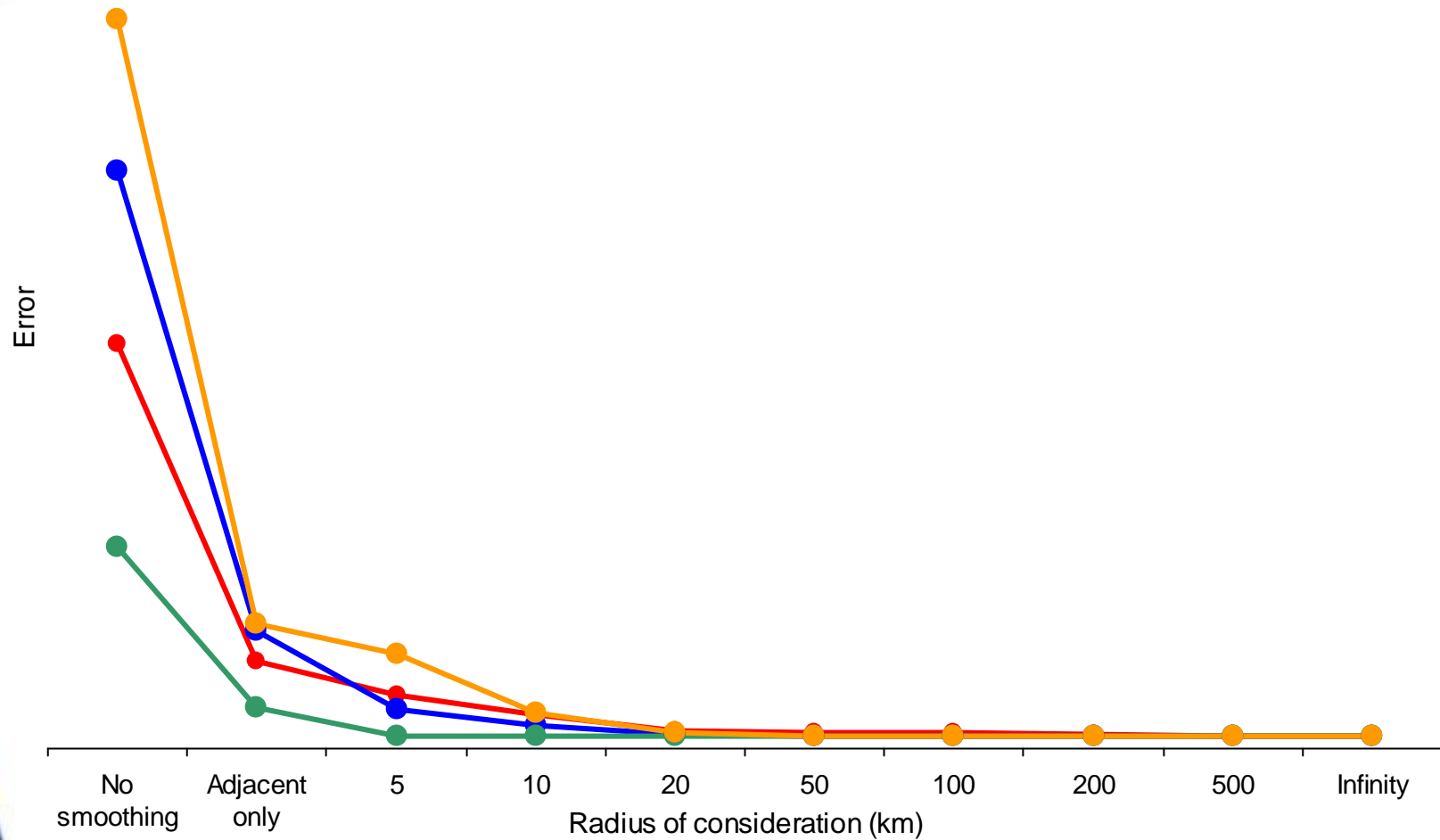
- Different
 - weighting functions
 - metrics
 - splits of the data
- Performance
- What to do when there is
 - no boundary data
 - no zip codes
- Geodemographic factors
- The competitive situation



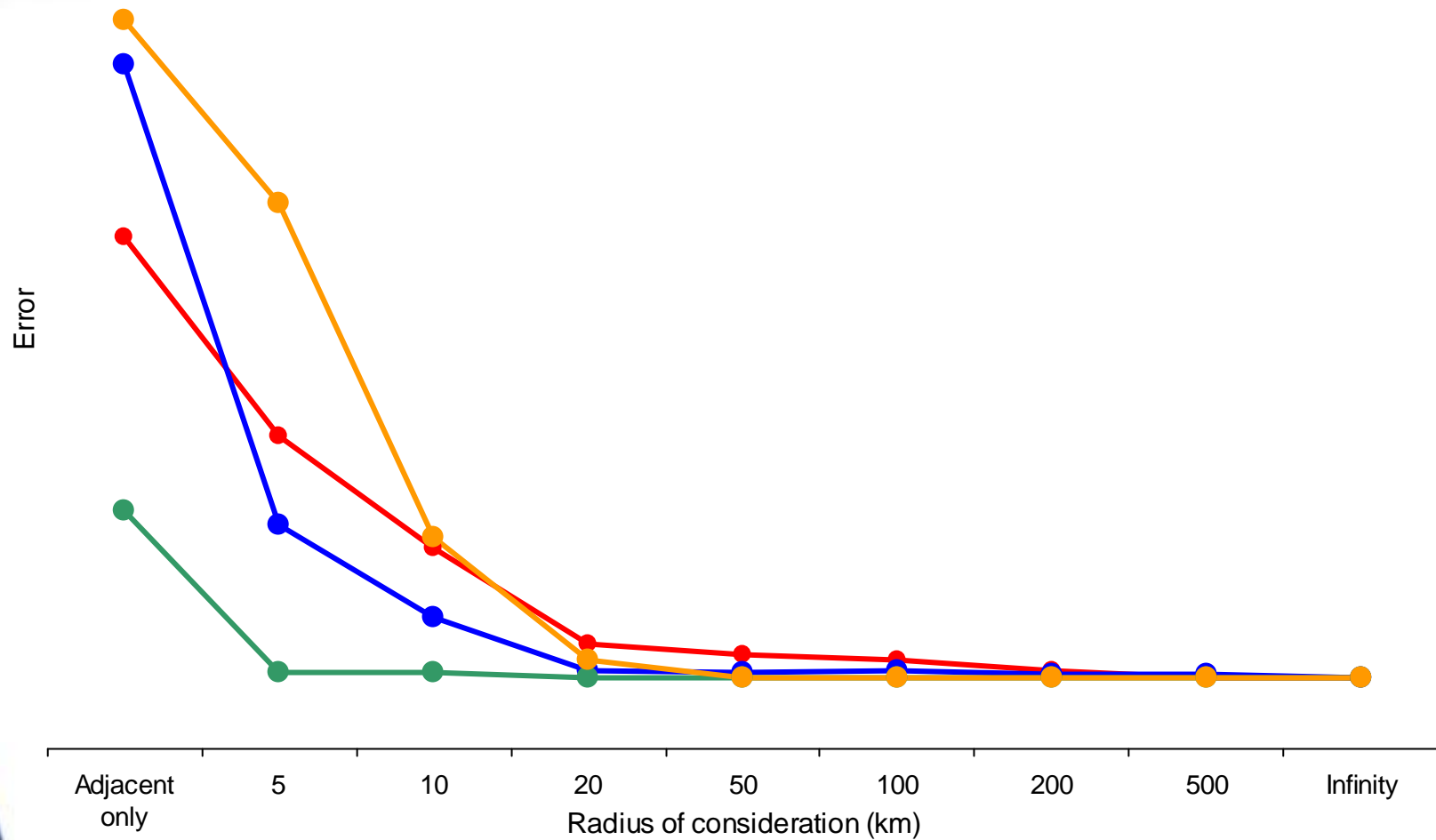
Limiting the definition of "neighboring"



Error with differing radii of consideration



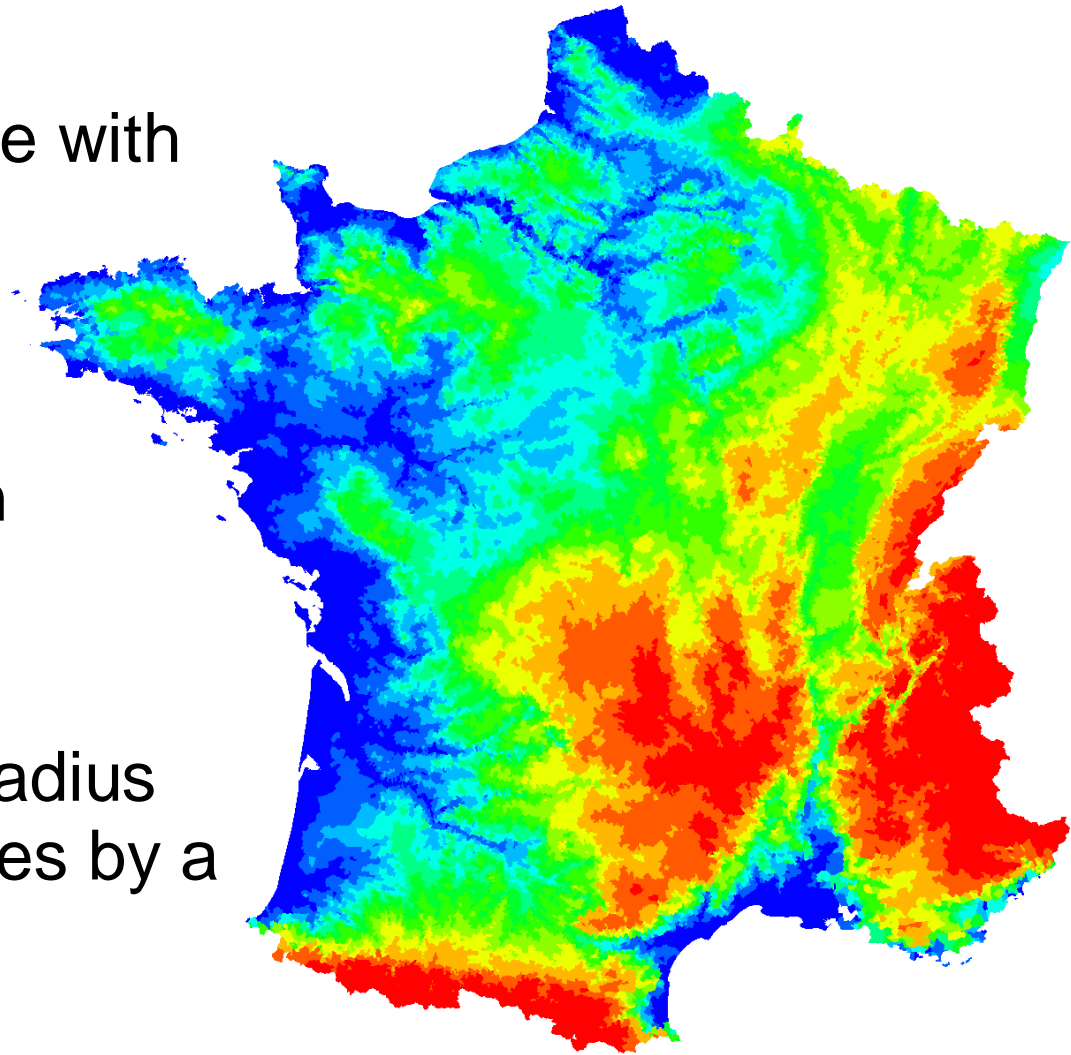
Error with differing radii of consideration





Computational short cuts

- Run times increase with $(\# \text{ regions})^2$
- There are 36,500 communes in France, ie 1 billion calculations per iteration!
- Limiting to 50km radius decreases run times by a factor of 6

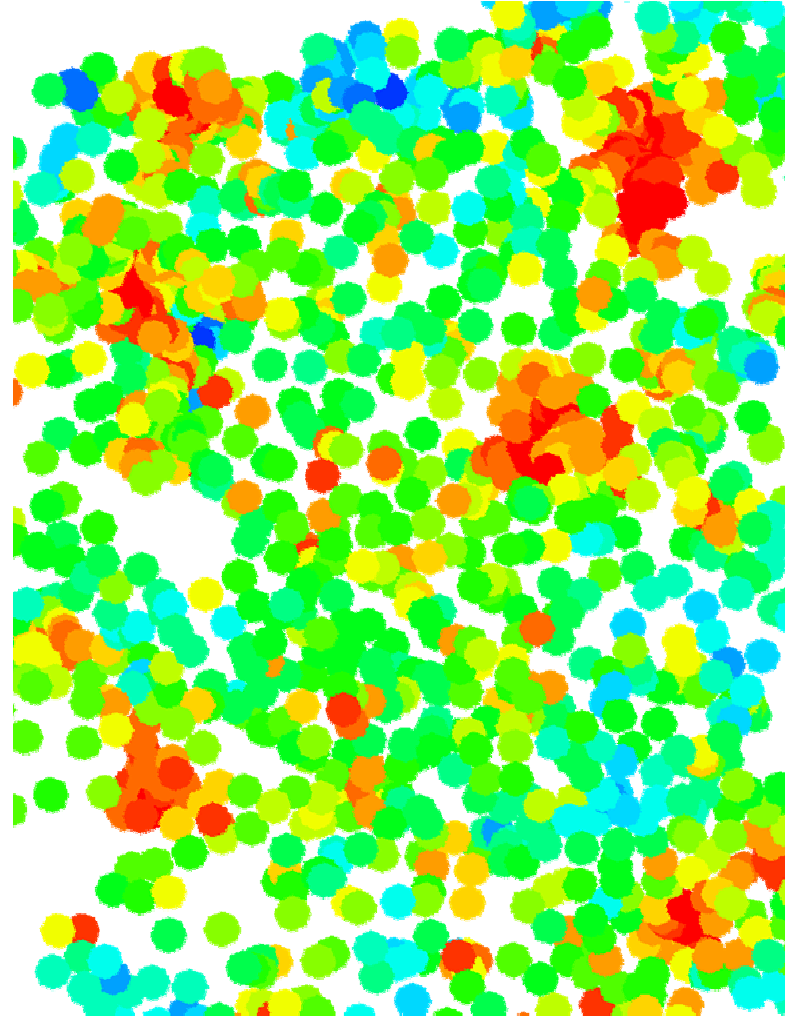
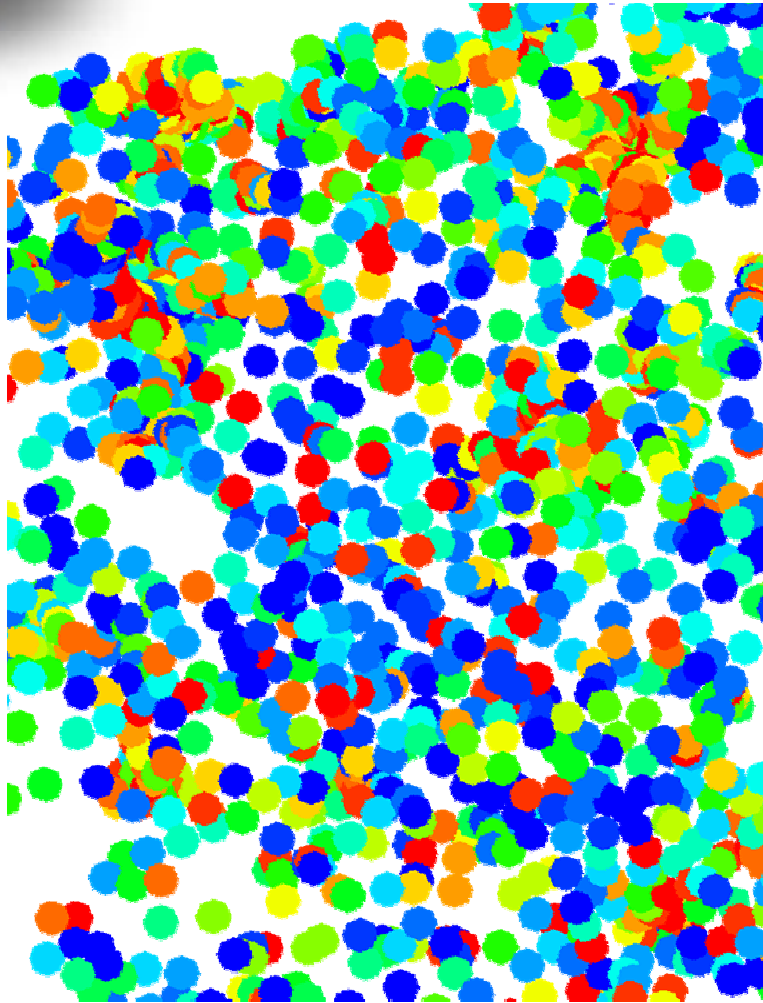




More details...

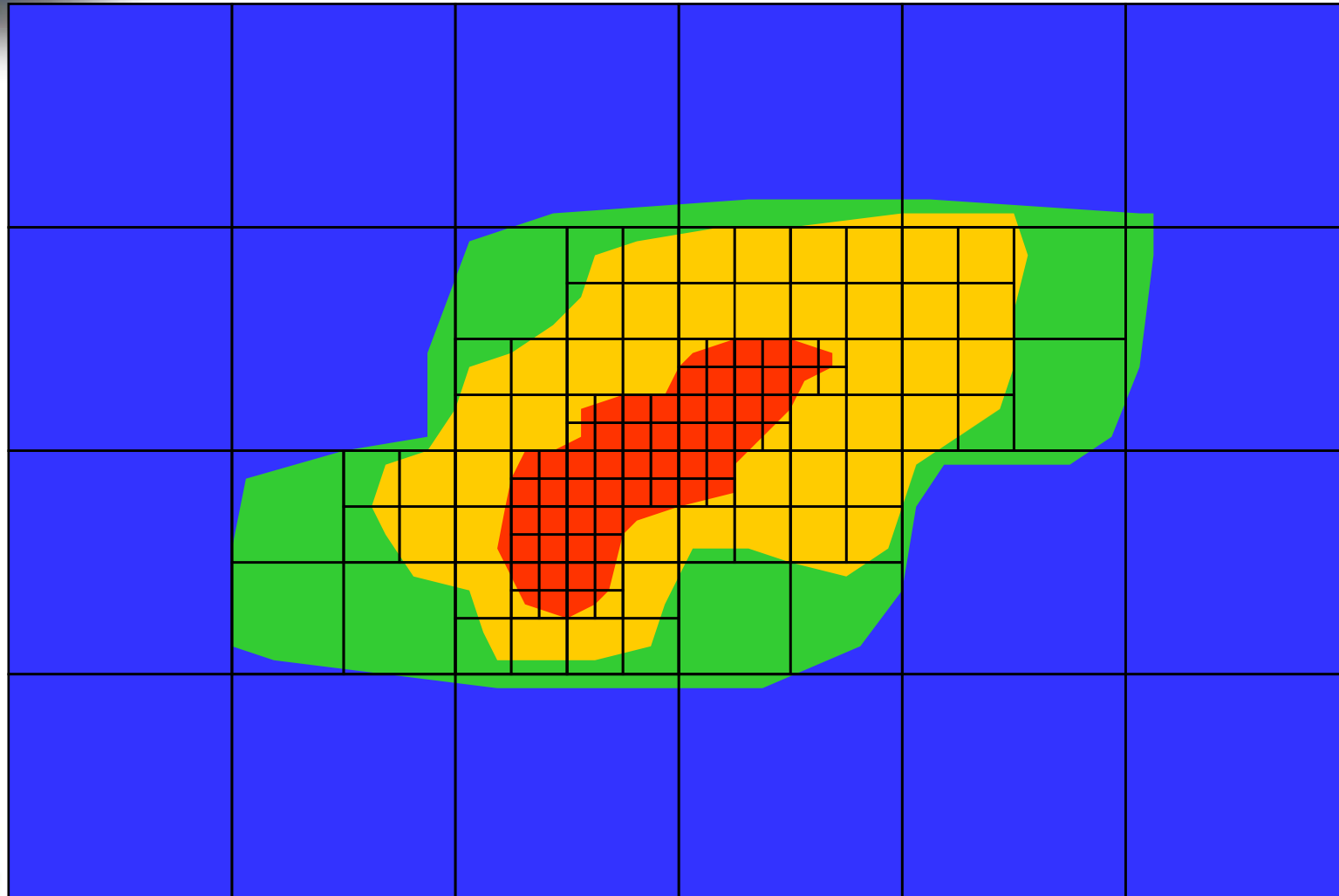
- Different
 - weighting functions
 - metrics
 - splits of the data
- Performance
- What to do when there is
 - no boundary data
 - no zip codes
- Geodemographic factors
- The competitive situation

When no boundary data is available (but x, y is)...

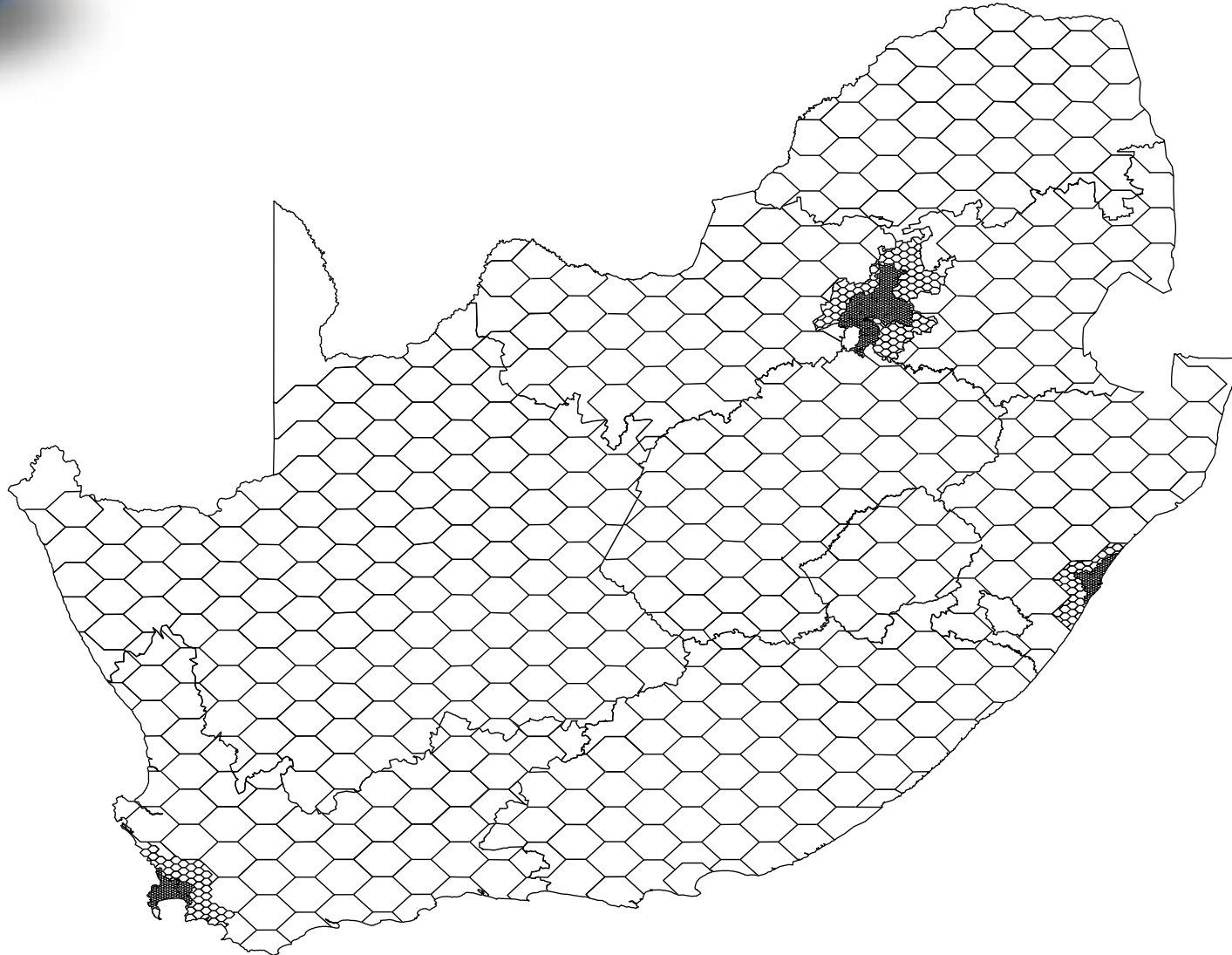




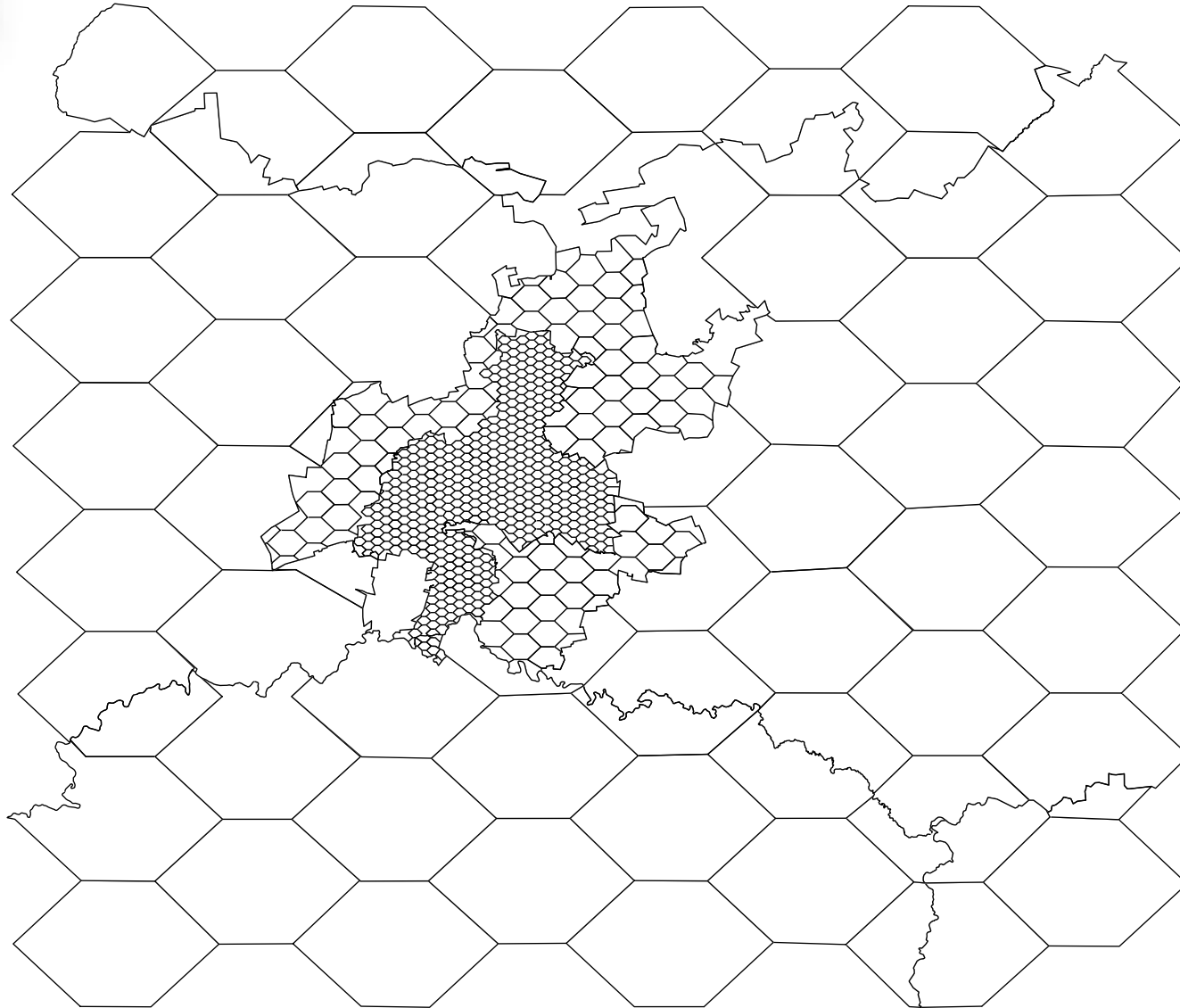
When no zip codes used...



When no zip codes used....



When no zip codes used...





When no post codes exist...

- (real South African results cannot be disclosed in this handout)



More details...

- Different
 - weighting functions
 - metrics
 - splits of the data
- Performance
- What to do when there is
 - no boundary data
 - no zip codes
- Geodemographic factors
- The competitive situation



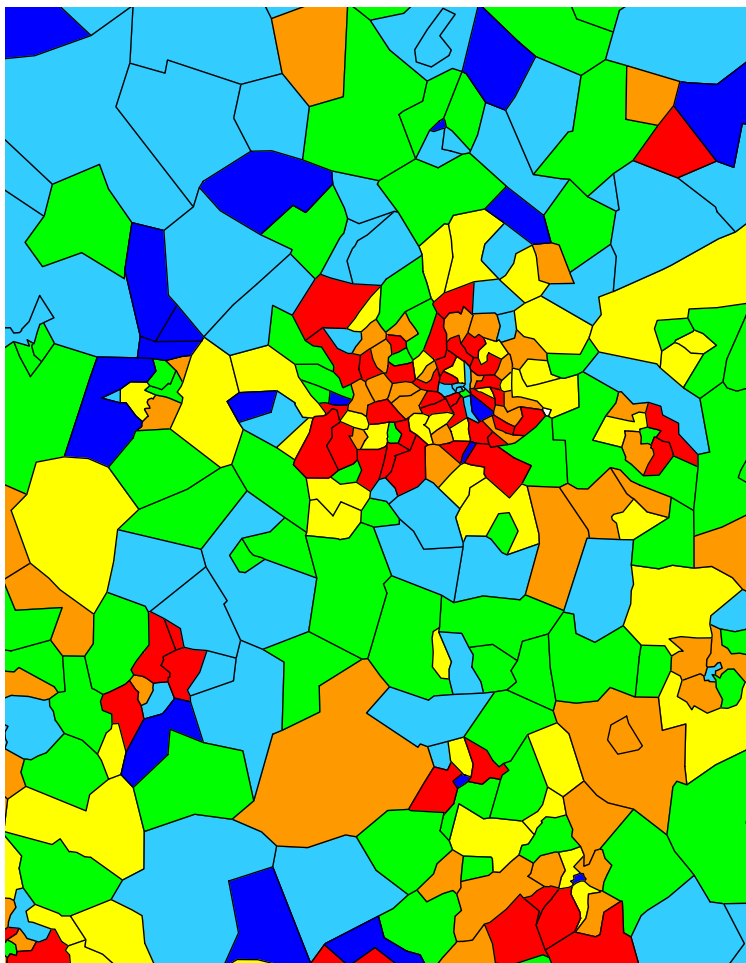
Geodemographic factors

- Can be very predictive
- Even simple measures of urban density can be interesting
- Can be used
 - (a) alongside zones derived as above
 - (b) to standardize experience prior to smoothing
- Investigate which yields most predictive zone
- Generally speaking, seek to standardize for factors which yield inherently smoother residuals

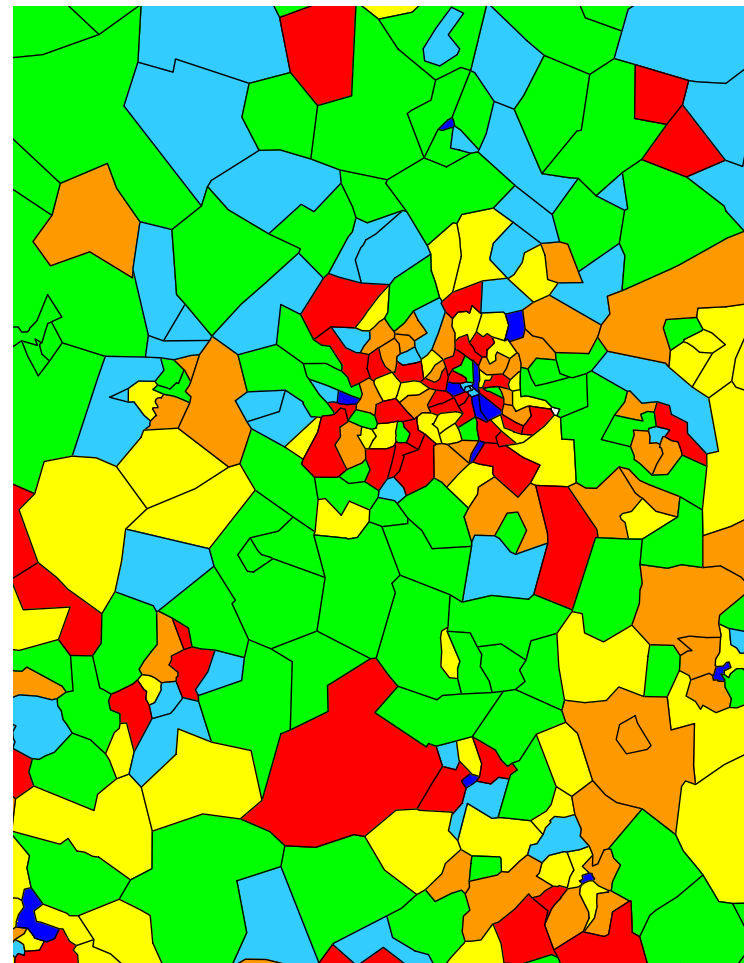


Unsmoothed residuals

Density not in
standardizing GLM



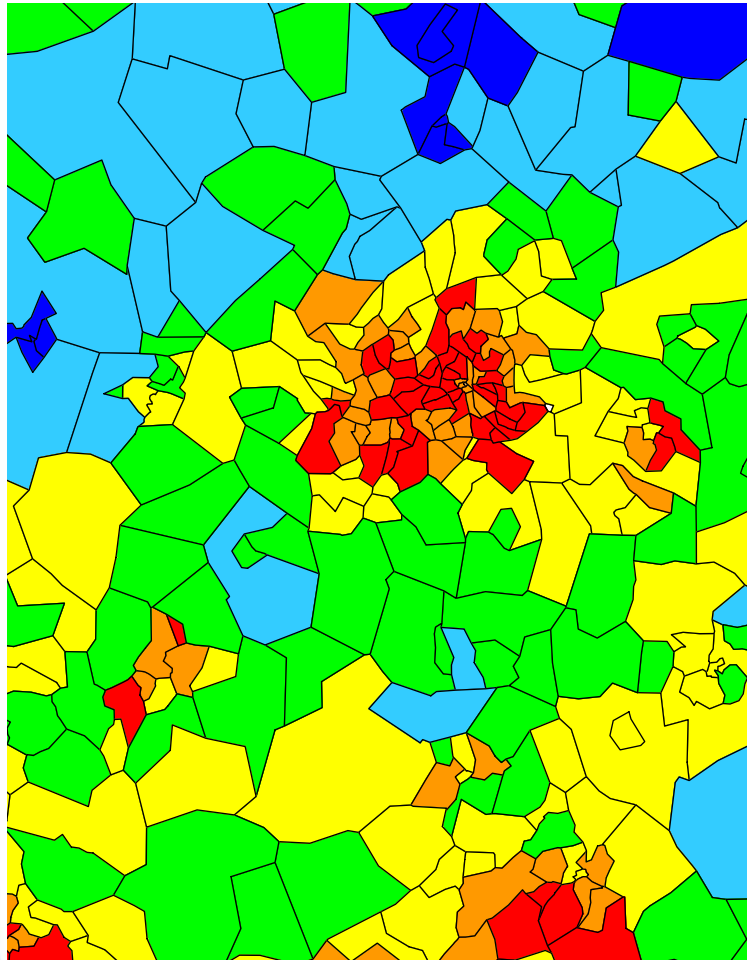
Density in
standardizing GLM



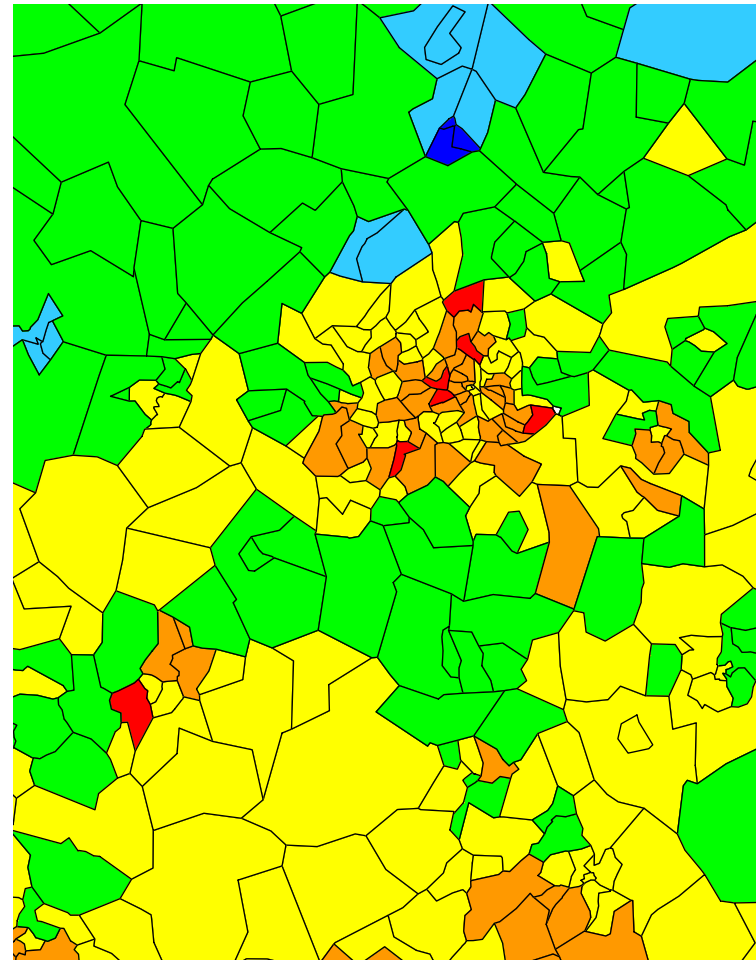


Smoothed residuals

Density not in
standardizing GLM



Density in
standardizing GLM





More details...

- Different
 - weighting functions
 - metrics
 - splits of the data
- Performance
- What to do when there is
 - no boundary data
 - no zip codes
- Geodemographic factors
- The competitive situation

Comparing smoothed results with existing rates and the market

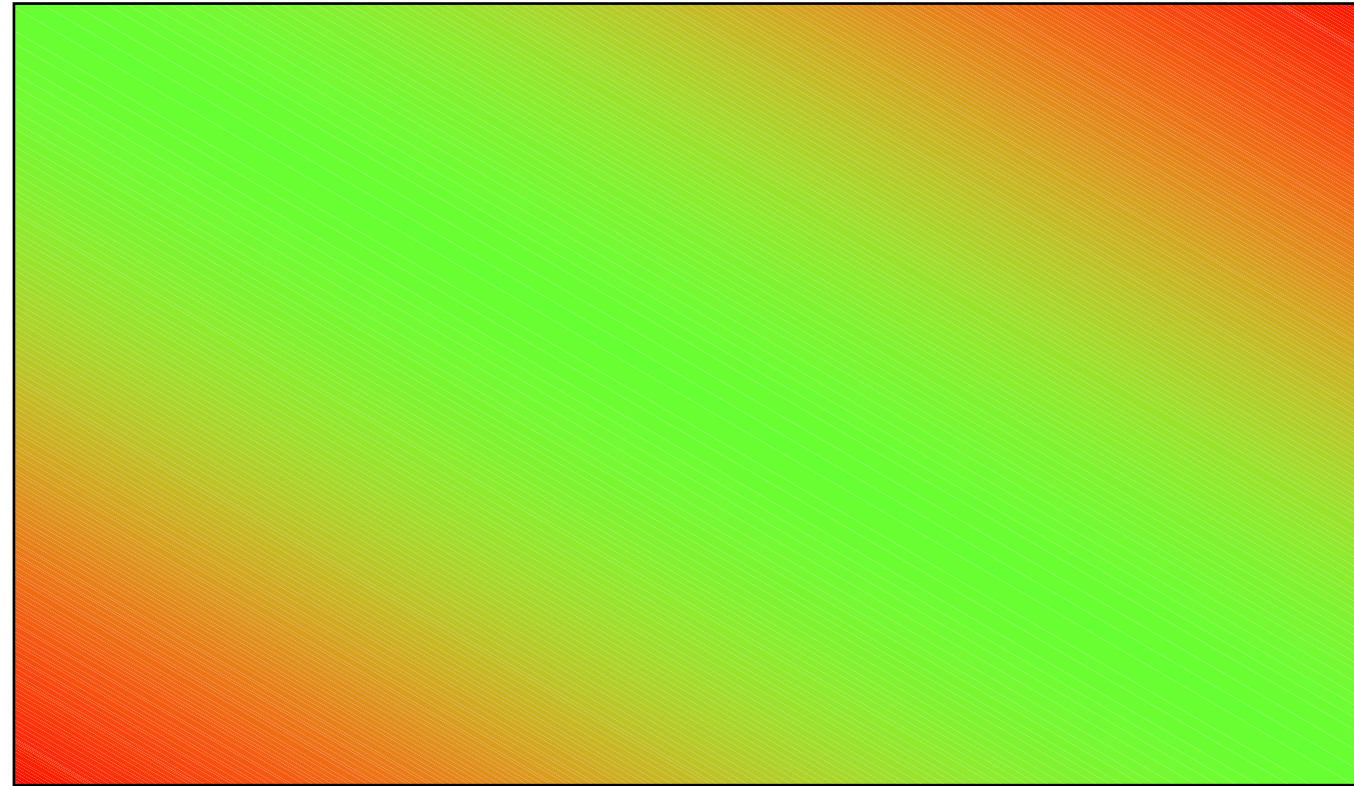
Company premium vs market

Below market

Above market

Theoretically desired change in premium

Increase
Decrease



Comparing smoothed results with existing rates and the market

Theoretically desired change in premium

Company premium vs market
Below market Above market



Comparing smoothed results with existing rates and the market

Company premium vs market

Below market

Above market

Theoretically desired change in premium

Increase

Decrease

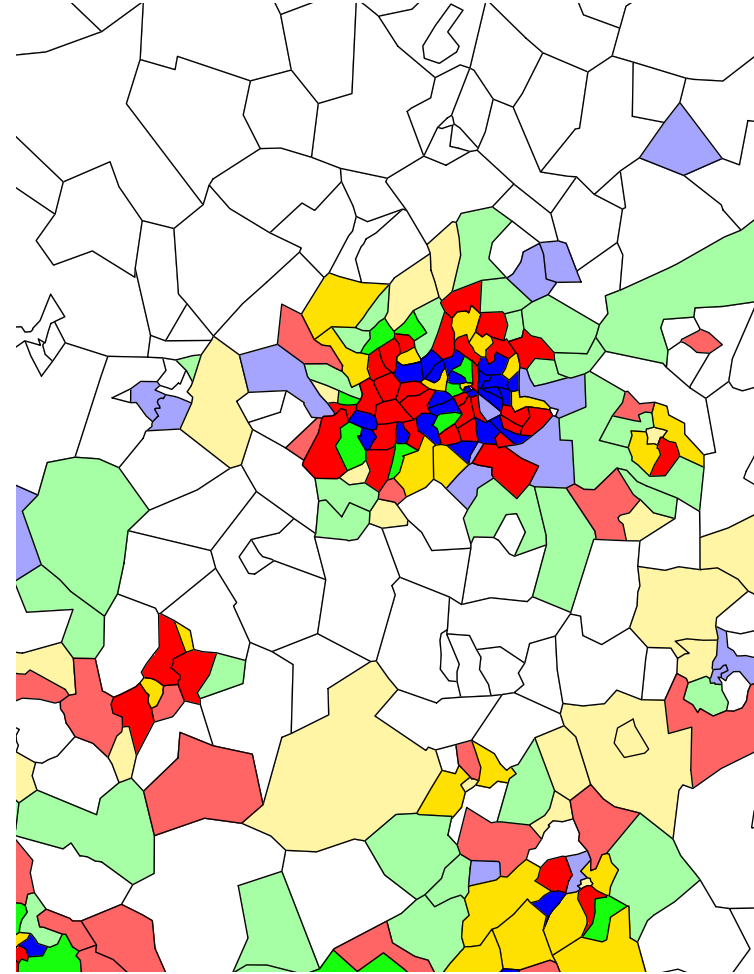
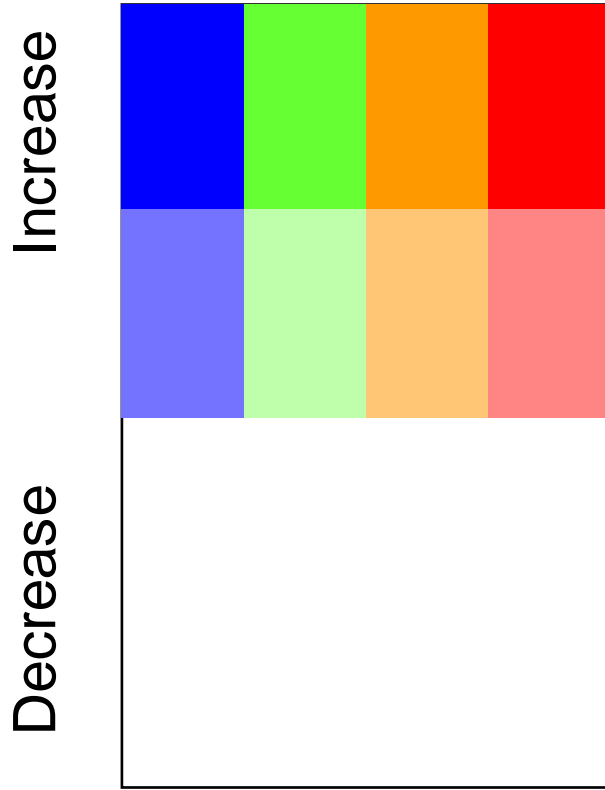


Comparing smoothed results with existing rates and the market

Theoretically desired change in premium

Company premium vs market

Below Above

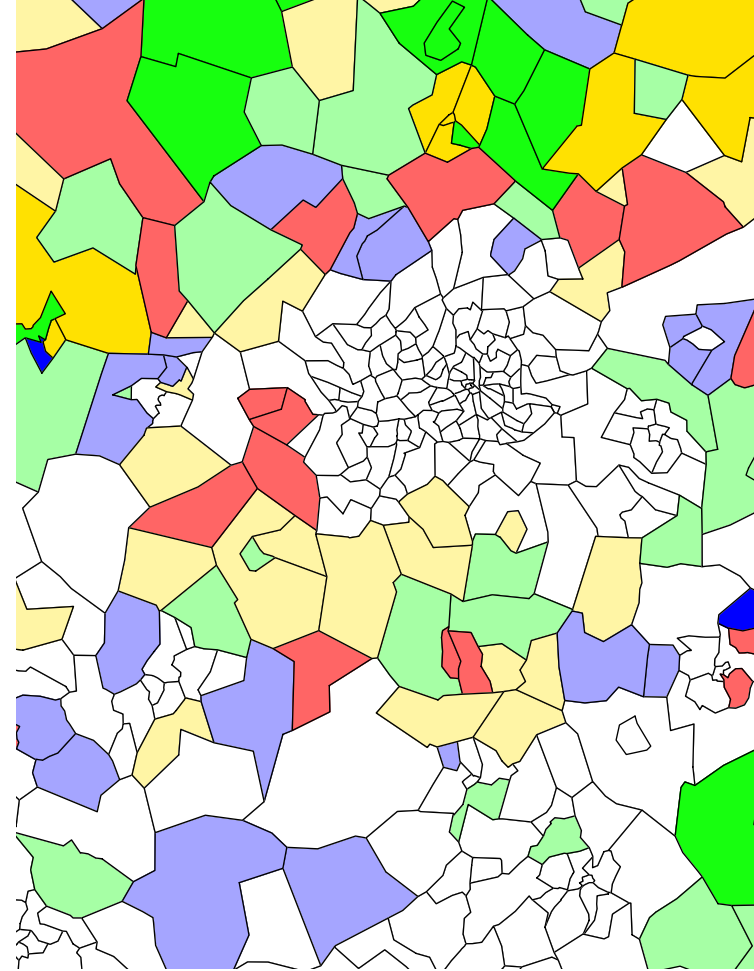
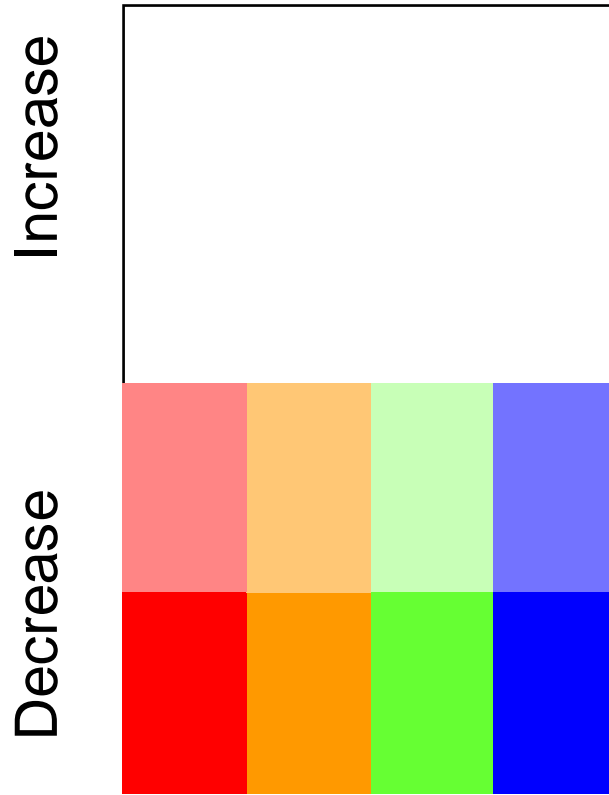


Comparing smoothed results with existing rates and the market

Theoretically desired change in premium

Company premium vs market

Below Above



Geographical Spatial Analysis in Personal Lines Territorial Ratemaking

2004 CAS Seminar on
Ratemaking

Duncan Anderson MA FIA

Watson Wyatt LLP



WWW.WATSONWYATT.COM/INSURANCE

