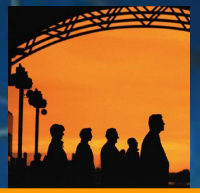
PL-7

Putting Your Company on the Map:

Determination of Statistically Indicated Territory Boundaries


WWW.WATSONWYATT.COM

2006 CAS Seminar on Ratemaking

Duncan Anderson MA FIA

Watson Wyatt Worldwide

Watson Wyatt Worldwide

Background

- 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)

Ingredients for a solution

Own experience of a "region"

Geodemographic factors

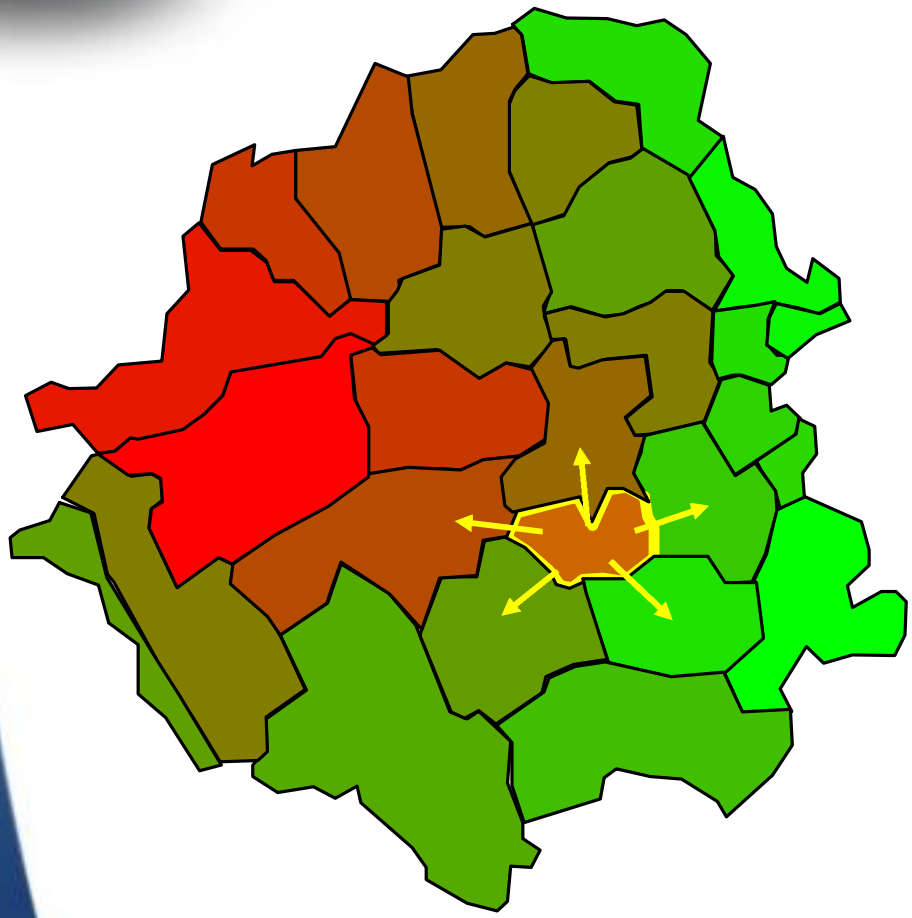
Experience from "nearby" regions

Standard rating factors

Competitor rates

Gut feeling

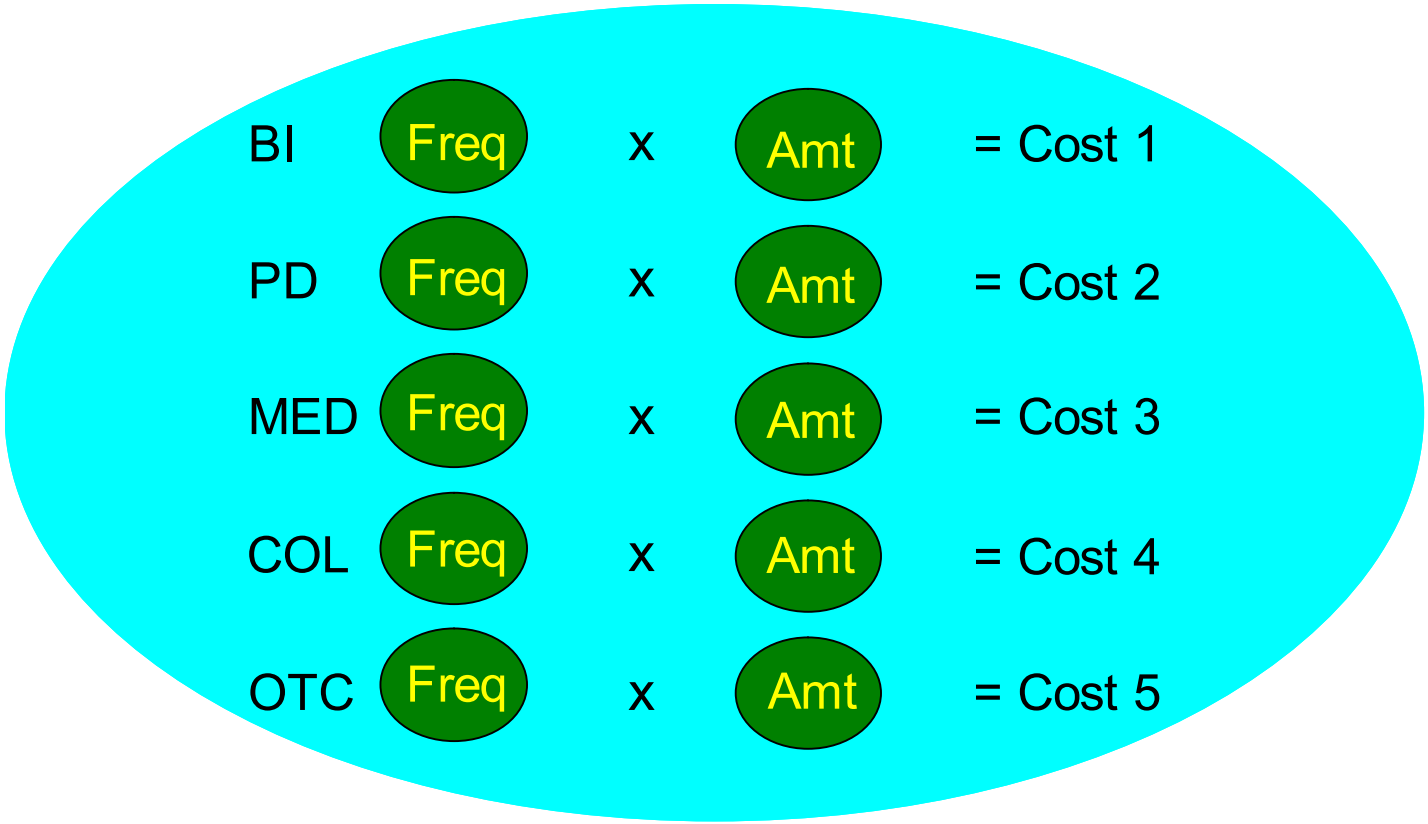
Proximity



- Key assumption is that "close" areas are similar
- May not be a perfect assumption
- Nevertheless it seems consistently to yield good results in practice

What to model?

- Select which element of experience to model



Two approaches to spatial smoothing

- Estimate effect of non-territory factors and then smooth residuals to derive new zones
  - + very practical
  - + can include differing distance metrics
  - + can incorporate credibility in a straightforward way
  - - distorted by non-systematic element of experience
  - - slight distortion from correlated factors
- Fit surface directly using maximum likelihood as part of GLM (ideally with splines)
  - + MLE
  - - harder to fit
  - - prone to over-smooth

Residual smoothing - a method

1. Assess true area risk as well as possible
2. Define "zones" containing areas of similar risk (may or may not be contiguous)
3. Determine loading applicable to each "zone"

Residual smoothing - a method

- 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"
- Smooth this to make it more predictive (at least in terms of rank ordering) of future experience 1
- Then categorize into zones 2
- And derive appropriate loadings for each zone 3

A model form

$r_i^* = Z.r_i + (1 - Z) .$ neighboring experience

where

r_i^* = smoothed residual risk

r_i = unsmoothed residual risk

Z = credibility function

A model form

$r_i^* = Z.r_i + (1 - Z) .$ neighboring experience

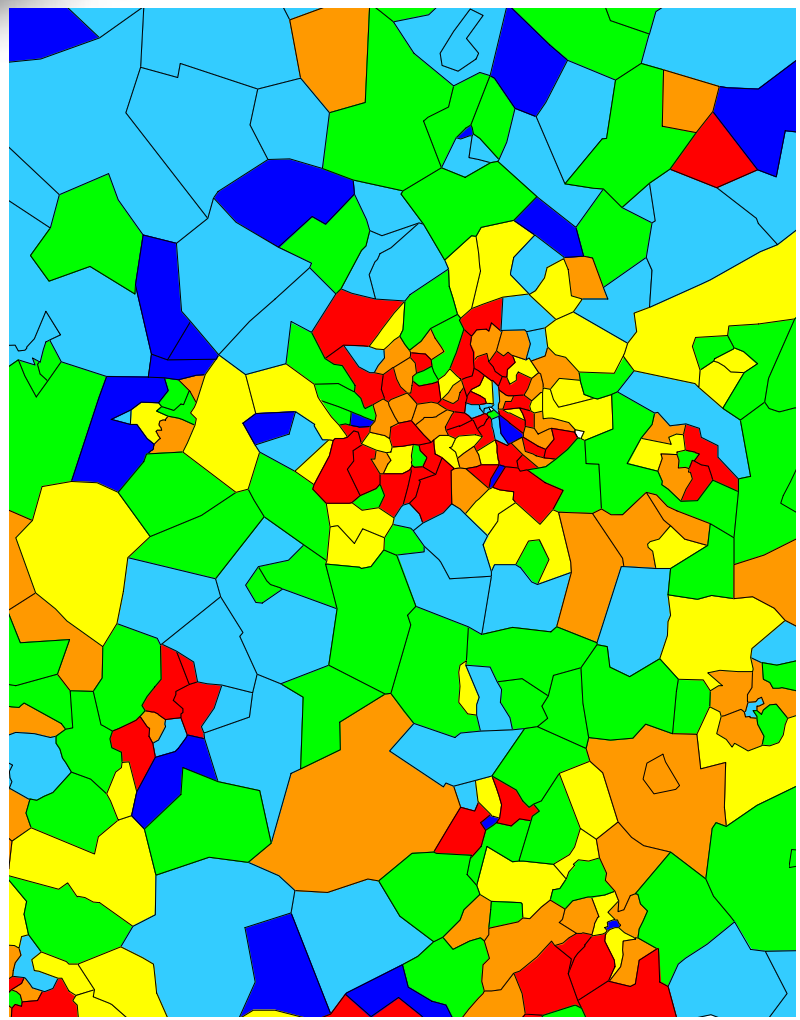
where

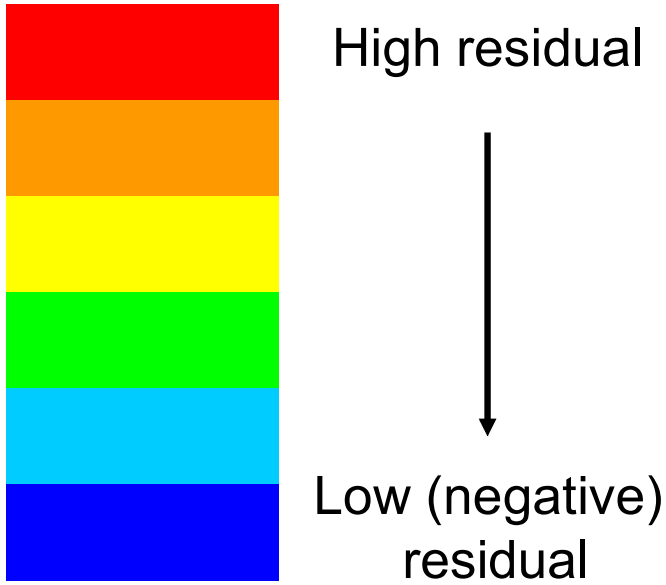
r_i^* = smoothed residual risk

r_i = unsmoothed residual risk

Z = credibility function

Example residual risk
UK homeowners contents theft frequency





A model form

$r_i^* = Z.r_i + (1 - Z) .$ neighboring experience

where

r_i^* = smoothed residual risk

r_i = unsmoothed residual risk

Z = credibility function

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

A model form

$r_i^* = Z.r_i + (1 - Z) .$ neighboring experience

where

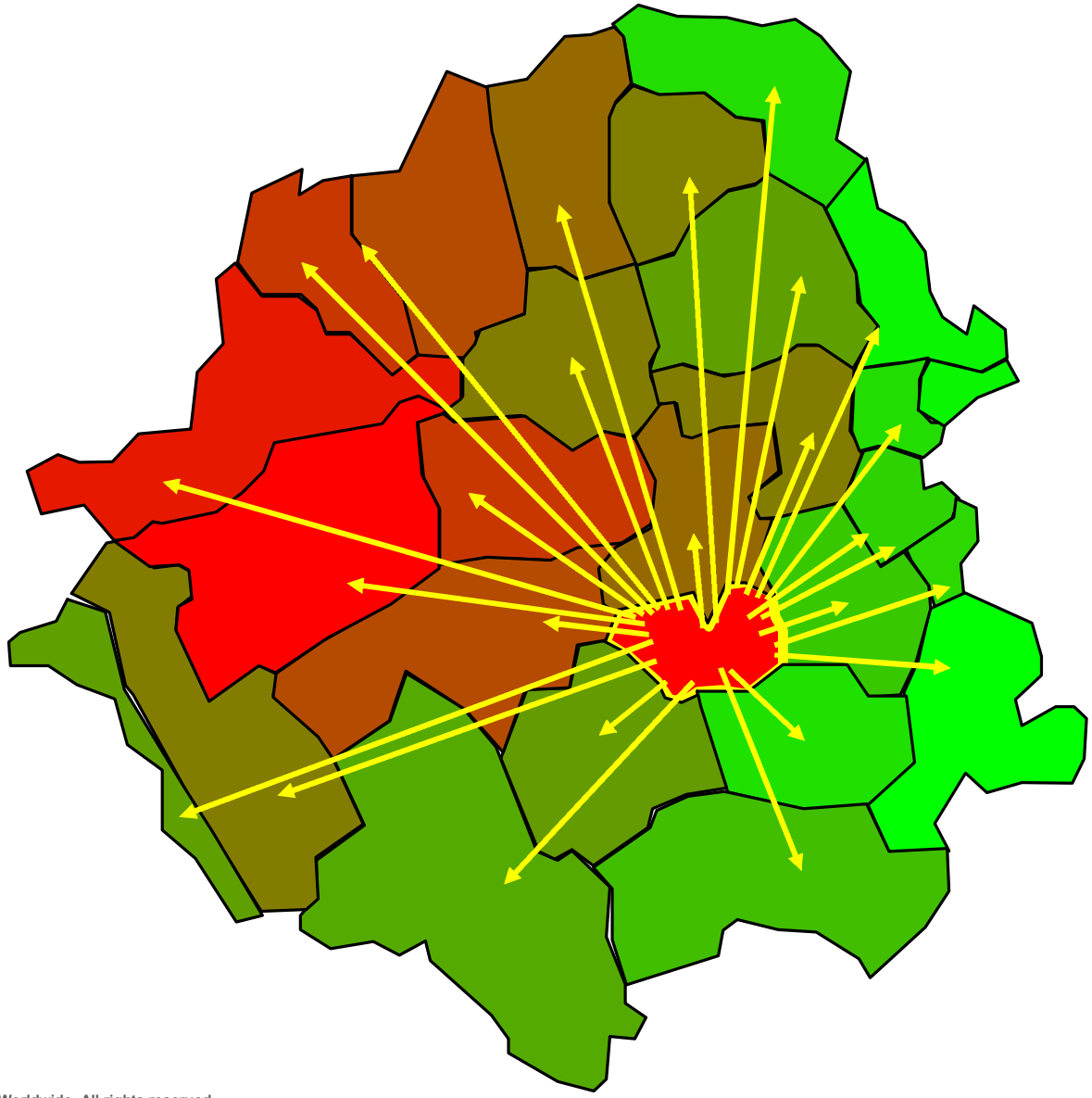
r_i^* = smoothed residual risk

r_i = unsmoothed residual risk

Z = credibility function

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

Definitions of "neighboring"



Model

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

where

r_i^* = smoothed residual r_i = unsmoothed residual

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

$$d_{ij} = \{ (x_i - x_j)^2 + (y_i - y_j)^2 \}^{½}$$

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

Parameters

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

where

r_i^* = smoothed residual r_i = unsmoothed residual

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

$$d_{ij} = \{ (x_i - x_j)^2 + (y_i - y_j)^2 \}^{½}$$

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

Finding the parameters

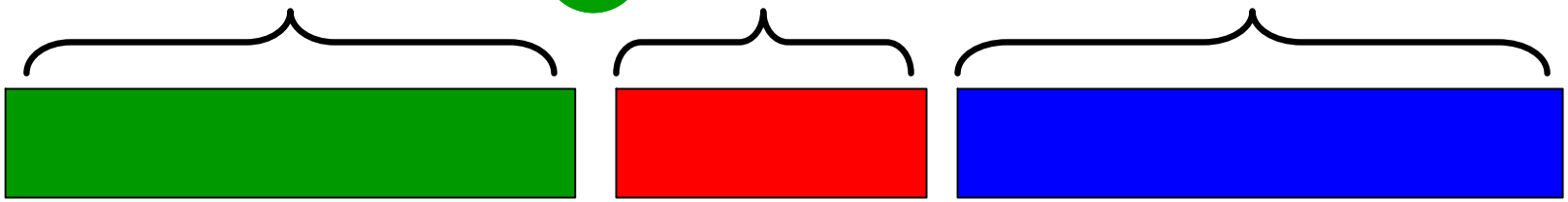
Calculate residuals

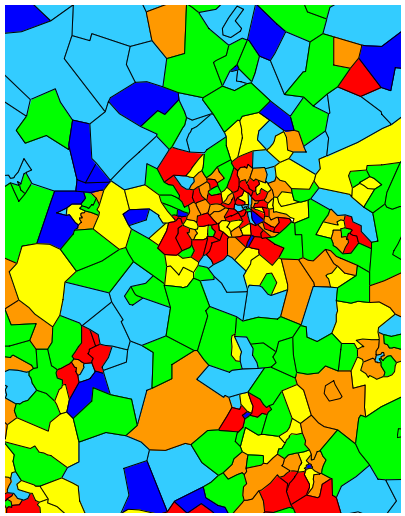
1

Seek parameters which minimize error

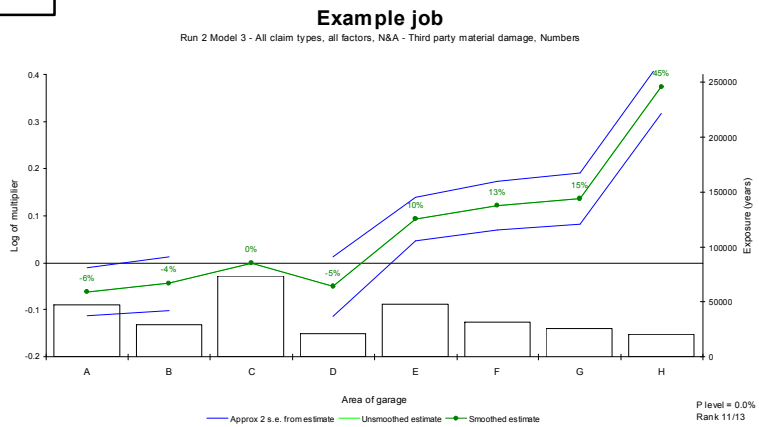
3

Save for determining zoning relativities



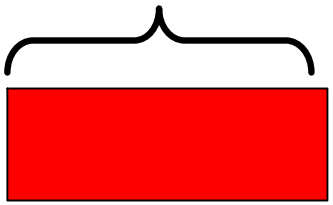


a, m, n, b



Finding the parameters

Seek parameters which minimize error



a, m, n, b

Error =

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

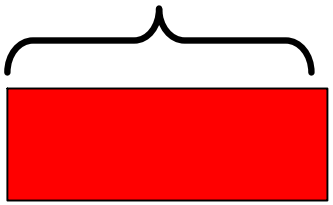
or

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

etc

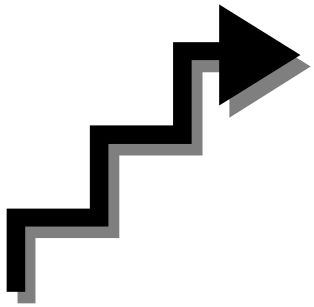
Finding the parameters

Seek parameters which minimize error

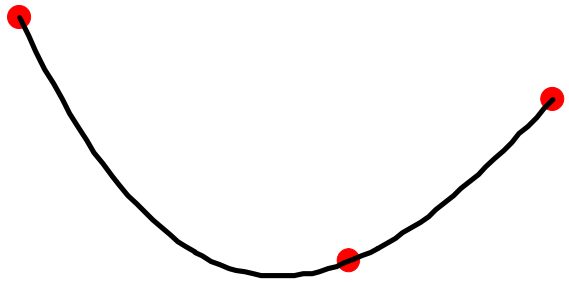


a, m, n, b

- Simple search



- Golden search

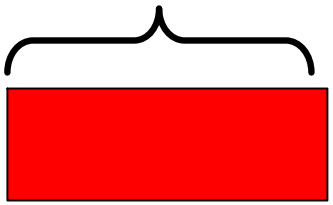


- Newton-Raphson

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

Finding the parameters

Seek parameters which minimize error

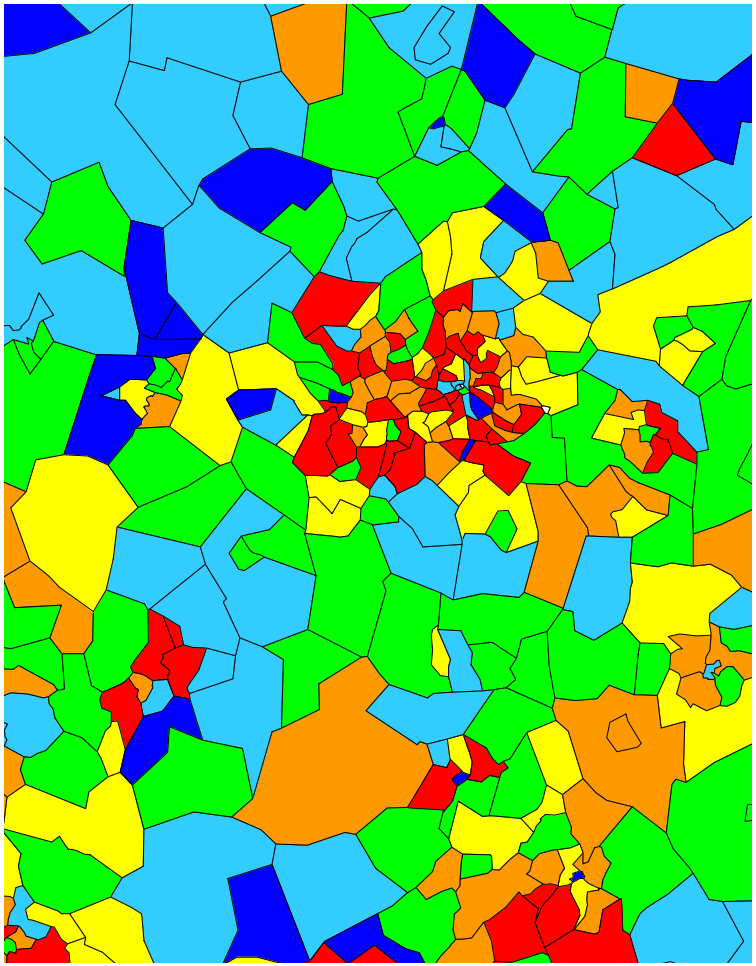


a, m, n, b

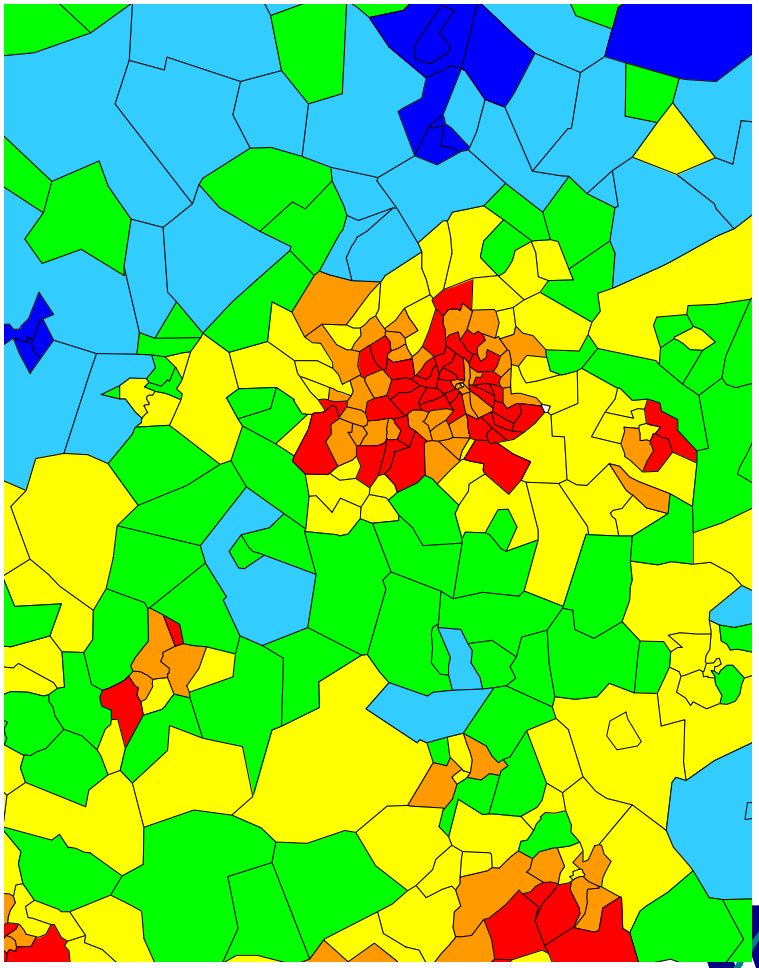
	n	e for Z=20%
USA	2.5	127
USA	1.9	106
France	2.0	104
France	1.9	146
Italy	1.4	87
Netherlands	1.8	61
South Africa	2.2	106
Spain	2.1	17
UK	1.9	146
UK	2.2	152
UK	1.8	78

Example results

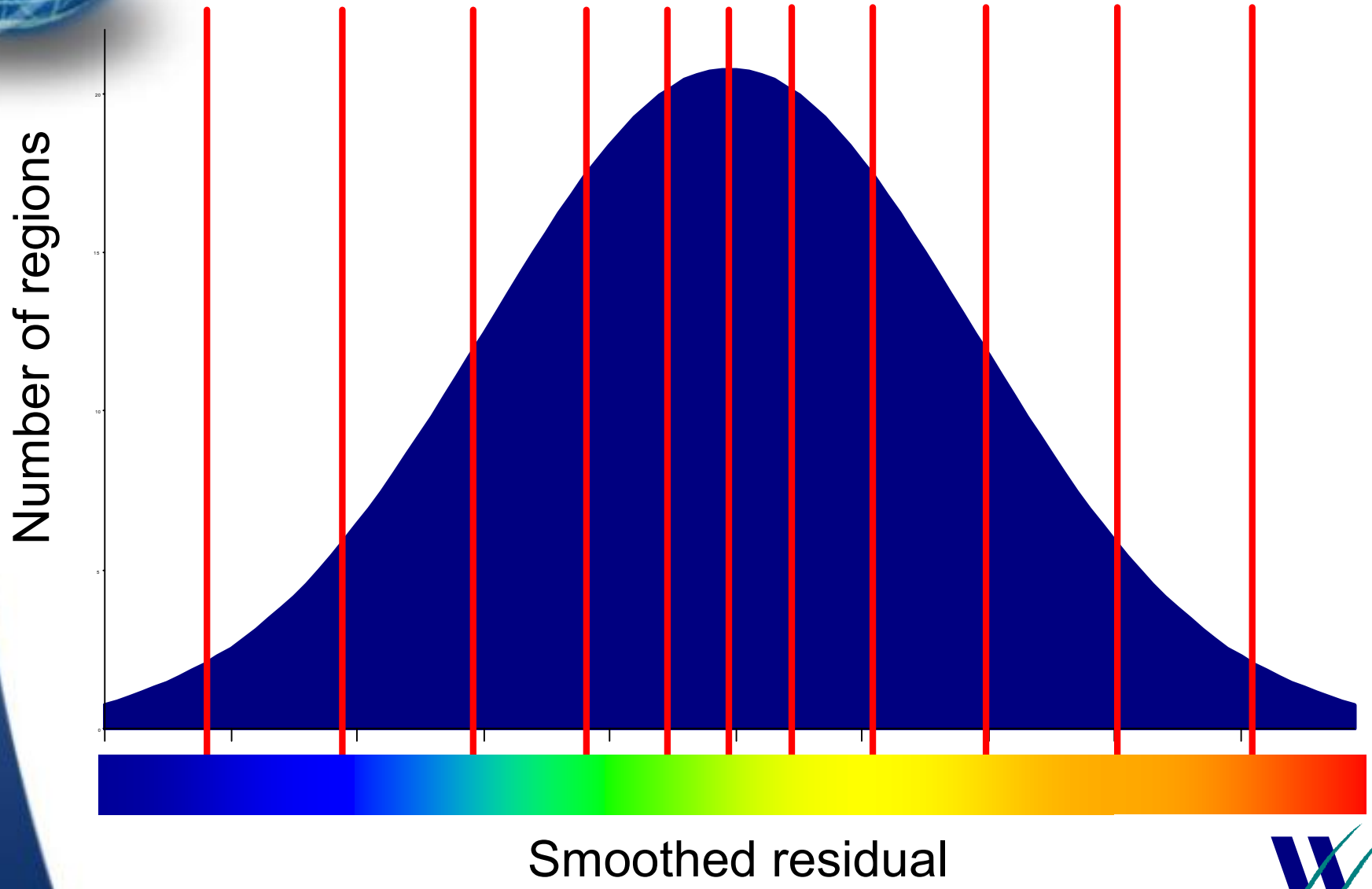
Unsmoothed residuals



Smoothed residuals



Creating zones



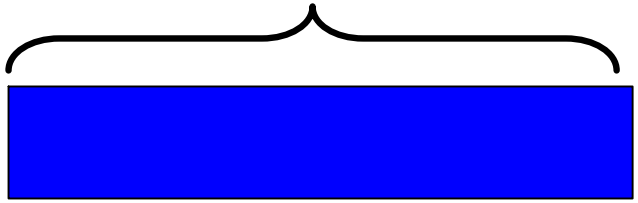
Creating zones

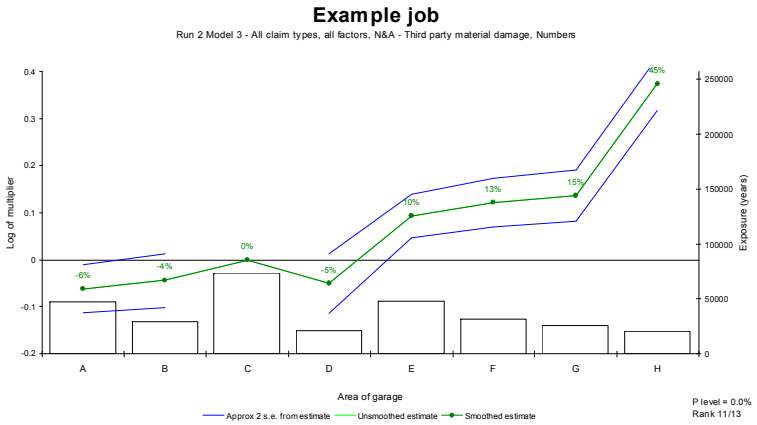
- Equal risk / equal exposure
  - generally mixture works best
- Algorithmically / manually
  - often manual method most pragmatic
- With / without regard to contiguity
  - ignoring contiguity more predictive
  - regulatory or sometime commercial considerations may dictate otherwise

Finding the parameters

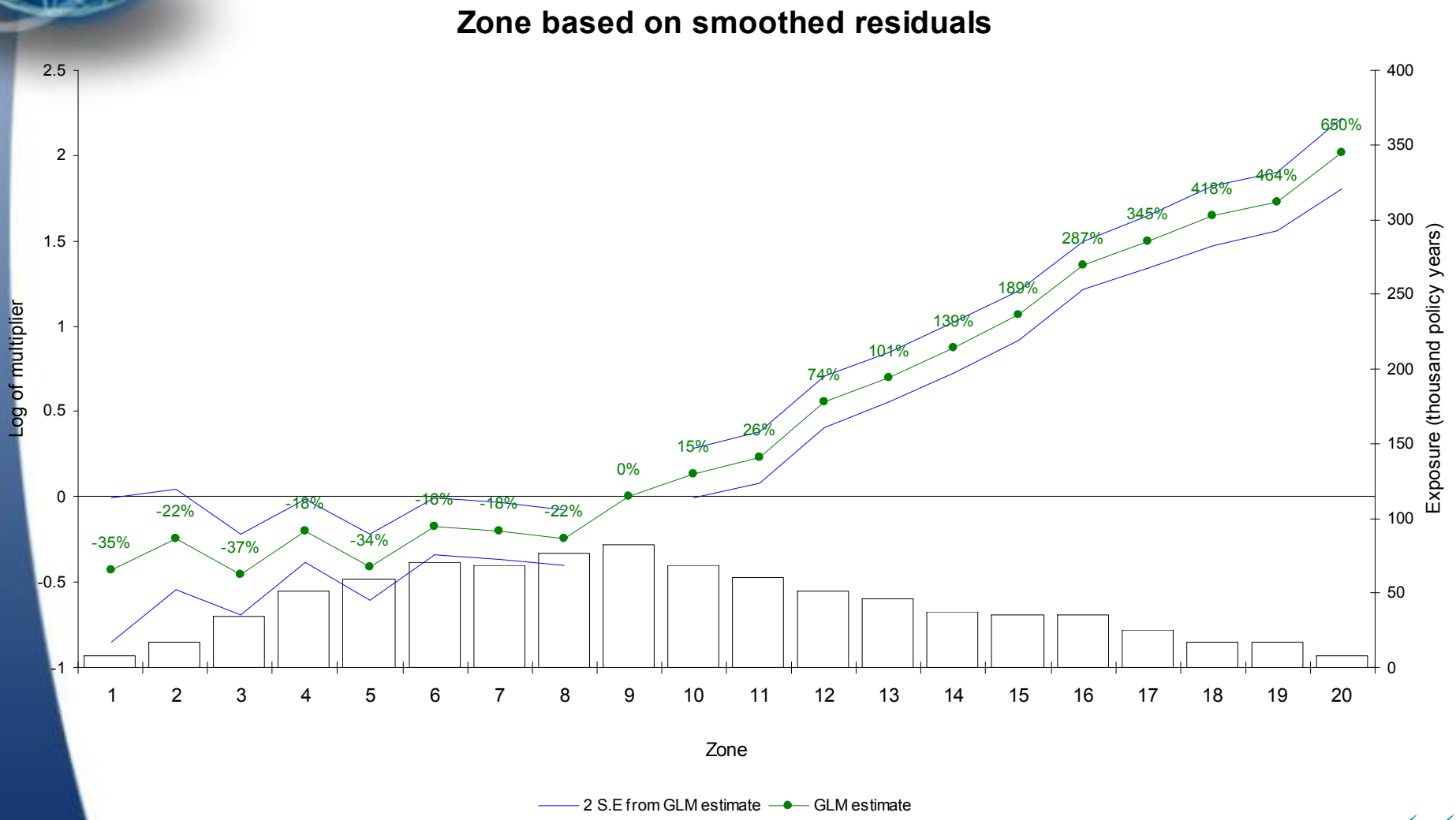
- Fit new zone definition in GLM to assess true predictive power
- Fresh data required to avoid self-fulfilling prophesies
- Compare against existing territory definition

Save for determining zoning relativities



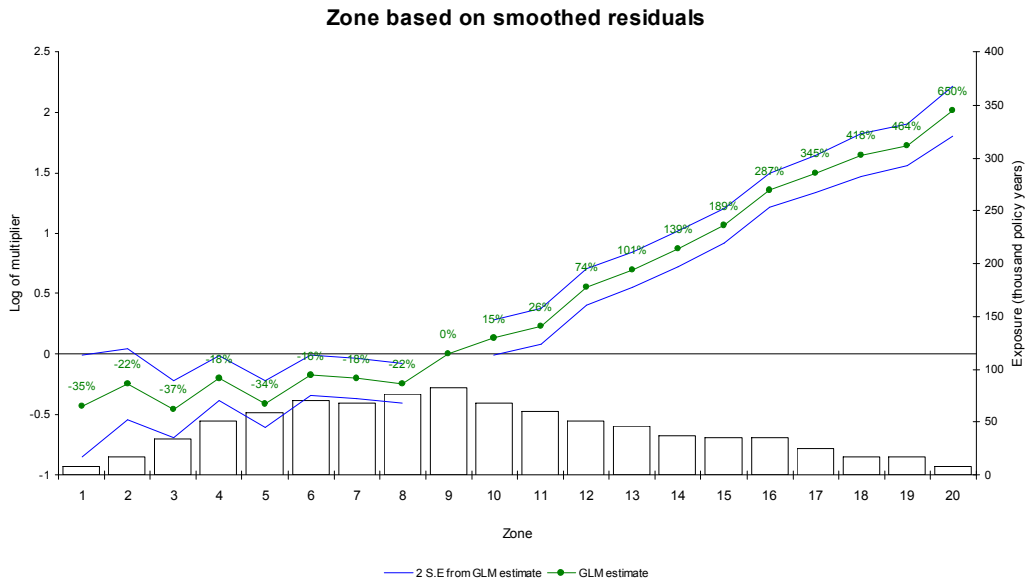


Finding the parameters
Effect of smoothed residual zone on fresh data

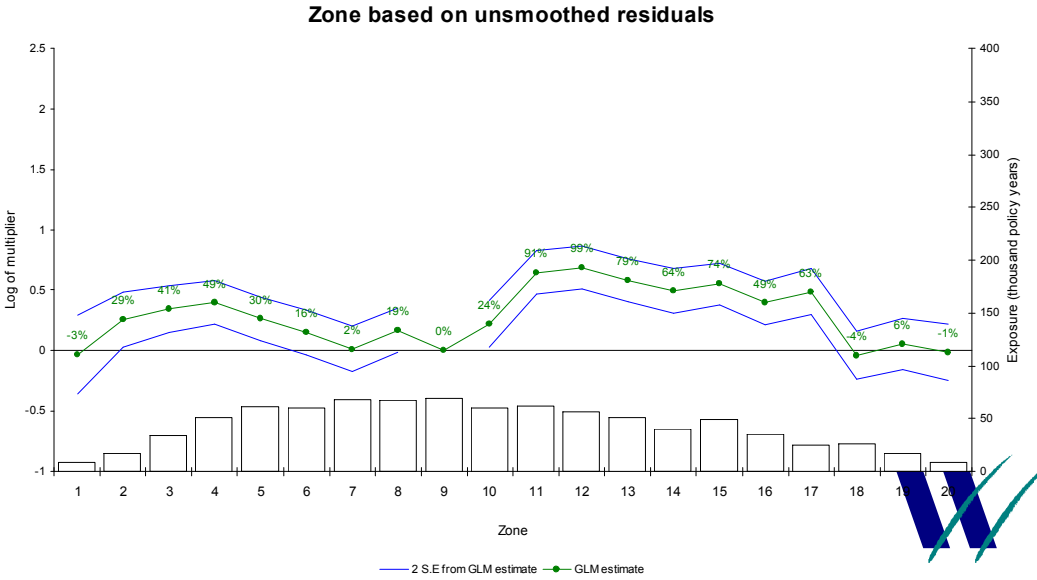


Finding the parameters
Effect of smoothed vs unsmoothed residual zone

Zone based on smoothed residuals



Zone based on unsmoothed residuals

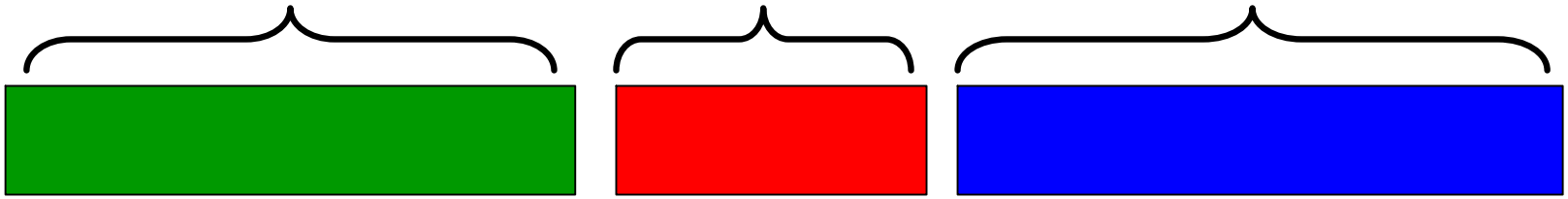


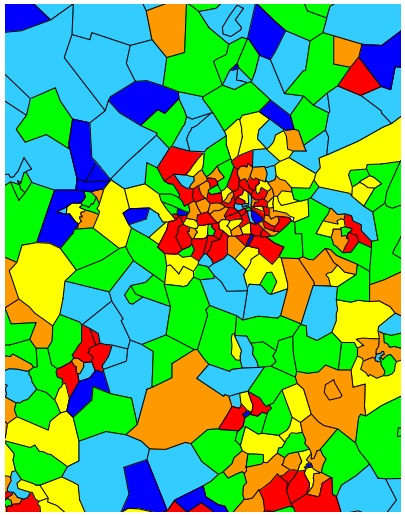
Making use of all the data

Calculate residuals

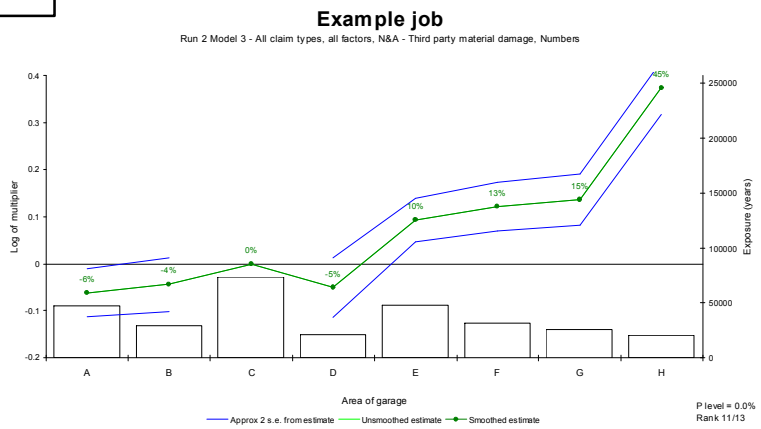
Seek parameters which minimize error

Save for determining zoning relativities





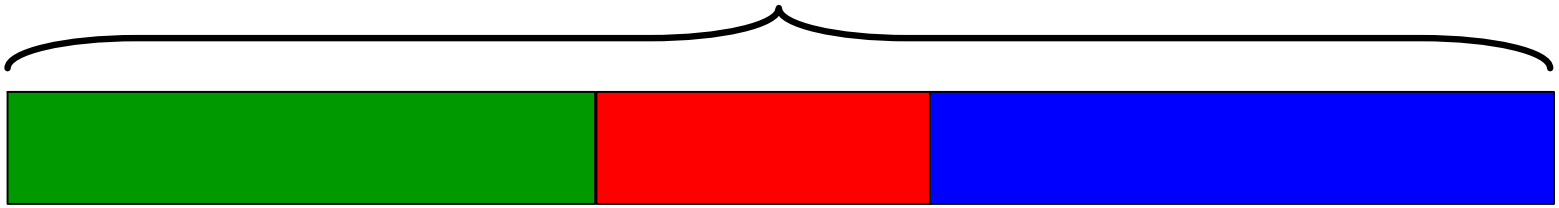
a, m, n, b

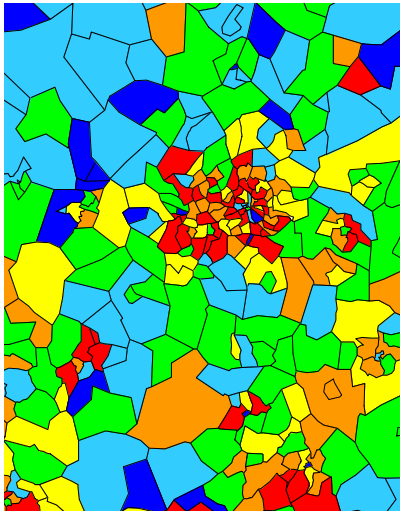


Freeze

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





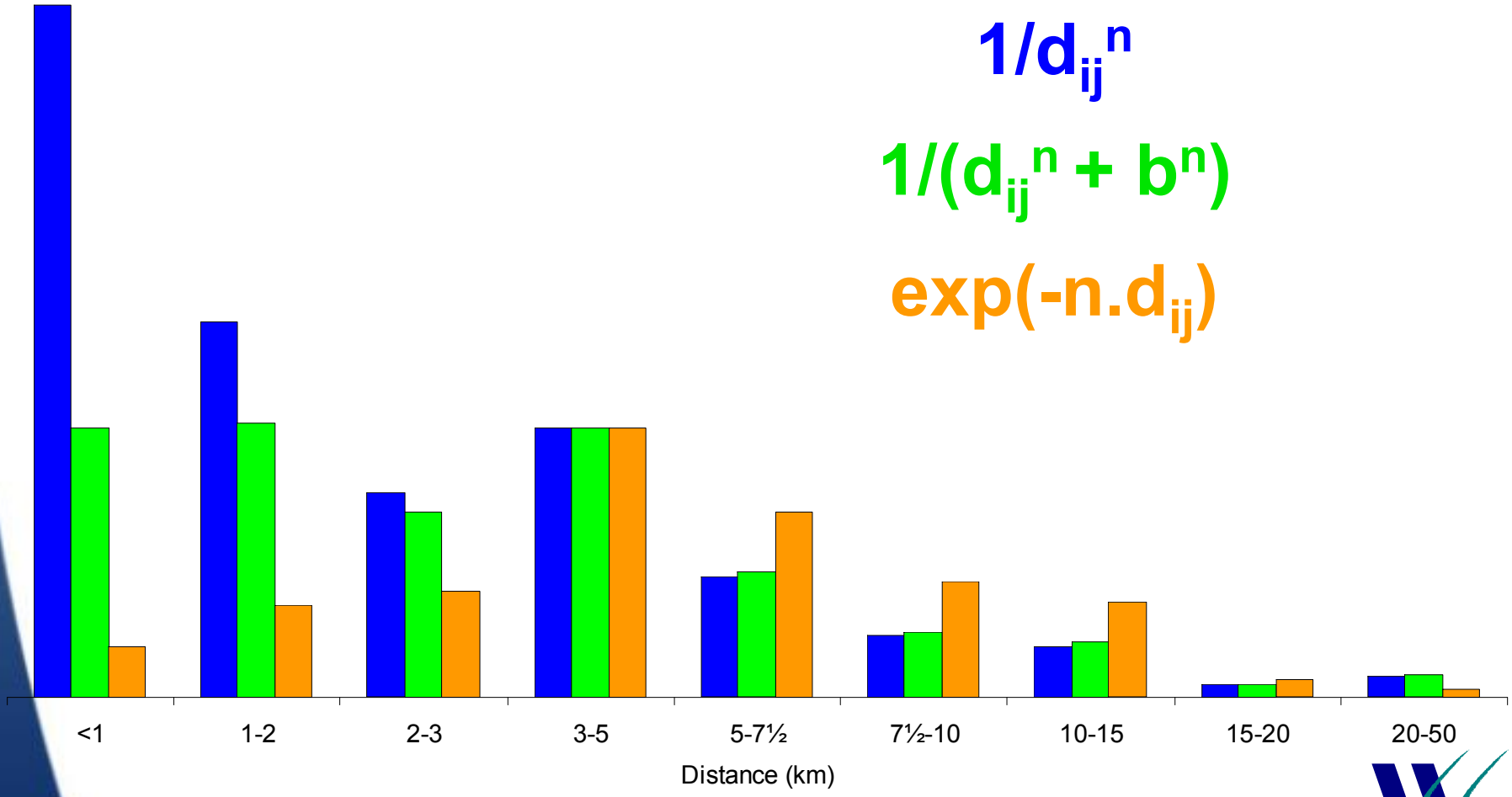
More details

- Data required
- Distance function
- Different distance metrics
  - including other things
  - adjacency
- Limiting the radius of assumed influence
- Contiguity clustering
- Residual approach vs MLE approach

Data required - examples

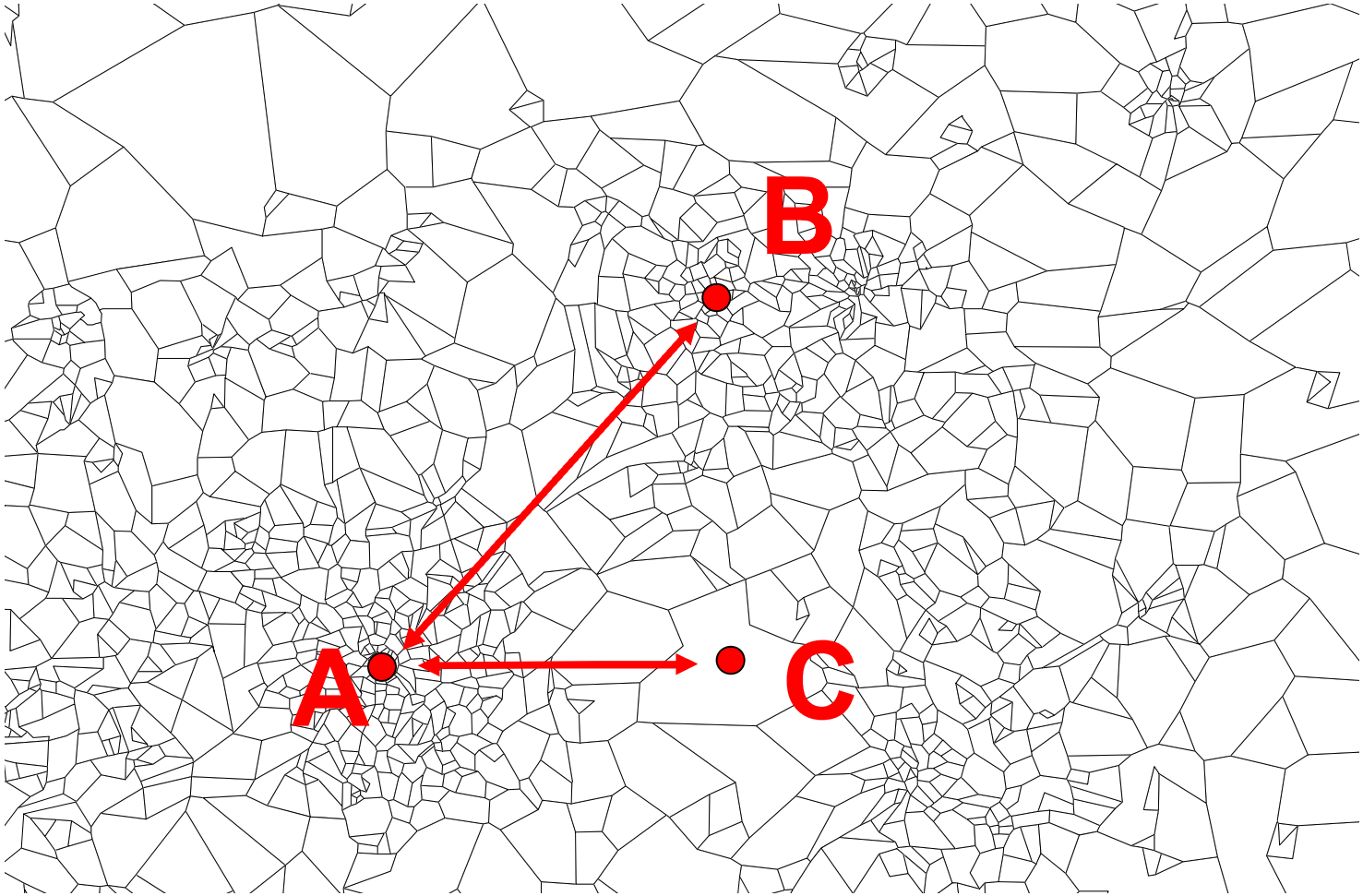
Cannot be disclosed in handout

Different weighting functions
Influence of neighbors in total - example urban area



Different metrics

Is A "closer" to B than C ?



Different metrics

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

where

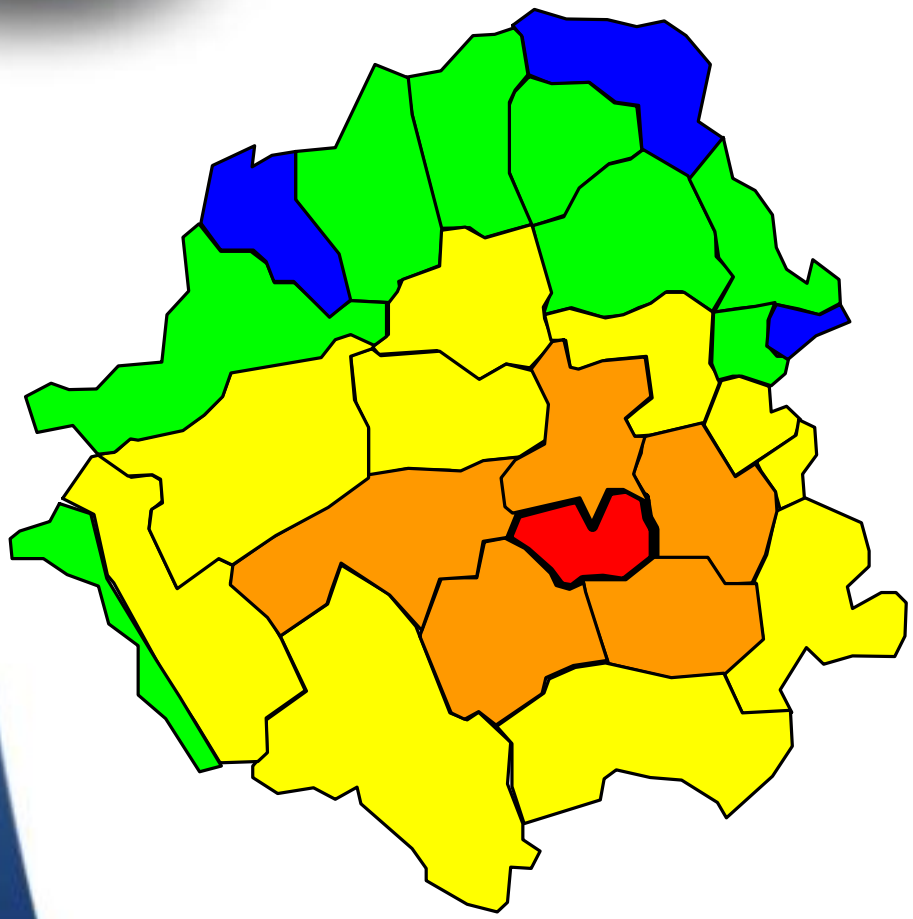
r_i^* = smoothed residual r_i = unsmoothed residual

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

$$d_{ij} = \{ (x_i - x_j)^2 + (y_i - y_j)^2 + (s.q_i - s.q_j)^2 \}^{½}$$

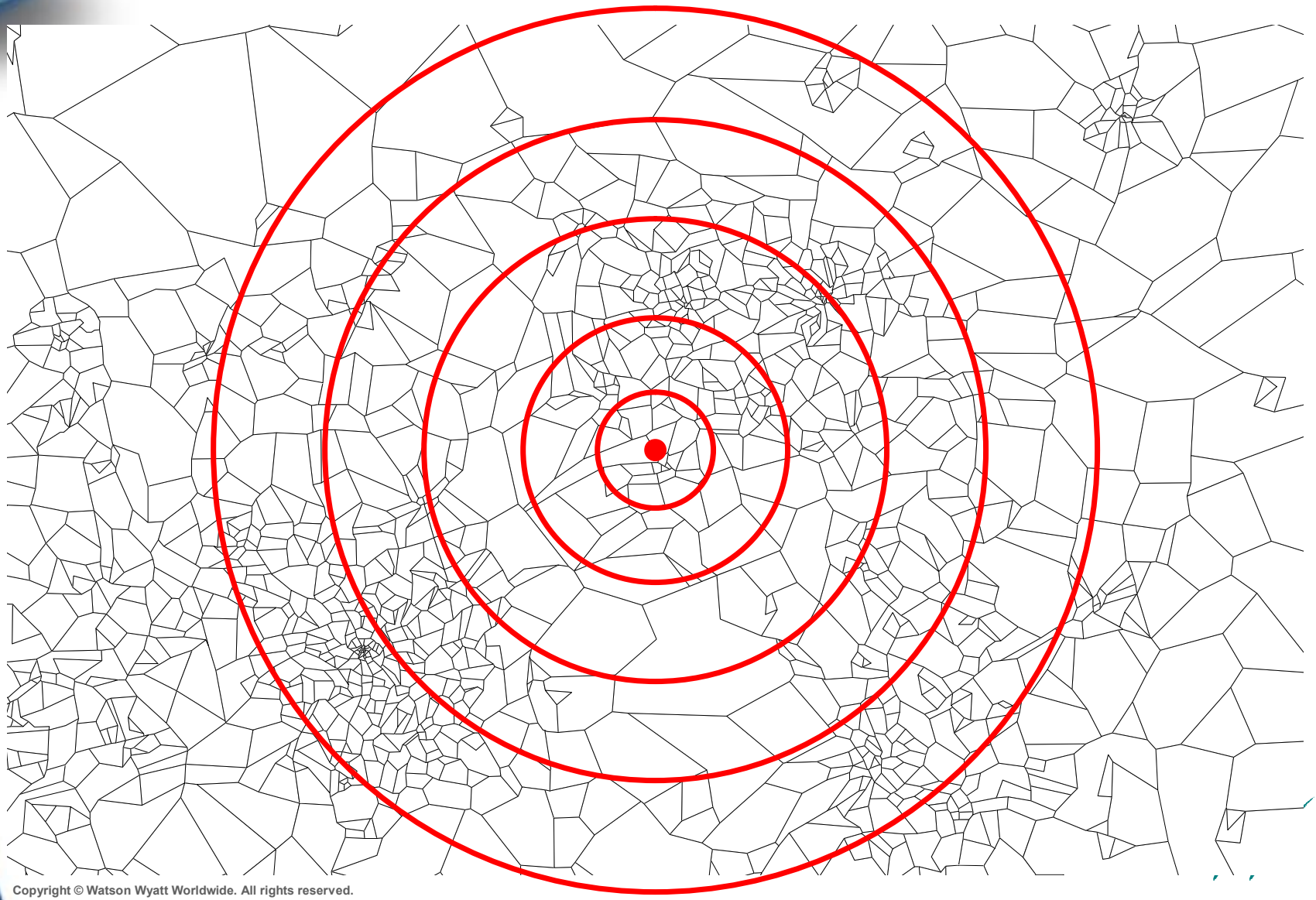
$$f(d_{ij}) = 1/d_{ij}^n \text{ or } 1/(d_{ij}^n + b^n) \text{ or } \exp(-n.d_{ij}) \text{ etc}$$

Adjacency metrics

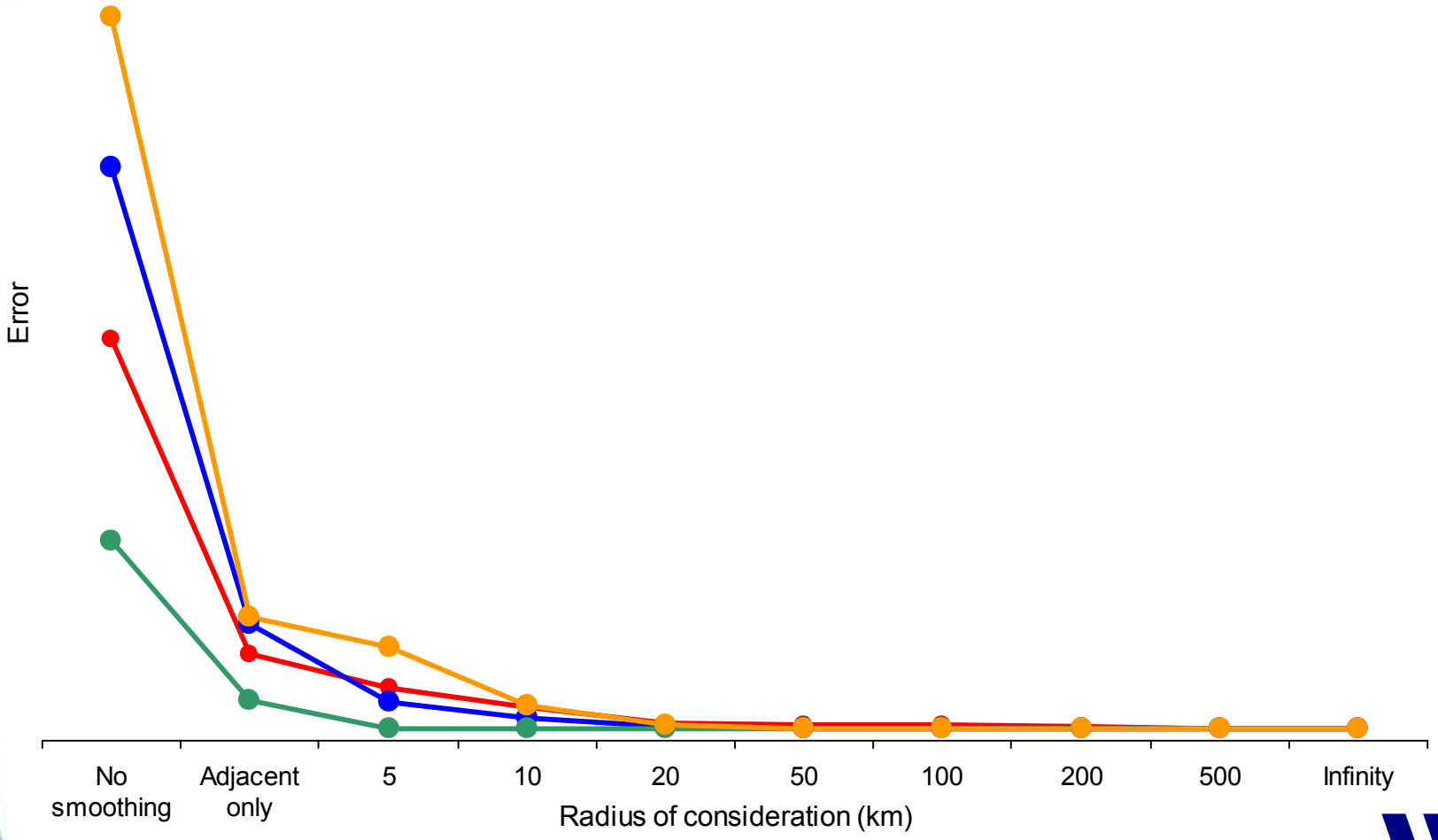


- Define distance by rings of adjacent areas
- Eg $f(d_{ij}) = 1 / t_{ij}^{n}$
  t_{ij} = number of ring
- Can work well for claim types such as theft
- Covers greater distance in rural areas than in urban areas

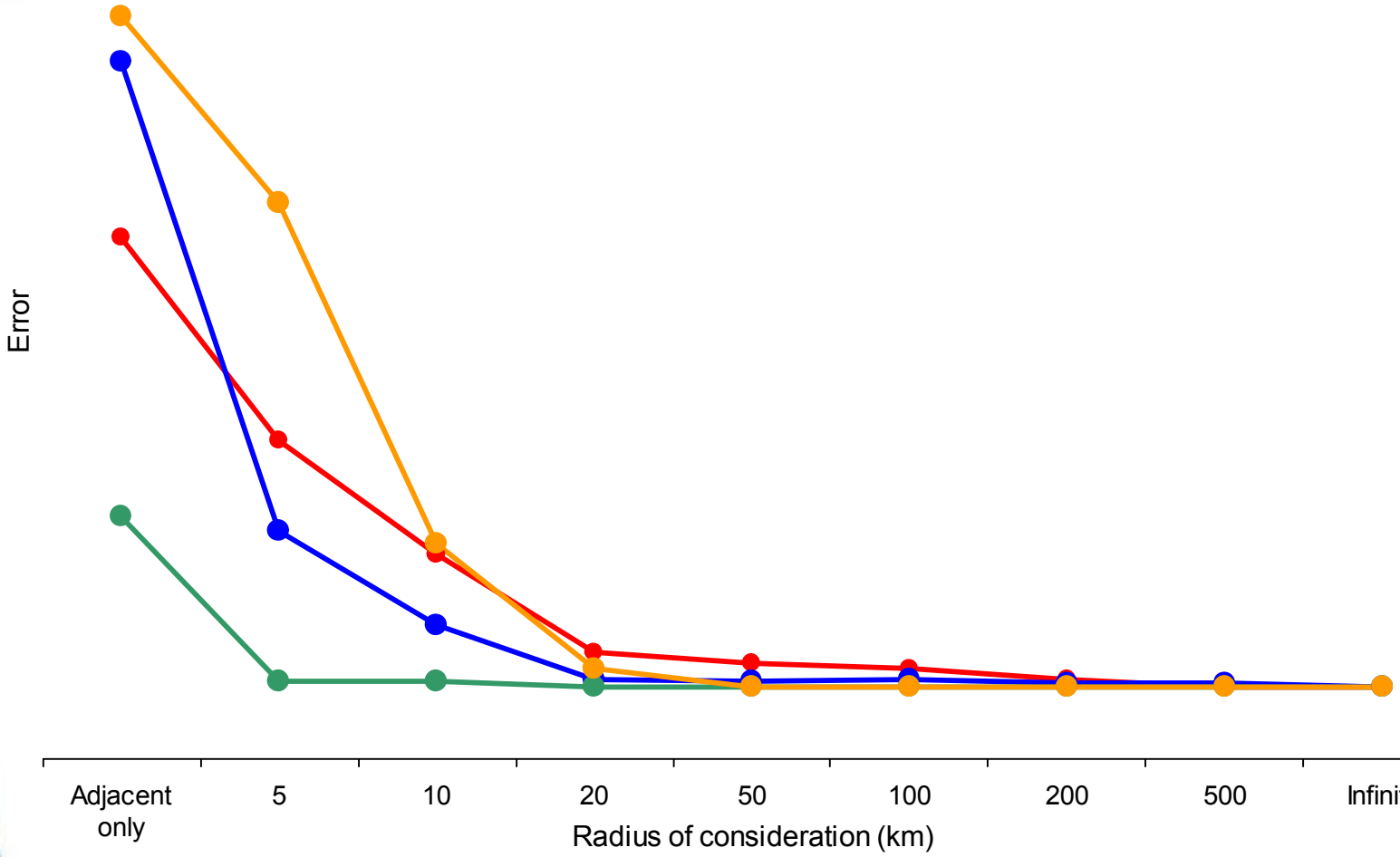
Limiting the definition of "neighboring"



Error with differing radii of consideration

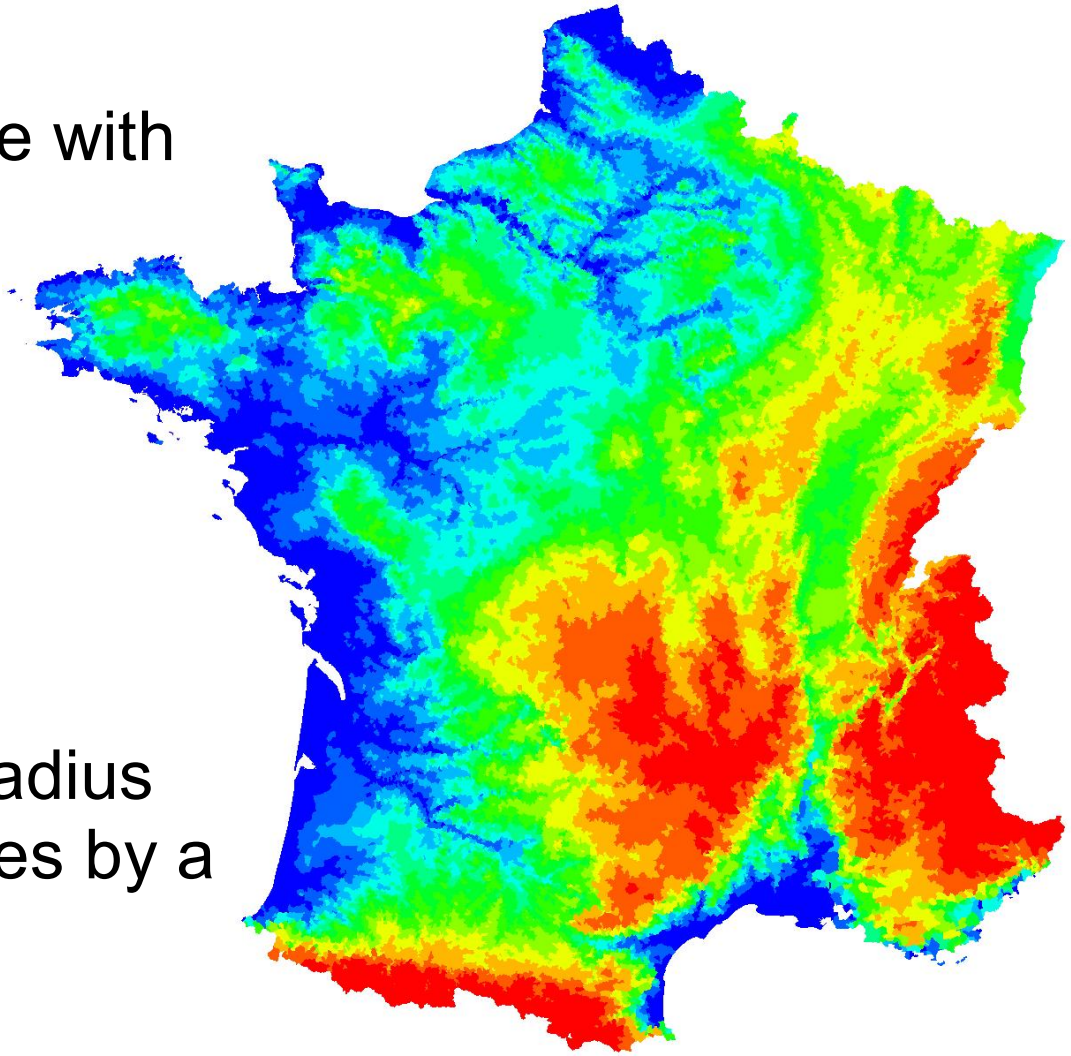


Error with differing radii of consideration



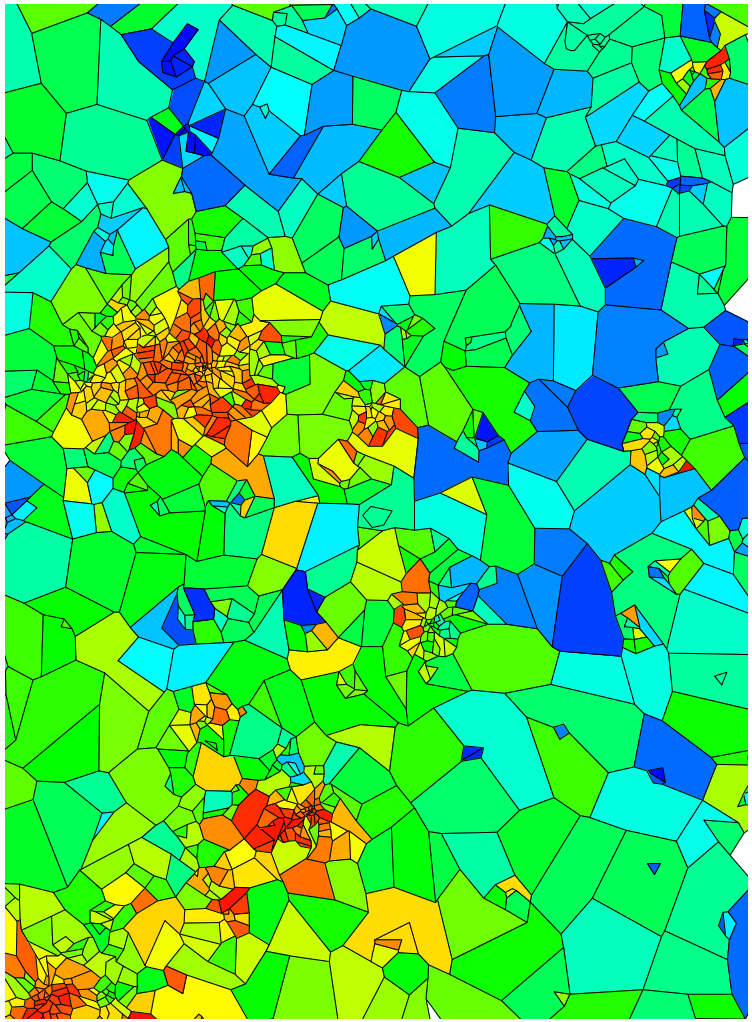
Computational short cuts

- Run times increase with (# 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

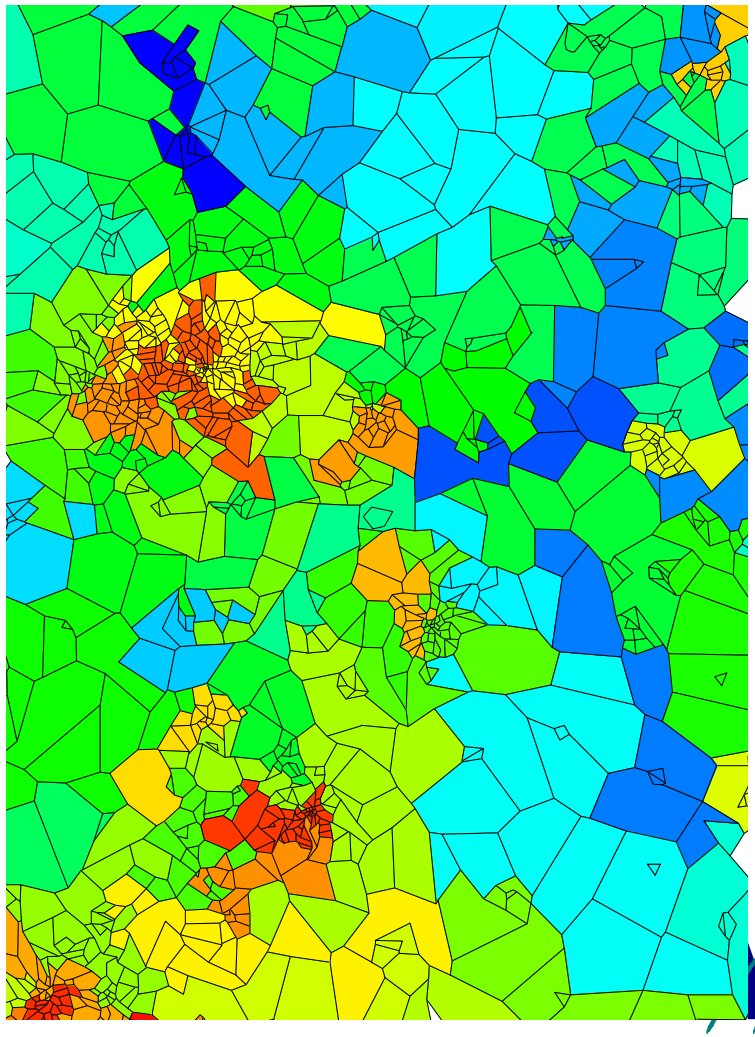


Contiguity clustering

Smoothed residuals



Clustered

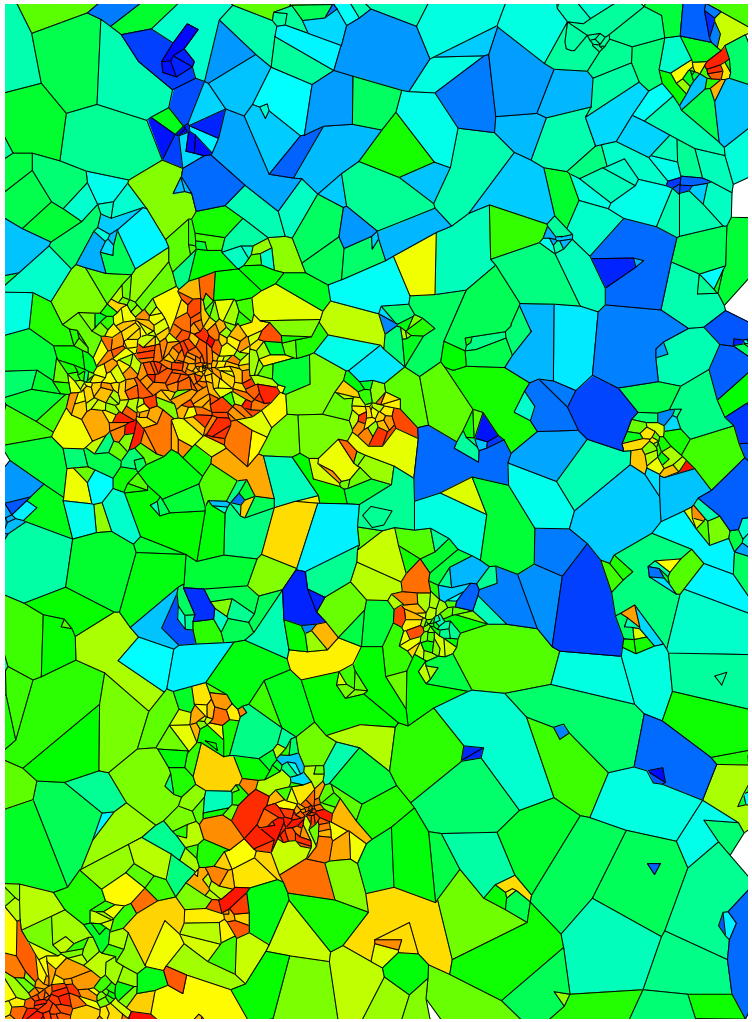


Fitting an MLE surface

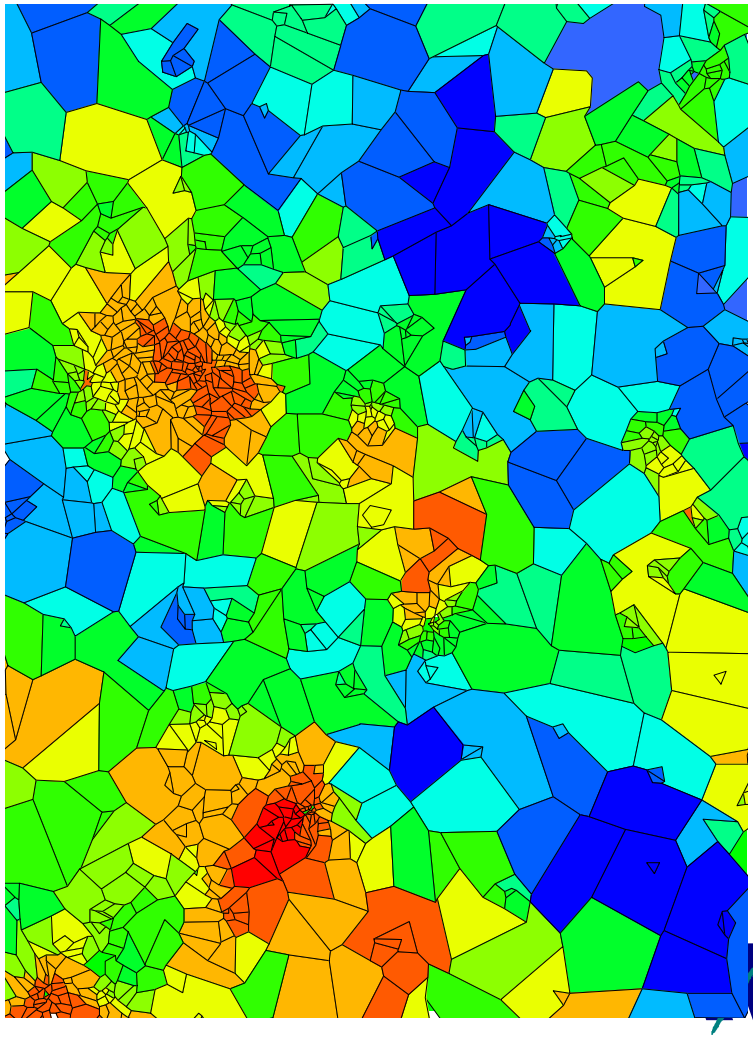
- Fits directly in GLM along with other factors
- Polynomials impractical - splines produce better fits
- Fit as function of x, y, f(x,y)
- "Patchwork quilt" of 2D splines best but computationally challenging

Fitting an MLE surface

Smoothed residuals



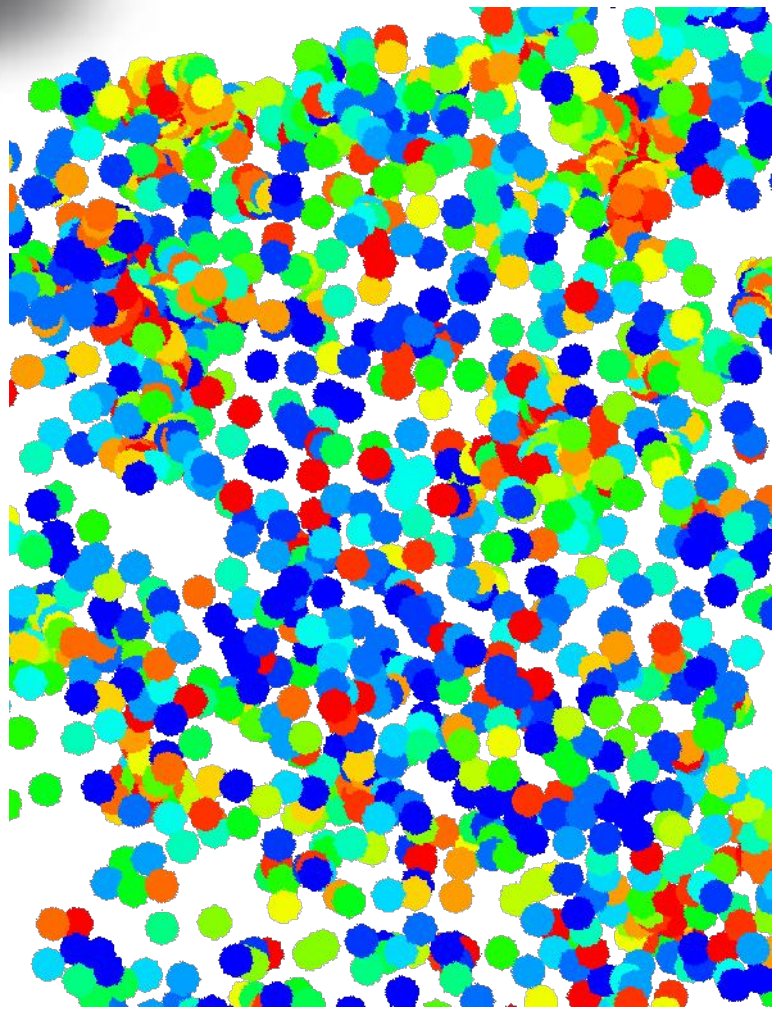
MLE surface

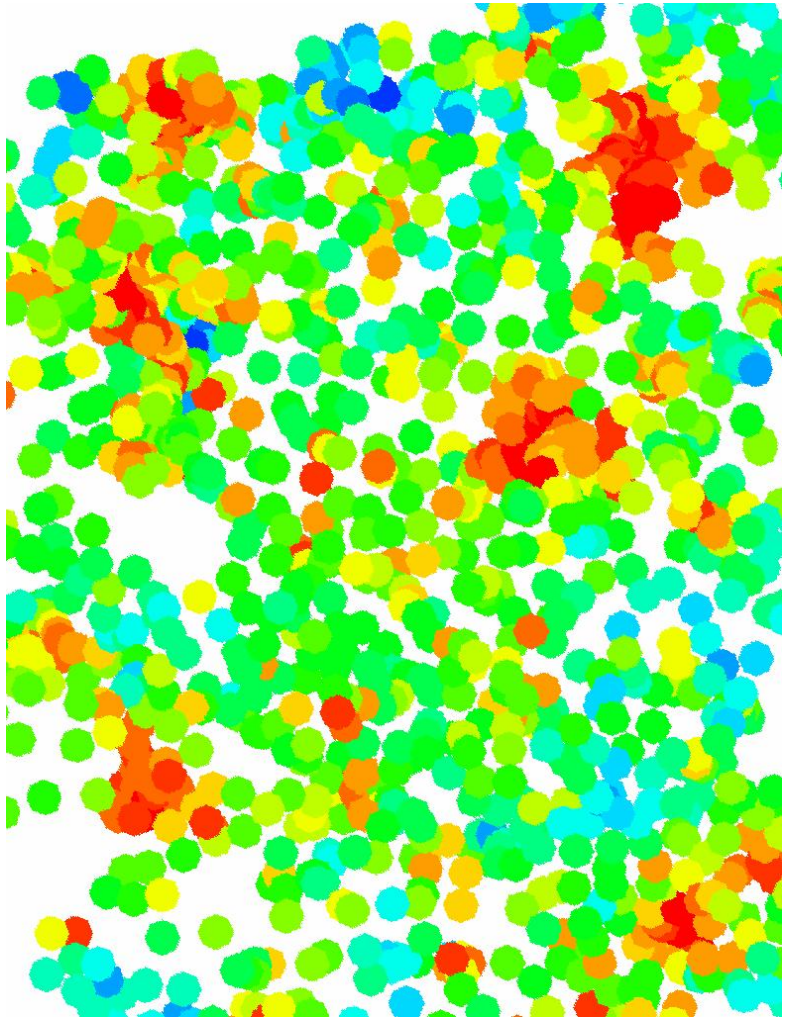


Practical issues

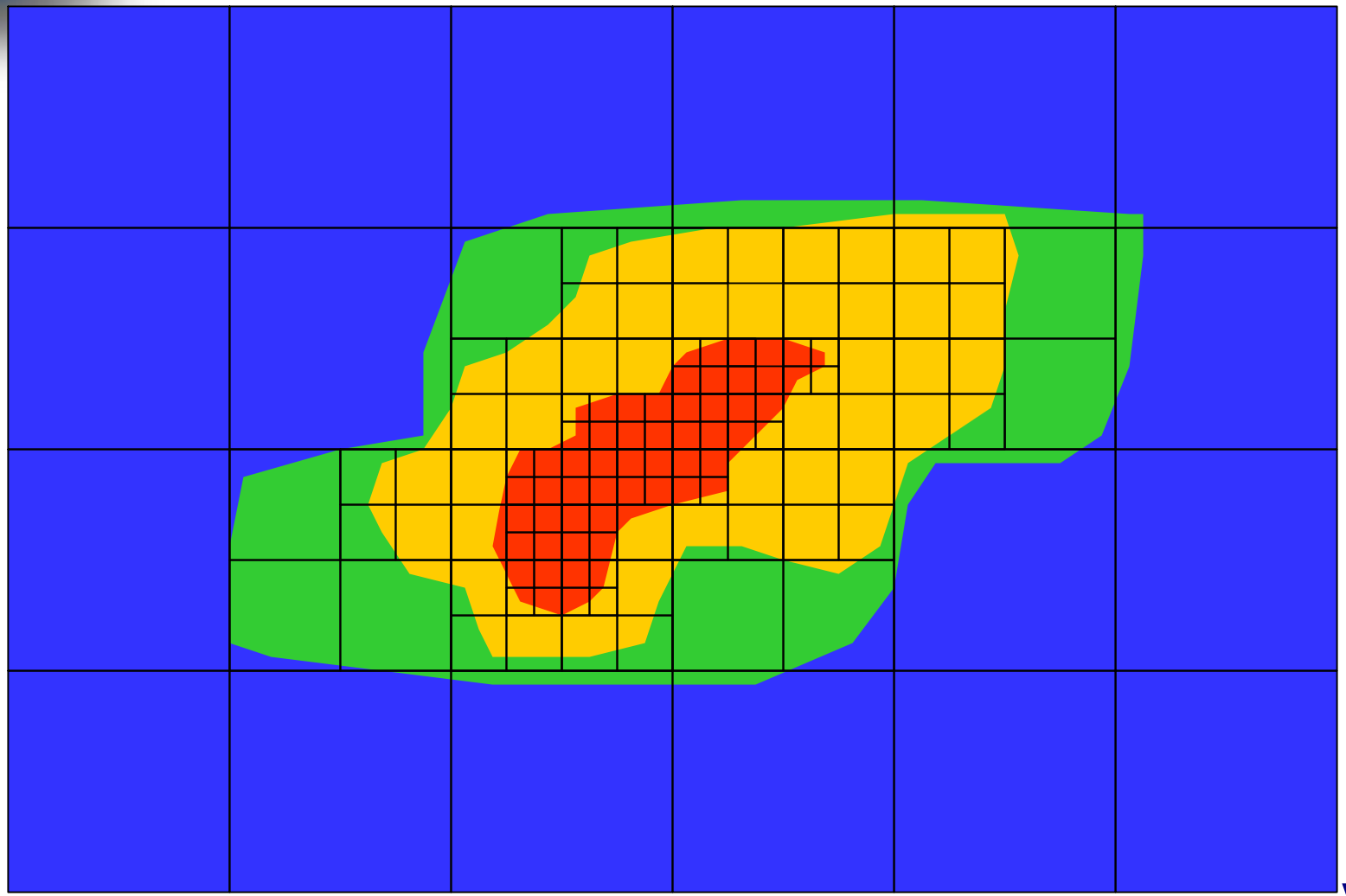
- What to do when there is
  - no boundary data
  - no zip codes
- Geodemographic factors

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





When no zip codes used…



Geodemographic factors

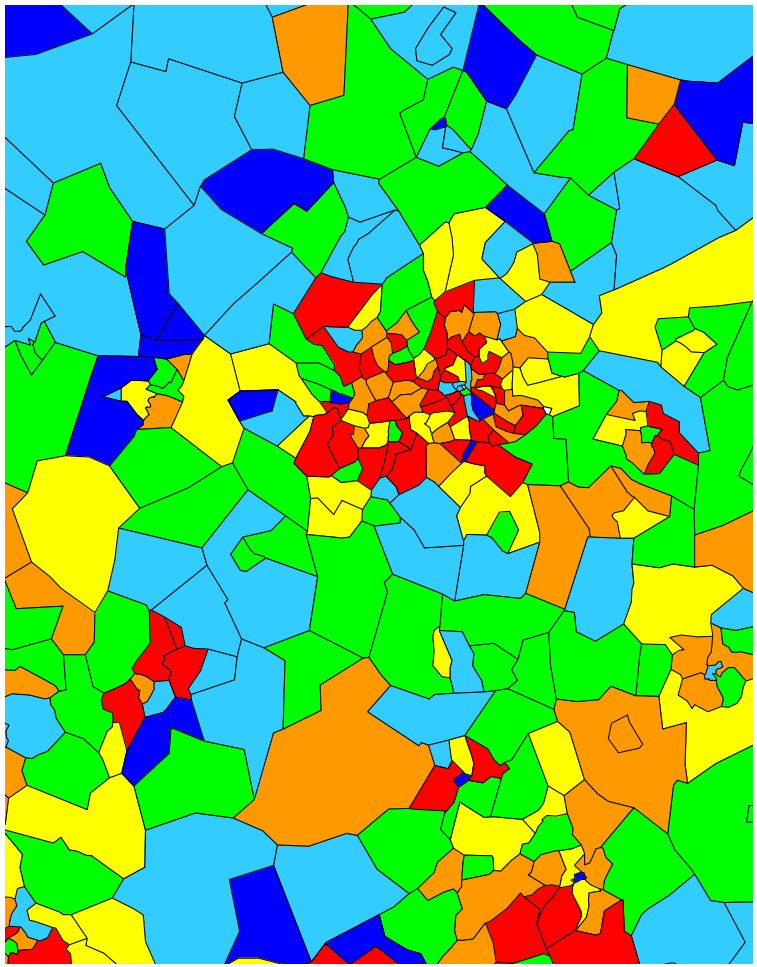
- Can be very predictive
- Even simple measures of urban density can be interesting
- Can be used

  (a) alongside zones derived as above

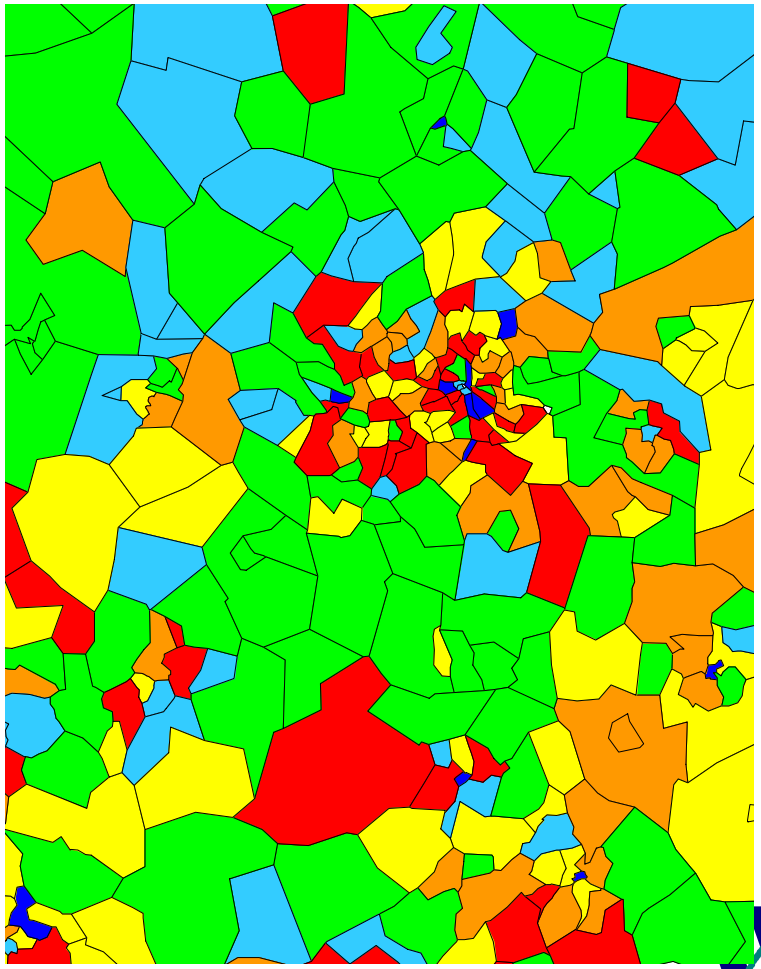
  (b) to standardize experience prior to smoothing
- Investigate which yields most predictive zone
- Generally speaking, seek to standardize for factors which yield inherently smoother residuals

Unsmoothed residuals

Density not in standardizing GLM

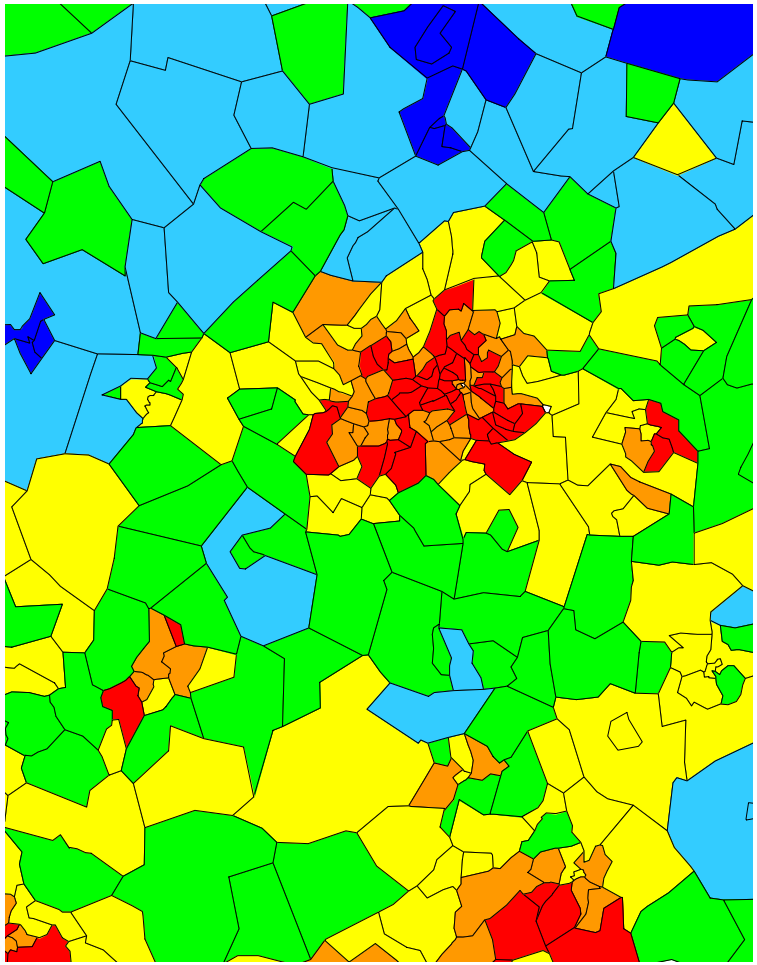


Density in standardizing GLM

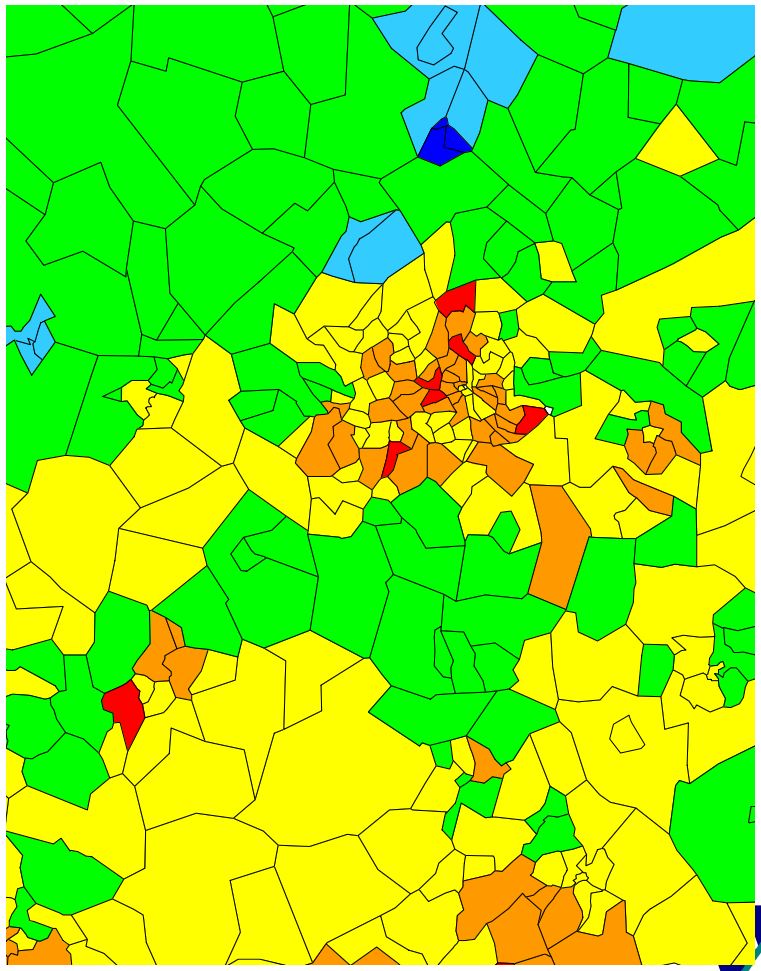


Smoothed residuals

Density not in standardizing GLM



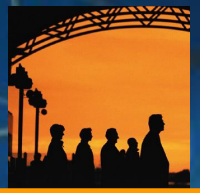
Density in standardizing GLM



PL-7

Putting Your Company on the Map:

Determination of Statistically Indicated Territory Boundaries


WWW.WATSONWYATT.COM

2006 CAS Seminar on Ratemaking

Duncan Anderson MA FIA

Watson Wyatt Worldwide

Watson Wyatt Worldwide