On the geography of road accidents. Challenges and opportunities.

Isabelle THOMAS

Nov 24th 2015

Hong Kong

Geography

Point of view (unique)

- **(place** and **space)** x time

  Spatial variation, distribution, diffusion….

Objects (shared with other disciplines)

WHAT IS WHERE ? WHY THERE ? WHY CARE ?

Languages : several (maps > models)

What does geography measure?

LOCATION

PLACE (ATTRIBUTES)

(Physical, human; points, lines, S, Vol)

INTERACTIONS

(environment; people; places)

X TIME
X CULTURE

Measures/indices about

PEOPLE

PLACES

INTERACTIONS

ESDA

DESCRIPTION

Spatial STATISTICS

Statistical MAPS

Modeling

Spatial statistical analysis and hypothesis testing

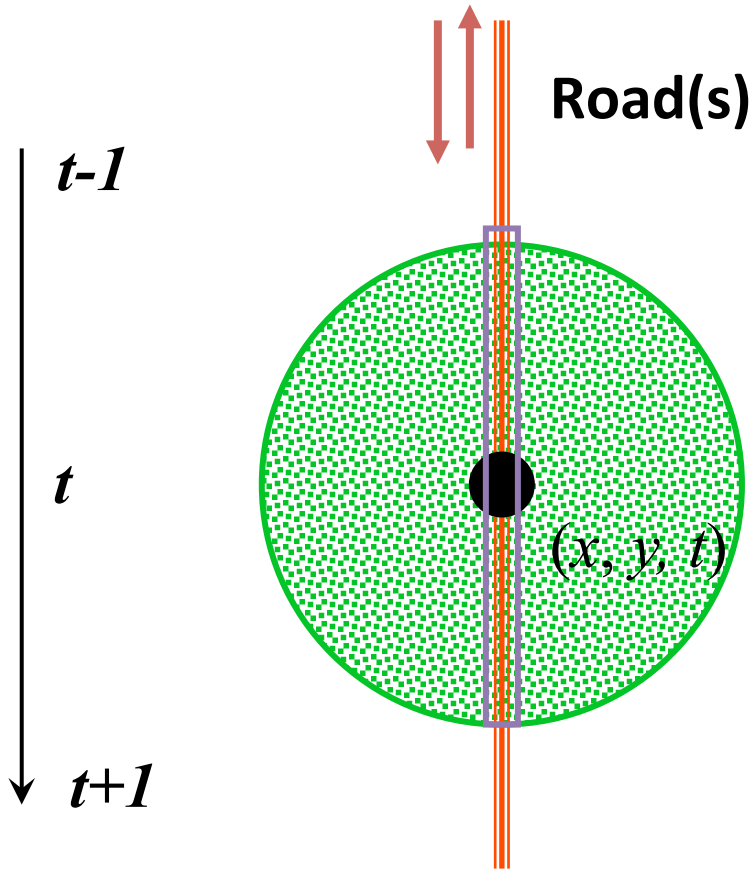
(Spatial) **modeling** and prediction

Statistical/mechanistic

LEVEL OF DIFFICULTY

Spatial is special

Complex events resulting from human, technical & environmental factors



PEOPLE
(*Roadusers*):
VEHICLE(S)

INTERACTIONS

PLACE(S)
(*Environment*)

(x, y)

- **LOCATION**
  - **Absolute** — *Latitude, longitude; an address*
  - **Relative** — *Distance, directions to other places*

Absolute LOCATION

(x, y)

Record, Process

View, Disseminate

GEOCODED ATTRIBUTES OF OUR ENVIRONMENT

Knowing where things are
in relation to other things.

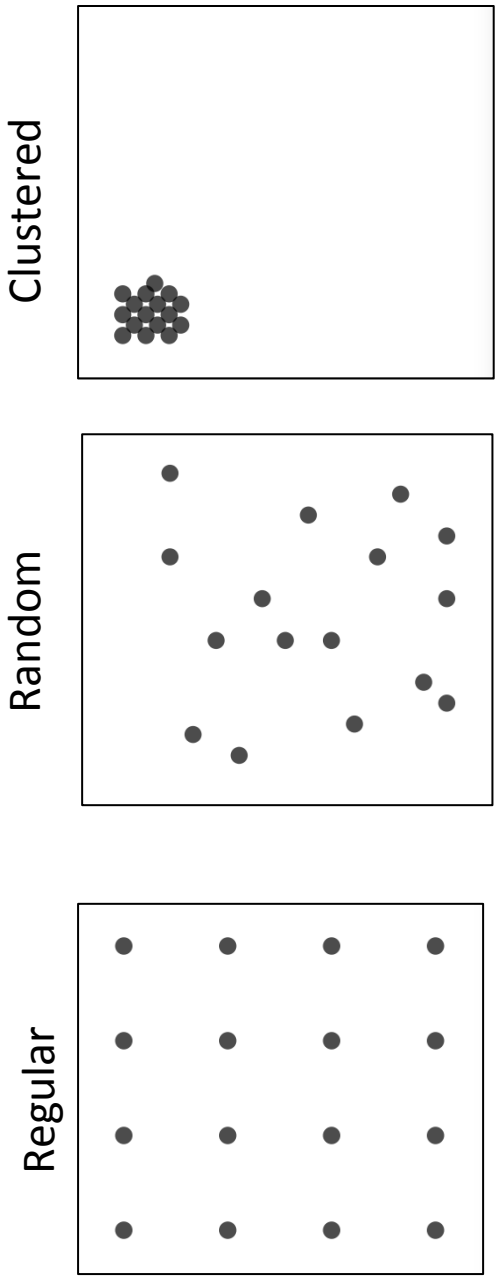
Better geography ?
Taking more informed decisions ?

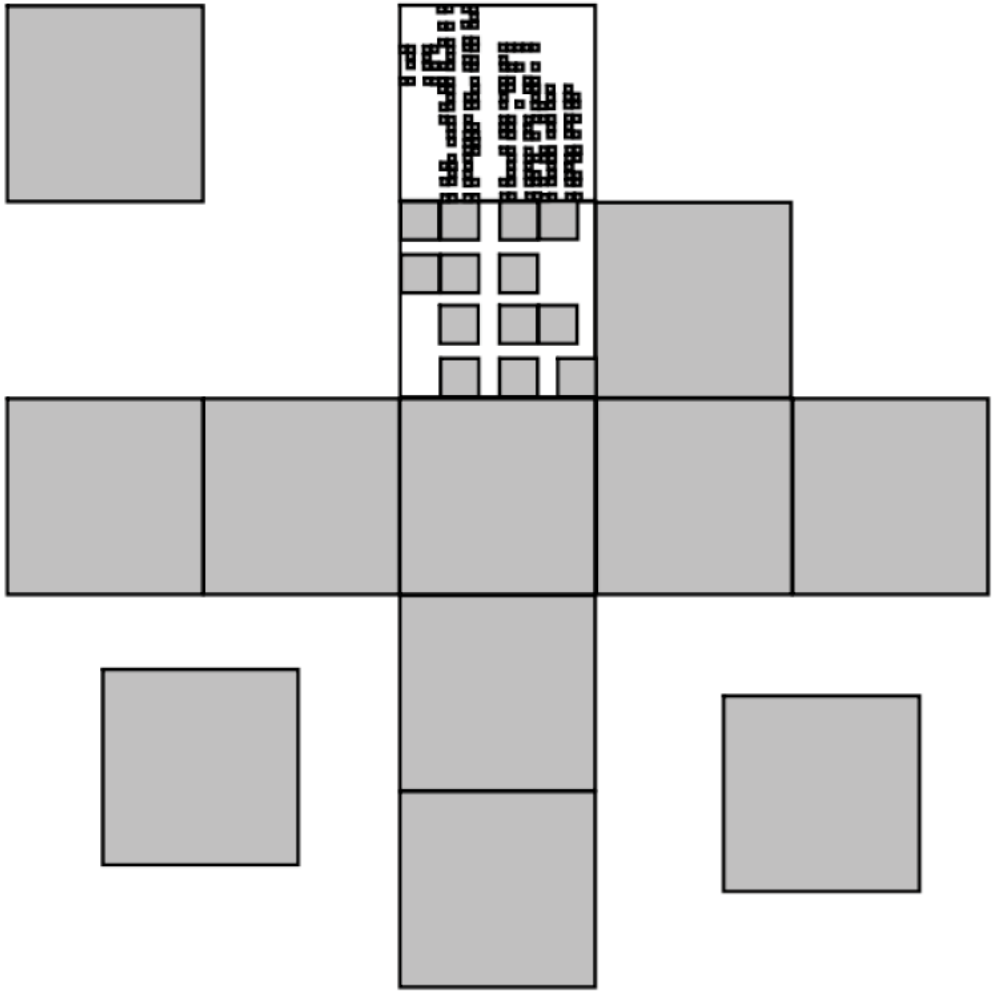
Point pattern analysis

Simplest data: point locations

How to quantitatively describe ?
MORPHOMETRY

Location + distance + direction
Points, lines, surfaces ?





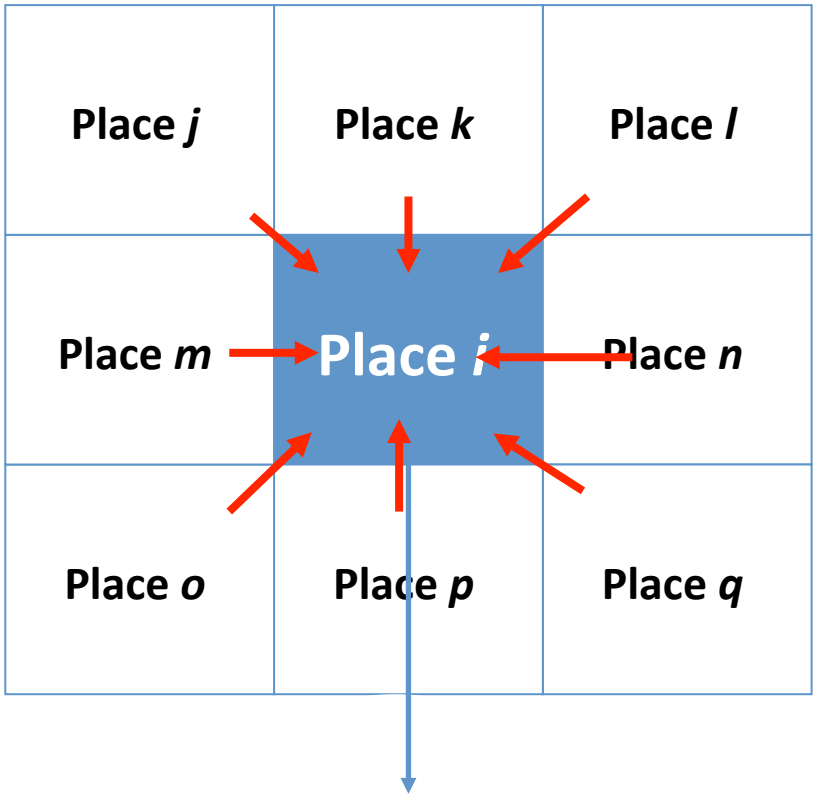
Source : Tannier et al., 2006

Challenge 1
Morphometry

- **RAcc**

- Environment (Built-up S, 3D)
- Non-built-up (green/blue)
- Networks (Access-ability)
- Borders:
  characterizing and extracting

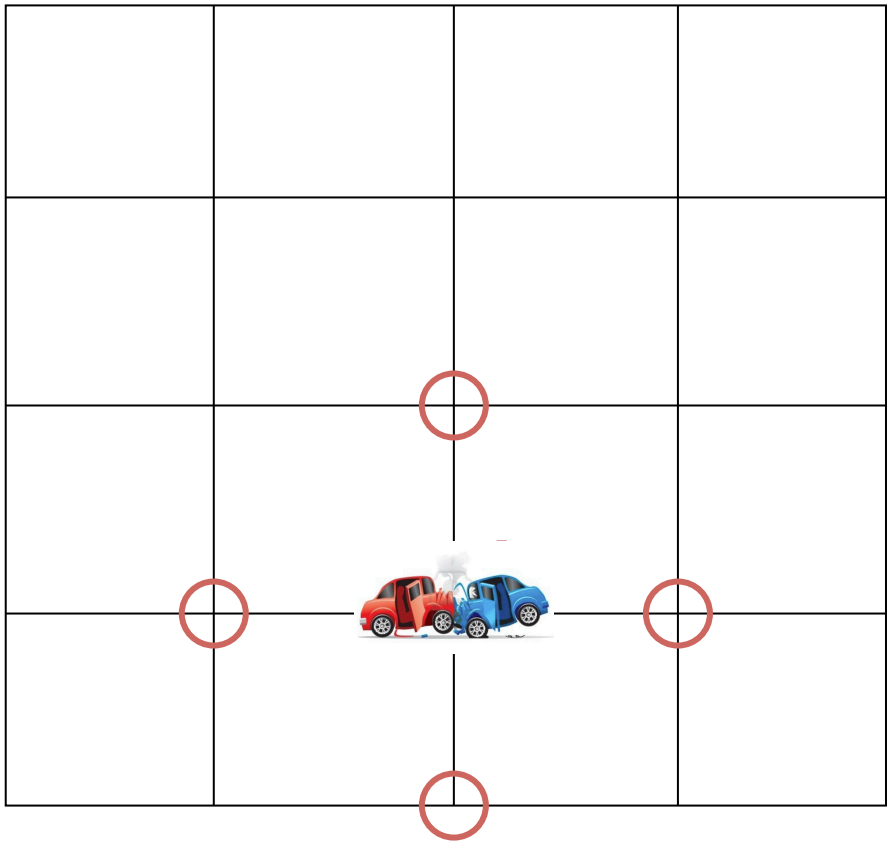
You cannot isolate a place



First Law of Geography
(**Tobler**)

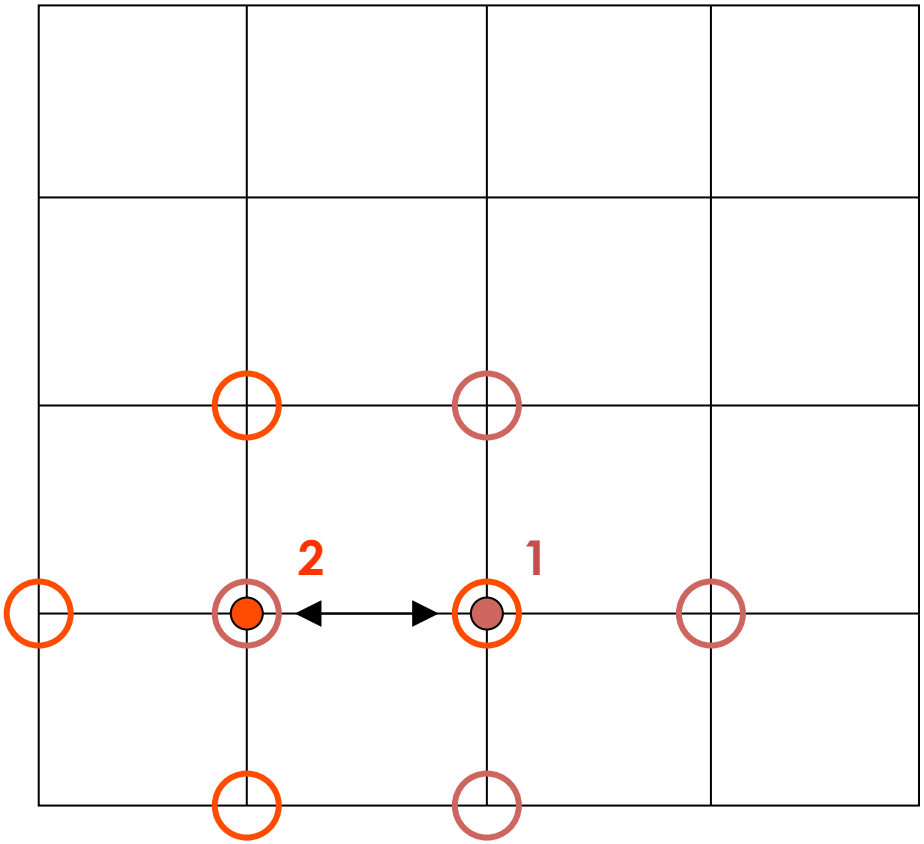
After : GISPopSci

Challenge 2
Close things



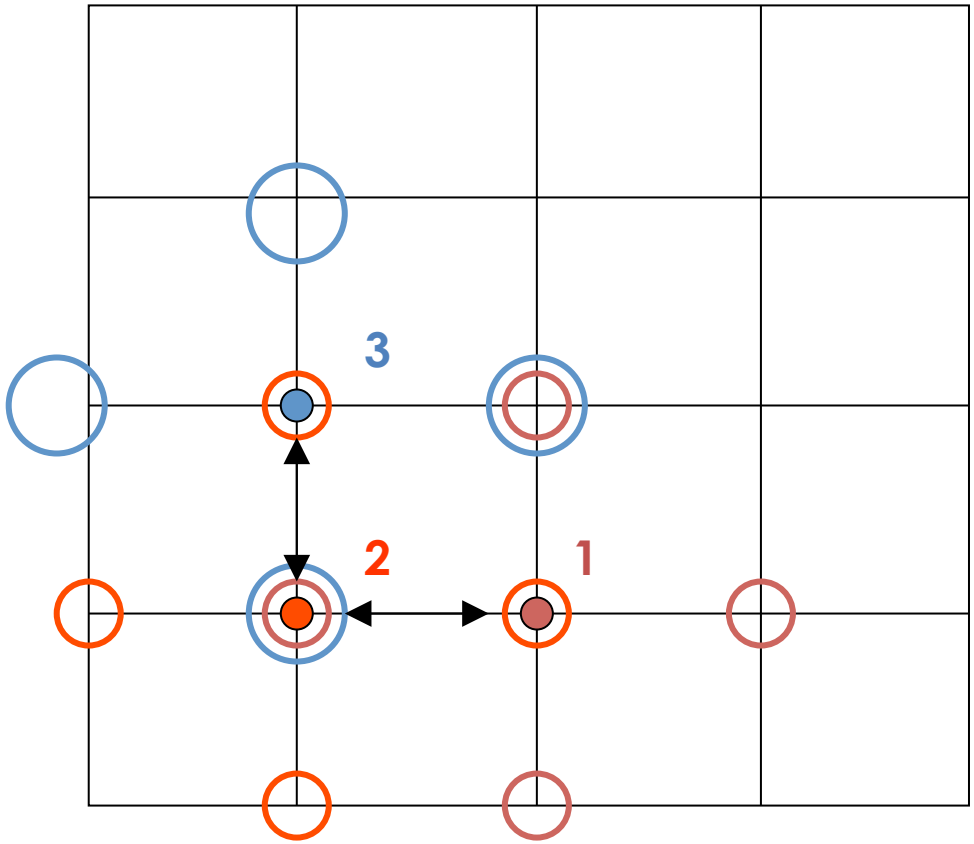
ce : S. Gibbons

Spatial autocorrelation (intuitive)



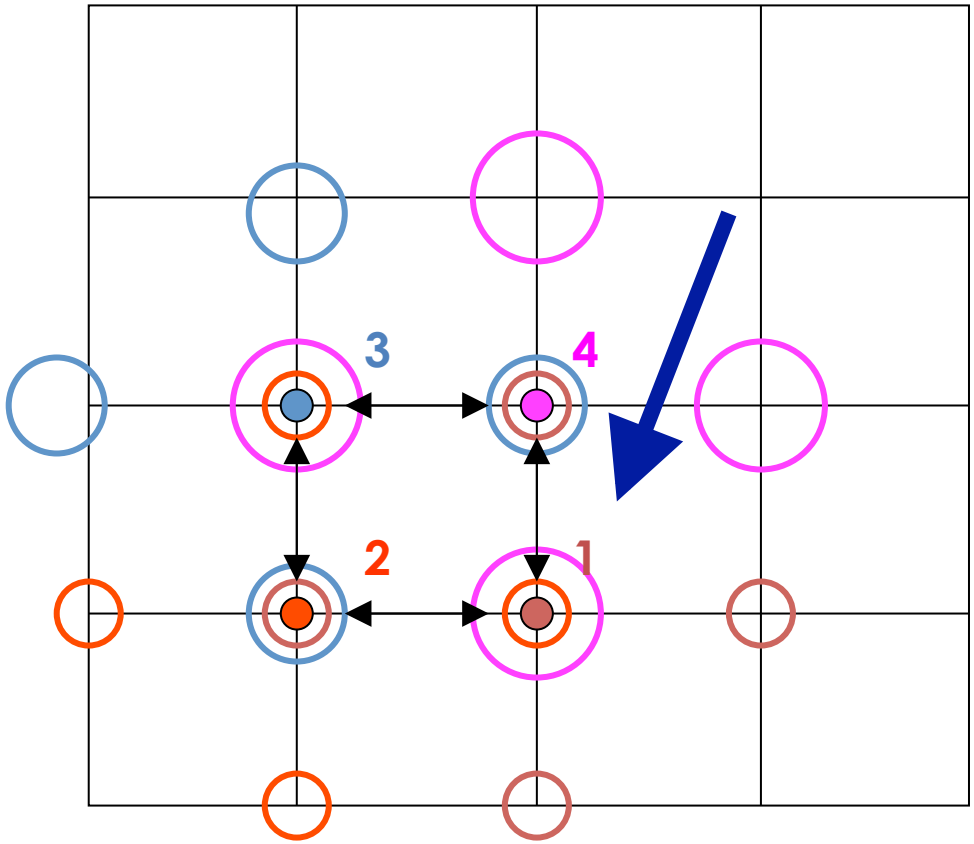
ce : S. Gibbons

Spatial autocorrelation (intuitive)



ce : S. Gibbons

Spatial autocorrelation (intuitive)



ce : S. Gibbons

Spatial autocorrelation X

≠ Spatial correlation $X - Y$

Statistical models : observations should be independent.

Controlling the problem can change equations and interpretation. Ignoring it **= biasing.**

What is Spatial autocorrelation
D. Griffith, 1992 – L'Esp. Géo.

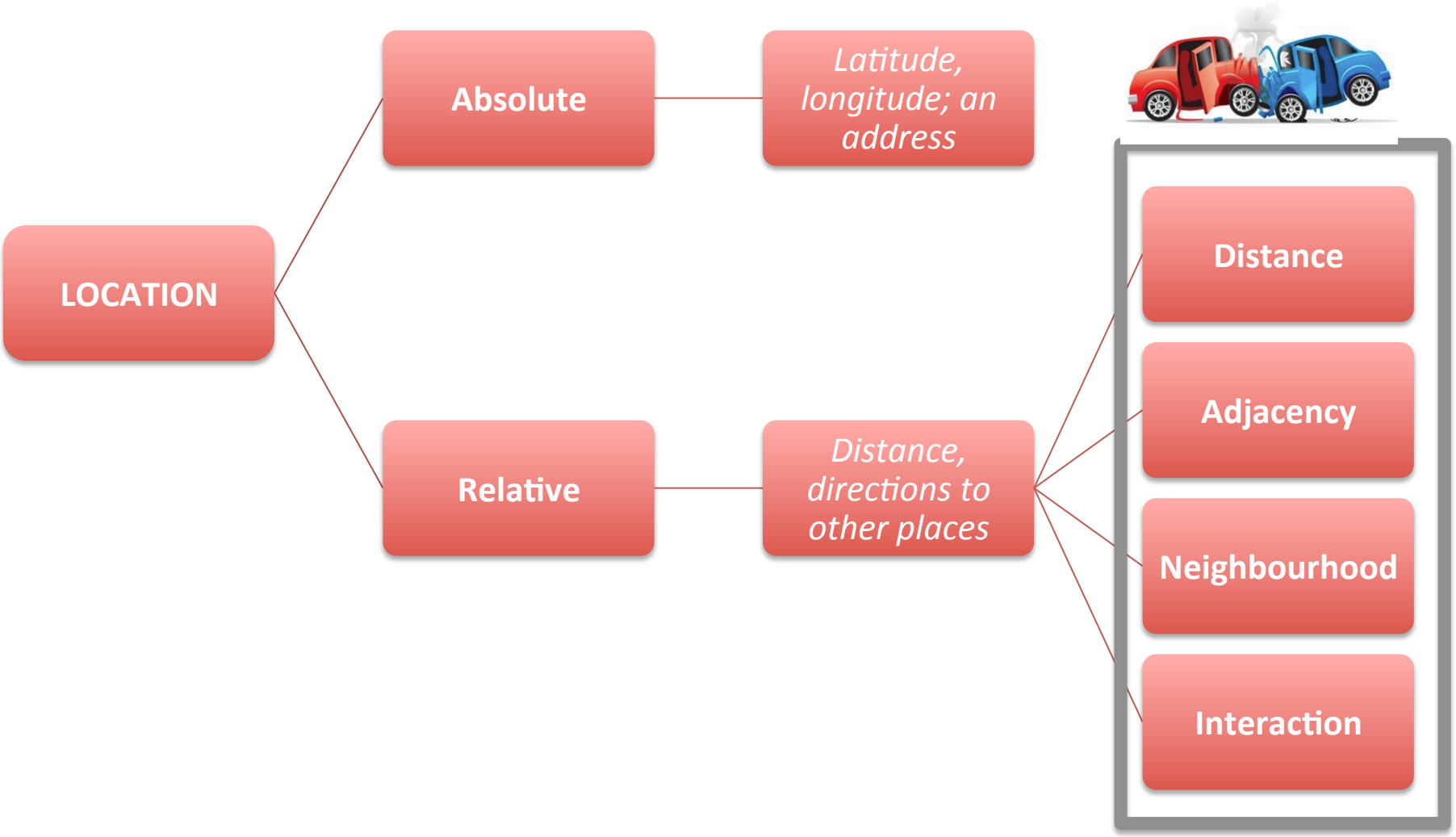
- **Moran's *I***

$$I = \frac{1}{p} \frac{\sum_i \sum_j w_{ij}(z_i - \bar{z})(z_j - \bar{z})}{\sum_i (z_i - \bar{z})^2}, \text{ where } p = \sum_i \sum_j w_{ij} / n$$

- **Extending SA concepts**
  - Neighbourhoods - lag models
  - Disaggregation (LISA)
  - Tests
  - Distances(…) instead of weights

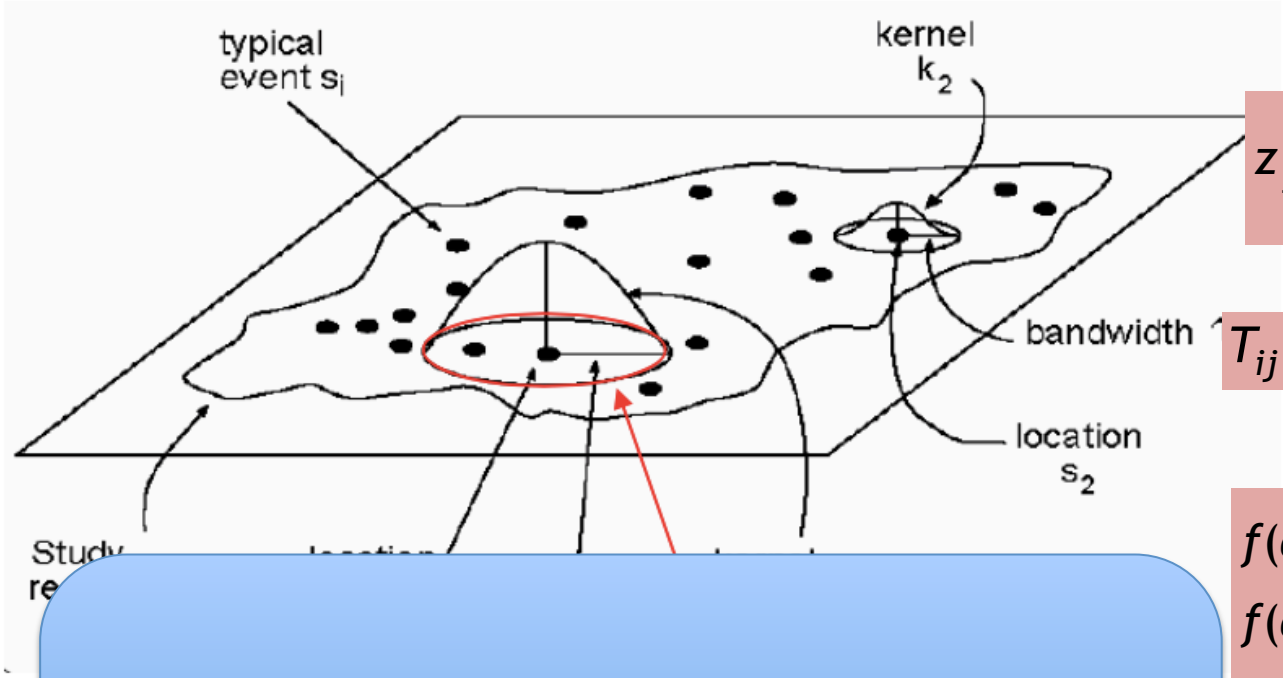
www.spatialanalysisonline.com

Challenge 3
Measuring distance



« Black zone » = ?

Kernel estimation



$$z_j = \frac{f(\{z_i\})}{{d_{ij}}^{\beta}}, \beta \geq 0$$

$$T_{ij} = A_i B_j O_i D_j f(d_{ij})$$

$$f(d) = e^{-d^2/2h^2}, or$$

$$f(d) = e^{-d/h}, or$$

$$f(d) = \left(1 - \frac{d^2}{h^2}\right)^2, d < r$$

$$f(d) = 0 \text{ otherwise}$$

OD ? 1D? 2D? 3D? Distance decay function ?

i = ? *(point ?Centroid?)*

Source : Ovtracht, 2014

Challenge 4
Scale

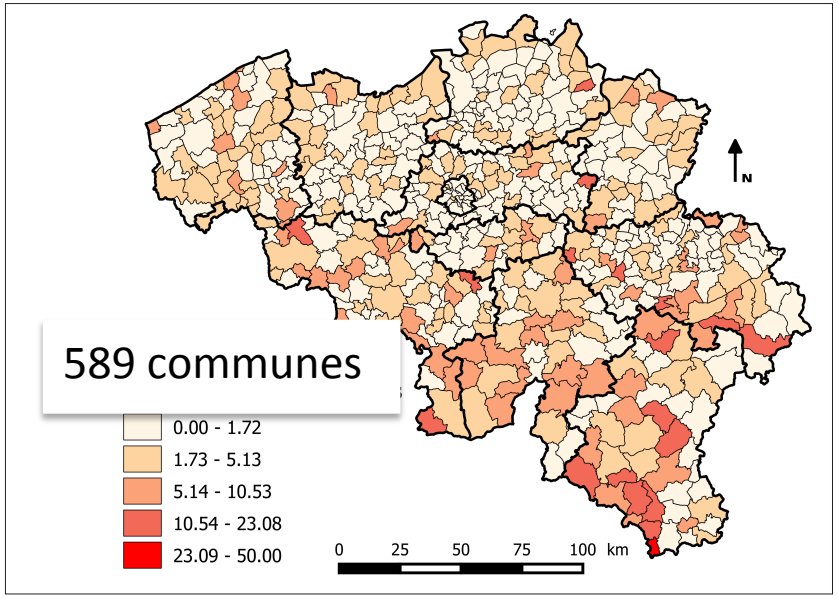
LOCATION — *Absolute/ Relative*

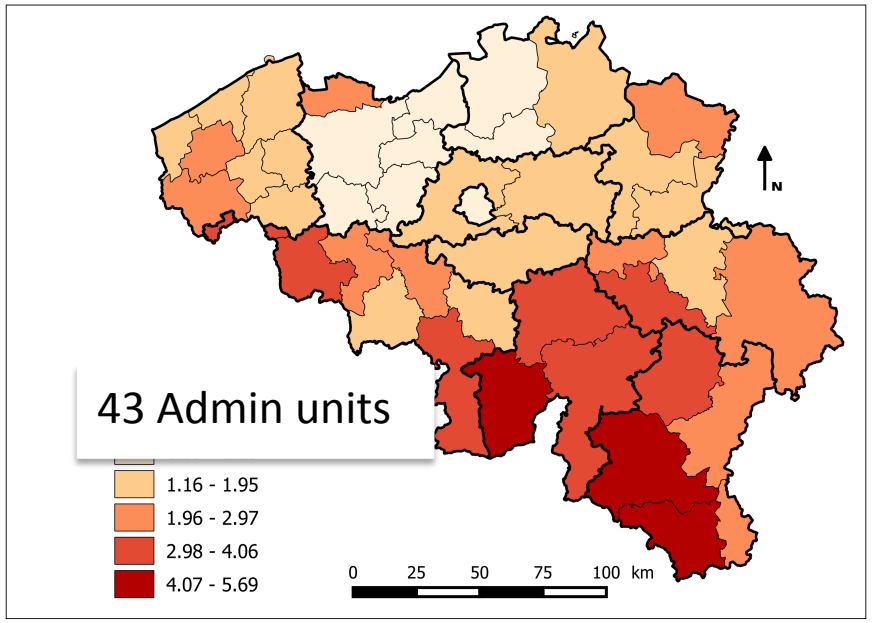
PLACE — **Attributes** — **Physical**, **Human**

(Moving) People to places.
« Complex-city »

How to measure?

Fatalities/100 accidents





SCALE:
2 aspects

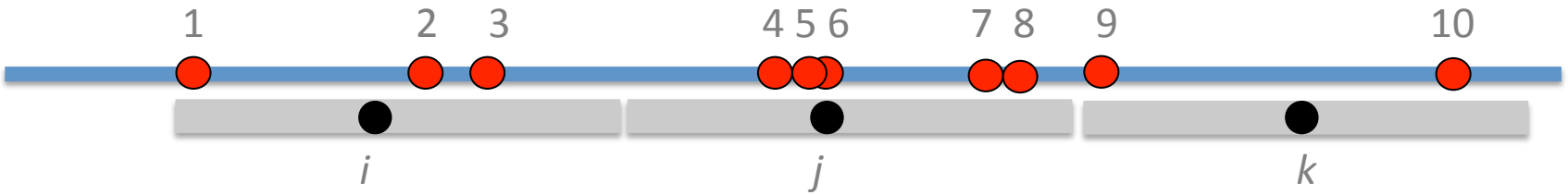
Grain (BSU): level of *spatial resolution* at which an object (or process) is measured

Size x *shape*

Extent: study area

Carto : Adam and Thomas

Agregation distance & scale



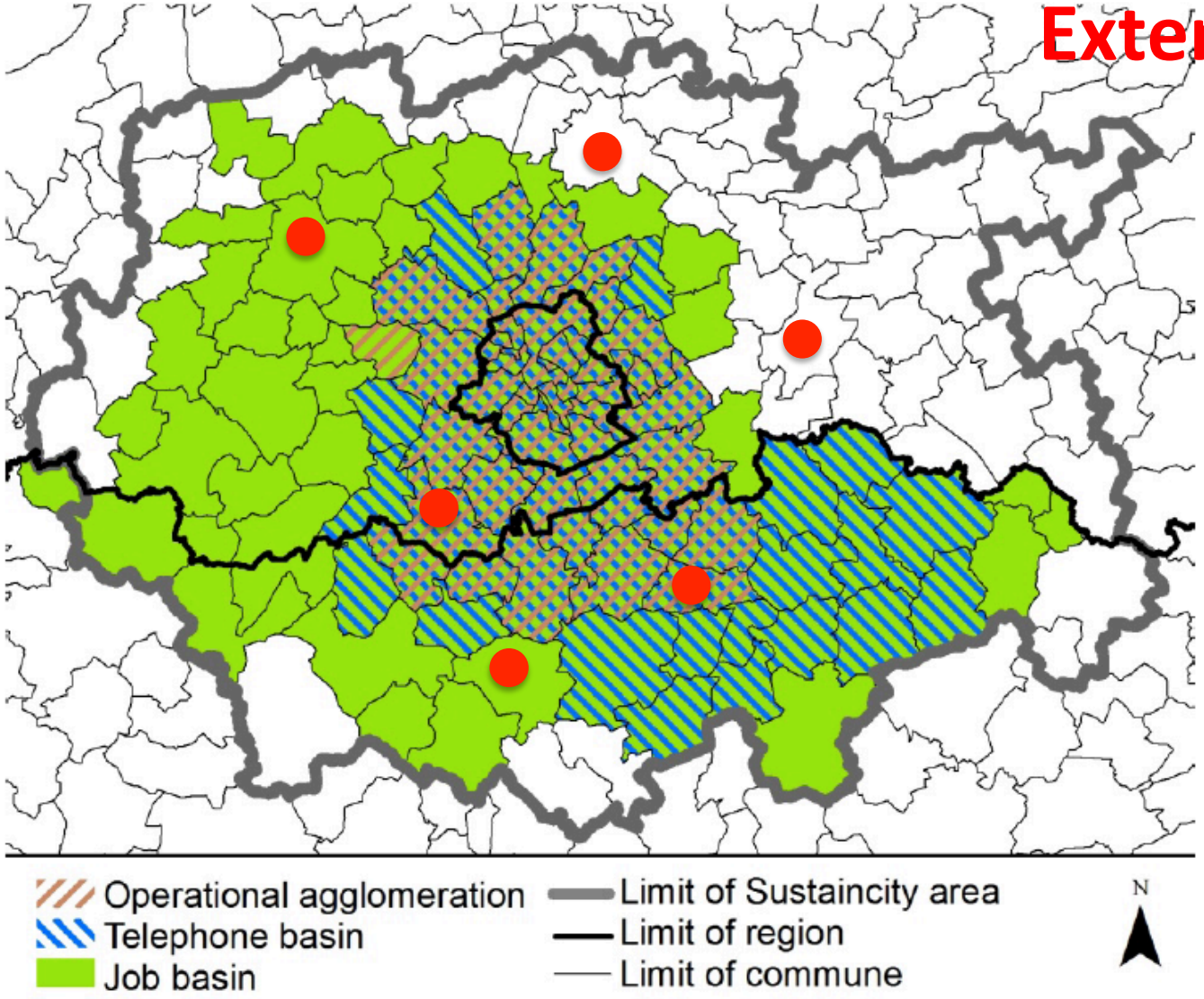
d(3, *j*) < d(*i*,*j*) < d(1, *j*)
d(1,*i*) = 0
9 is allocated to *k* while closer to 7 and 8

SCALE: Extent

Urban-periurban Delineation

Distances
Distance decay
Friction of distance

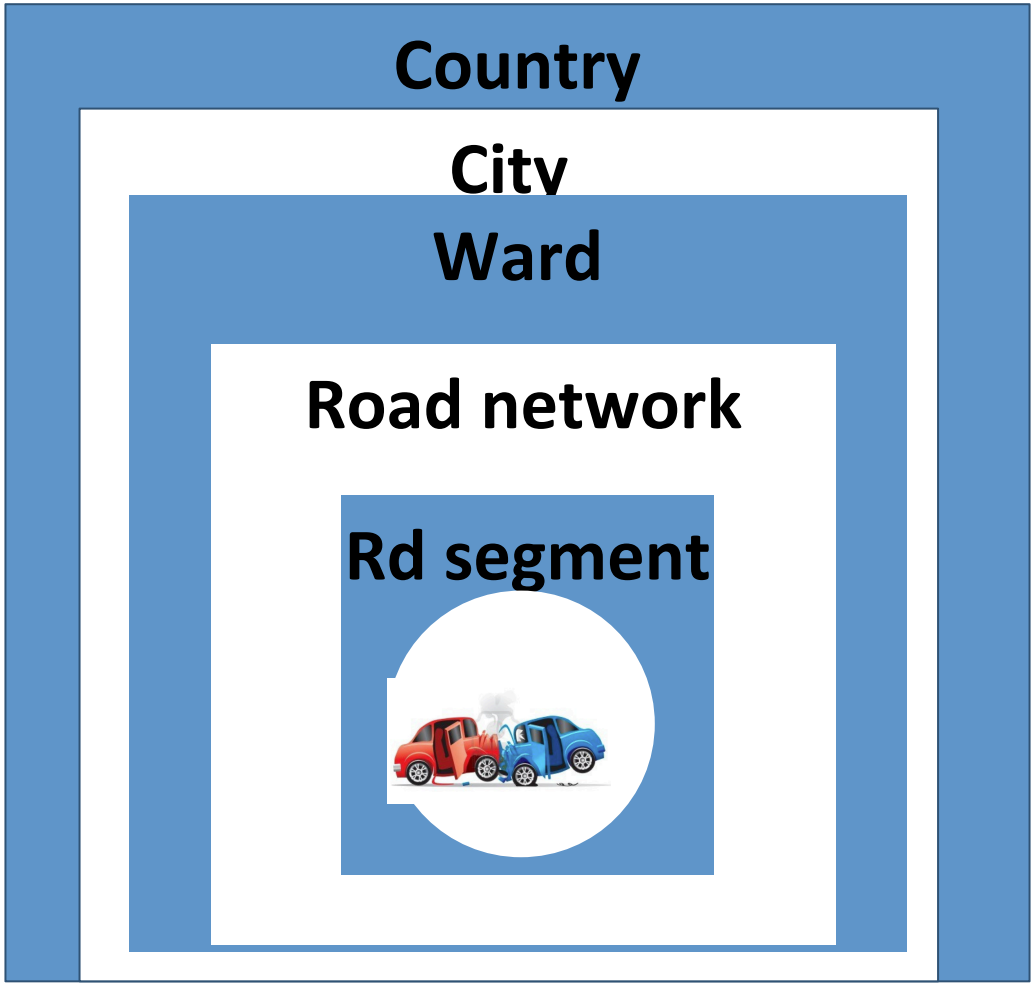
Mode choices
Accessibility
Mobility
(…)



Auteurs : Thomas et al; e-Belgeo , 2012

Scales are nested

A scale cannot be isolated



Do not generalise conclusions at other scales

Ecological & **atomistic fallacy**

Why being concerned about scale?

1. Patterns are dependent upon the scale of observation
2. Importance of explanatory variables changes with scale.
3. Statistical relationships may change with scale.
4. Patterns are generated by processes acting over various spatial (and temporal) scales

Fallacies of scale - No unique solution

Nested models, power laws, fractals, networks, …

What is special about spatial data?

LOCATION

Pitfalls

- **Scale** (nested)
- Unit definition (**MAUP**)
- **Spatial autocorrelation**
- Border (edge) issues
- **Heterogeneity of space**

…

Potentials

- **Distance**
- **Adjacency**
- Interactions
- **Neighborhood**
- **Complexity**
- **…**

Gut-feeling **« Correlation »** **Complex causation**

INTUITION | DATA DRIVEN patterns | MODEL & THEORY | VALIDATION | **RESULTS DECISION**

Increasing

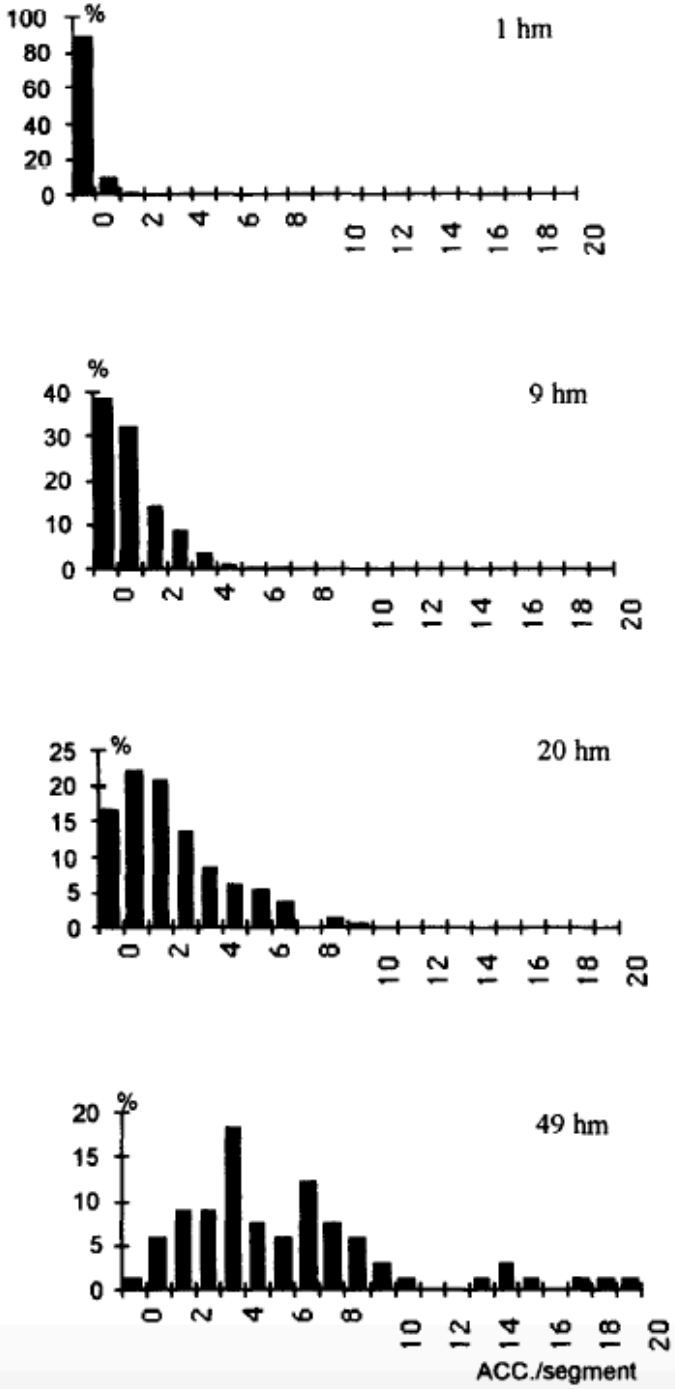
- **Effort and rigor**
- **Level of certainty**

Inspired from : IBM website

Poisson or not ?

- **Process** = Poisson
- **Measures**
  - 1hm (not a point !)
  - Poisson > Binomial
  - **Aggregation** effects

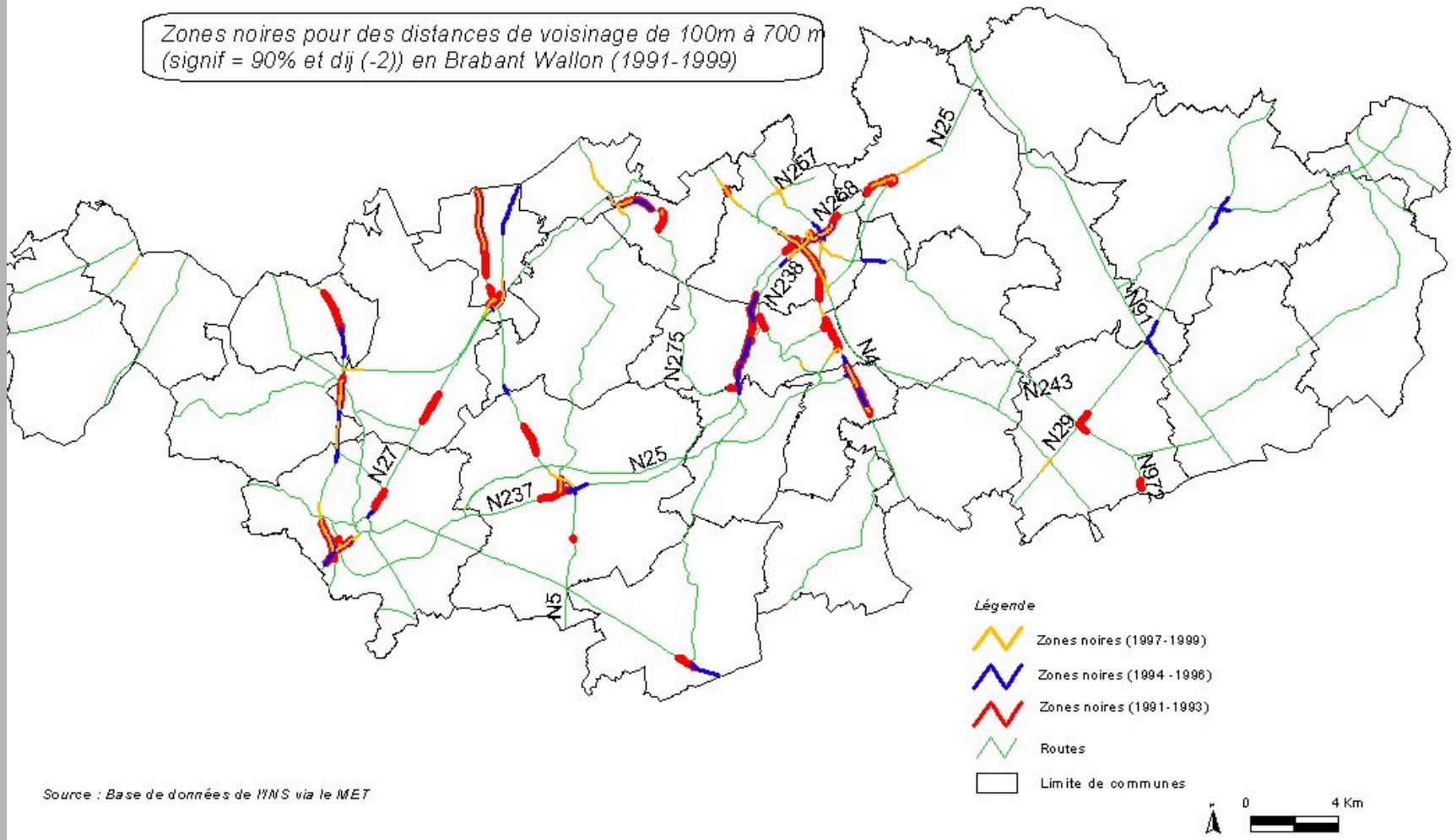
    ! Length of segments



Road accidents (*N29*)

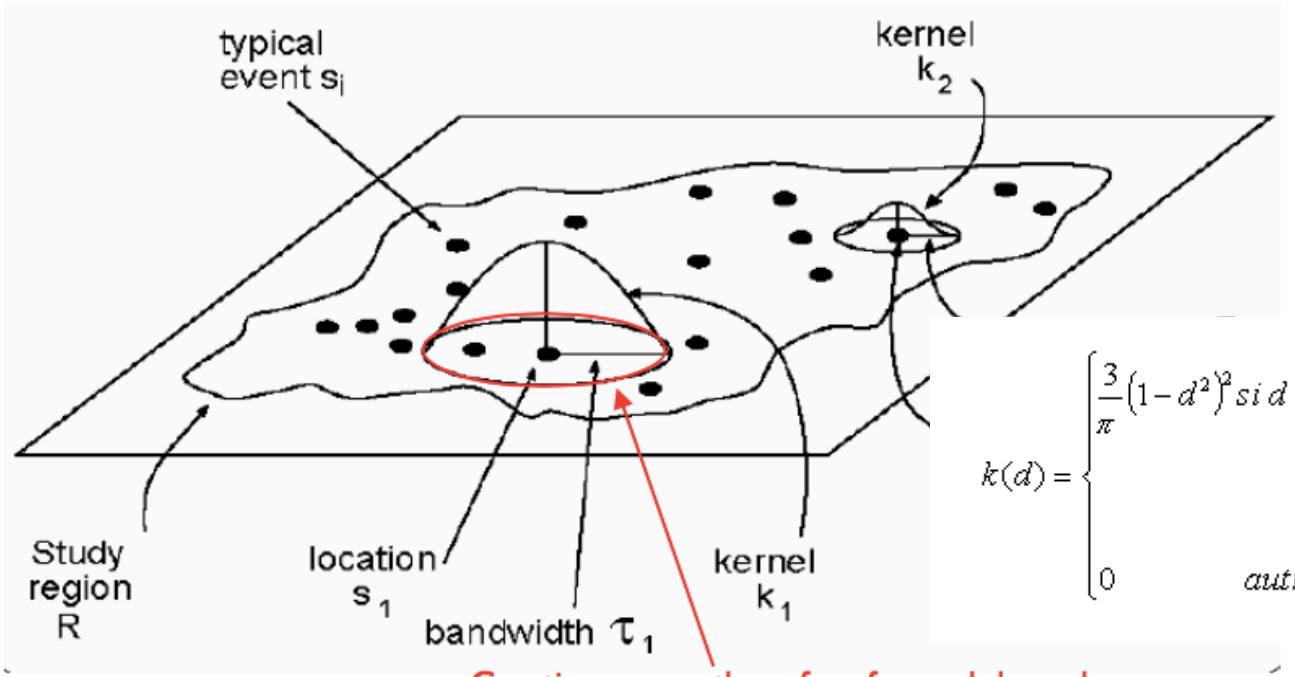
Moran for black segments

3.1 Point pattern analyses



Source: Eckhart,et al. 2004

Kernel estimation



$$k(d) = \begin{cases} \frac{3}{\pi}\left(1-d^2\right)^2 si\ d \le 1 \\ \\ 0 \qquad autrement \end{cases} \qquad avec\ d = \left(\frac{(X - X_i)}{r}\right)$$

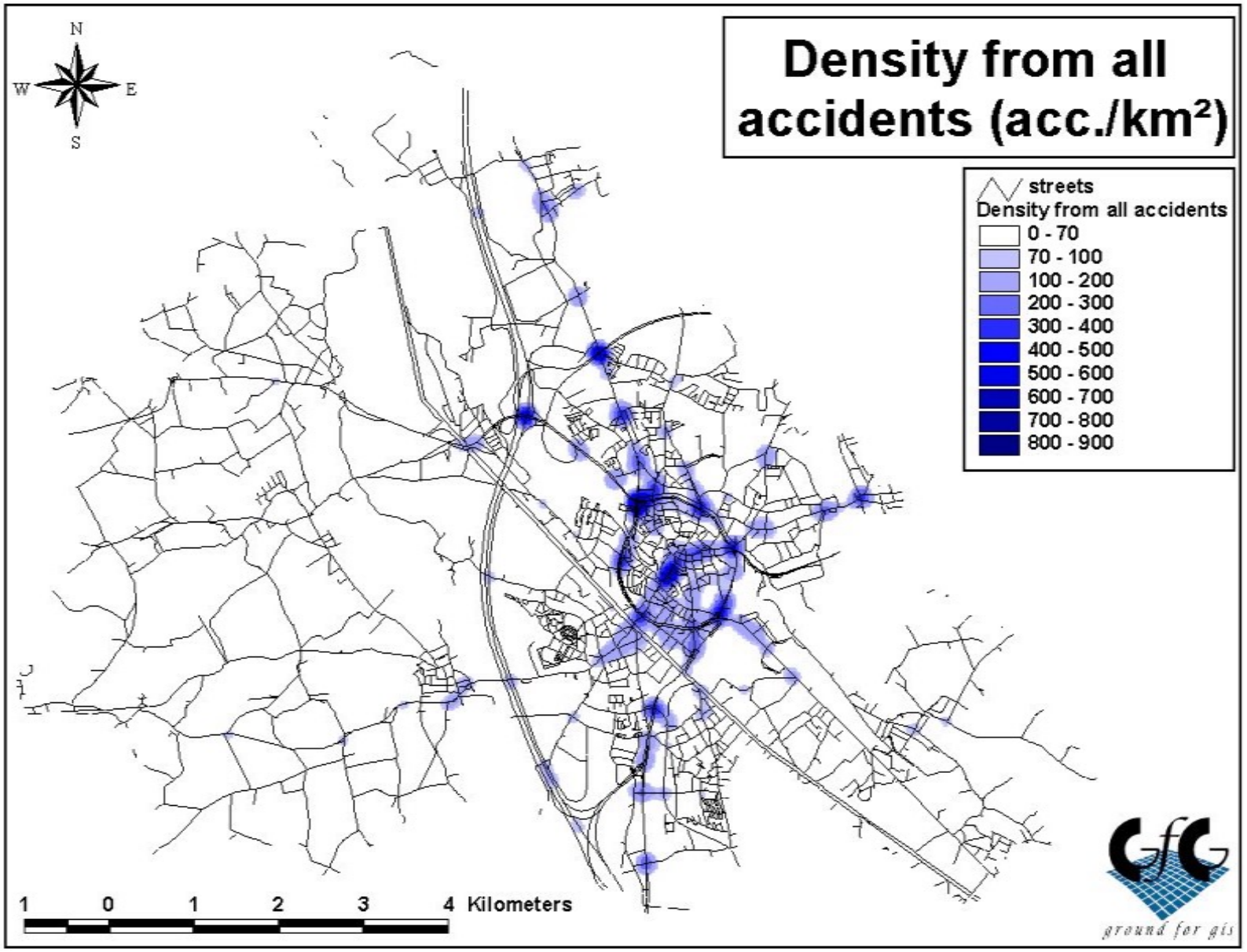
- Creating a smooth surface for each kernel
- Surface value highest in the center (point location) and diminishes with distance…reaches 0 at radius distance

$$\hat{\lambda}(X) = \frac{3}{\pi r^2} \sum_{d \le r} \left(1 - \frac{d_i^2}{r^2}\right)^2$$

Kernels
in 1D or 2D ?

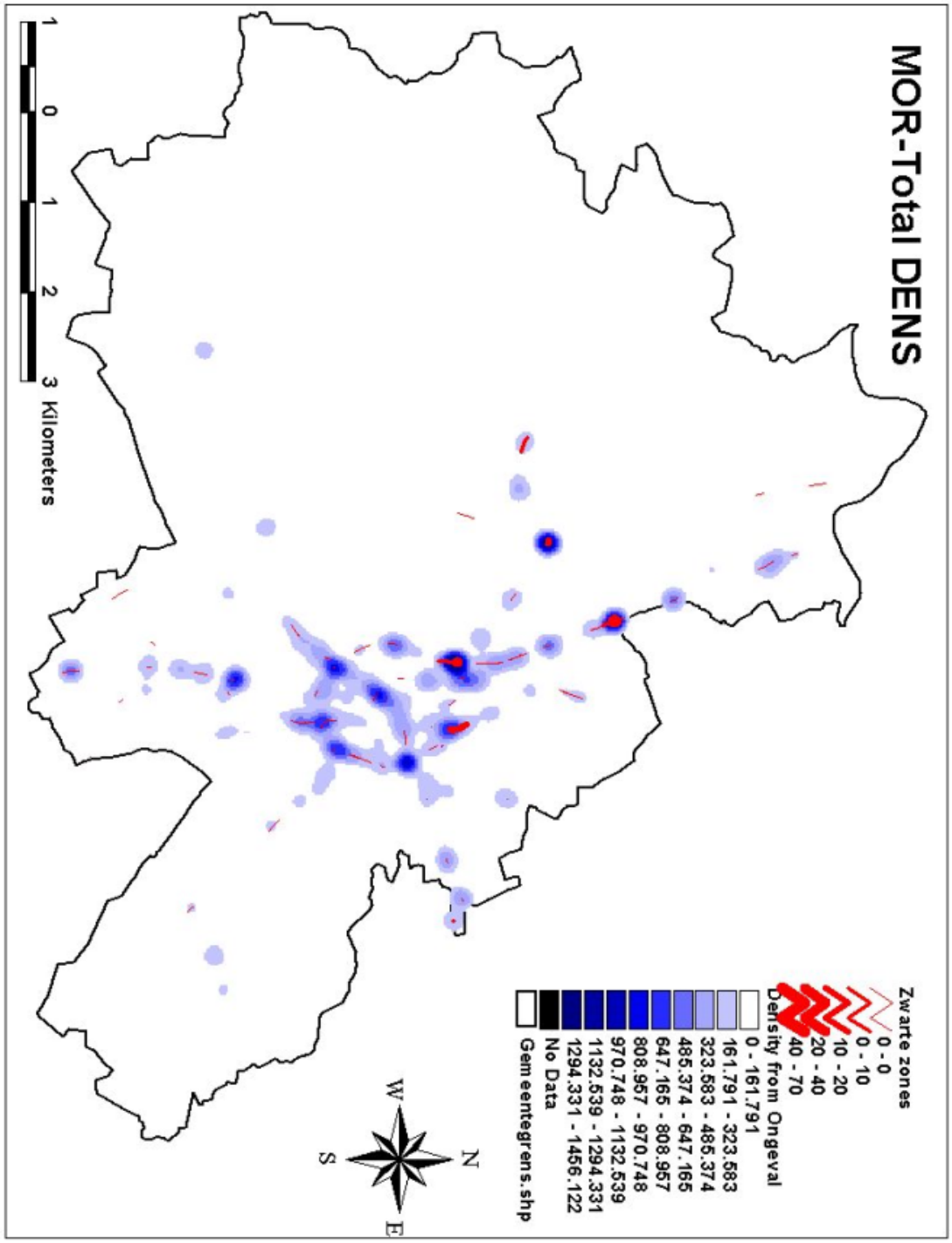
Source: Steenberghen, Defays, Thom

Mechelen



Source: Steenberghen, Defays, Thomas, Flahaut, 2010

3.1 Point pattern analyses



Source: Steenberghen, Defays, Thomas, Flahaut, 2010

3.2 Model for *i* = hecotmers

Y_i = ***1*** if hm belongs to a « black segment ».

Y_i = ***0*** otherwise

X_i Characteristics of the road
- Usage
- Physical properties
- Environment (landuse, …)

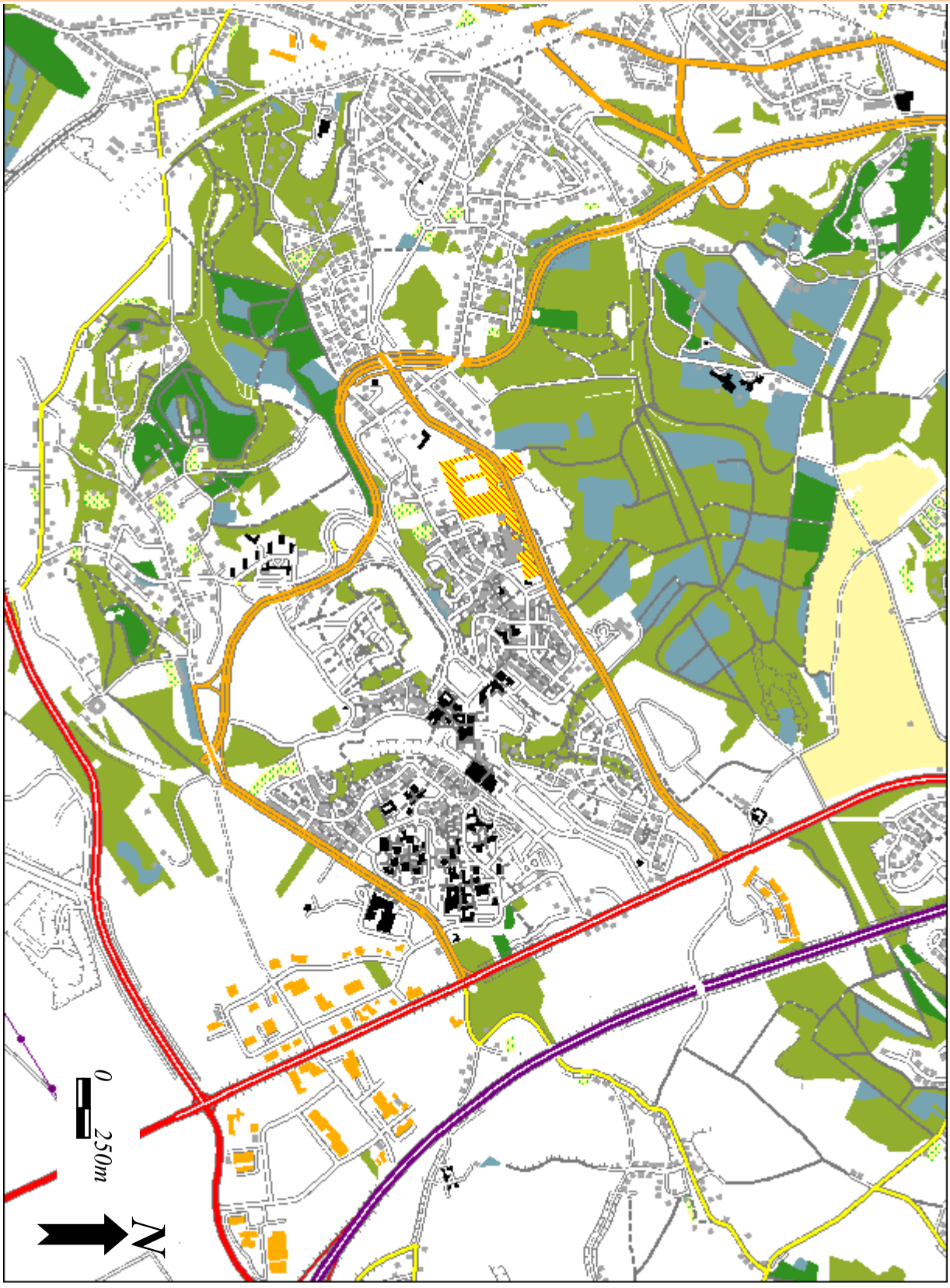
→ *Infrastructure & Environnement*

(Official data; Numerical Digital Terrain Model; IGN maps)

Logistic regression

Source : Flahaut, 2004

3.2 Model for *i* = hecotmers



Source : Flahaut, 2004

3.2 Model for i = hecotmers

Structure of the logistic model for regional roads

Overall significance		
Likelihood ratio	464.7***	
Score	477.7***	
Overall goodness-of-fit		
Deviance/d.f.	1.07[a]	
Generalized R^2 (Nagelkerke)	0.25	
Pseudo-R^2 (McFadden)	0.20	
Significance of each covariate	**Score**	**d.f.**
Traffic	84.2***	3
Proximity of firms	32.1***	1
Distance to a major junction	38.1***	3
Proportion of built area	28.2***	3
Type of road	18.4***	3
Distance to a change in the speed limit	17.4***	3
Adherence	8.9**	1
Type of road surfacing	13.4**	3
Distance to a change in the type of road	8.9*	3

$N = 3479$[3] from which $N_{Y=1} = 376$ (11%).

[a] Not significant at 95%.

* Significant at 95%.

** Significant at 99%.

*** Significant at 99.9%.

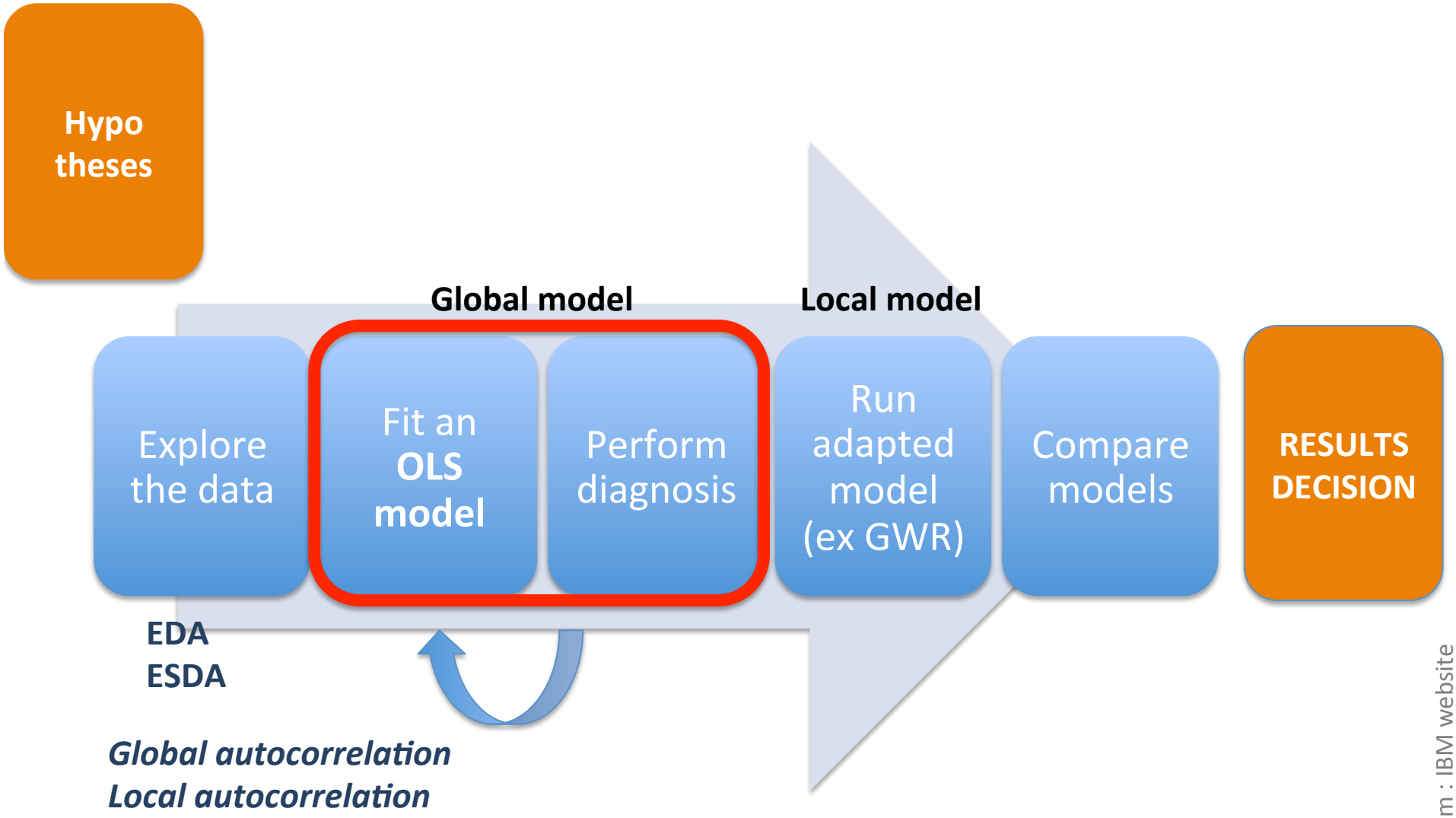
Source : Flahaut, 2004

Multi-level modelling?

Variables	*Multilevel model*	*Logistic model*
TRAFFIC	++	+++
VMAX (0m)– *_dist*		+++
LANES	–––	
LANES (0m) *_dist*		+
SURFACE		++
JUNCT (0m) *_dist*	+++	+++
ADHERENCE	+++	+
BUILT (30%)	+++	+++
FIRMS	++	+++
DIRECTION	––	
EMPLOYDENS (level 2)	+++	

(+ a positive relationship; – a negative relationship.

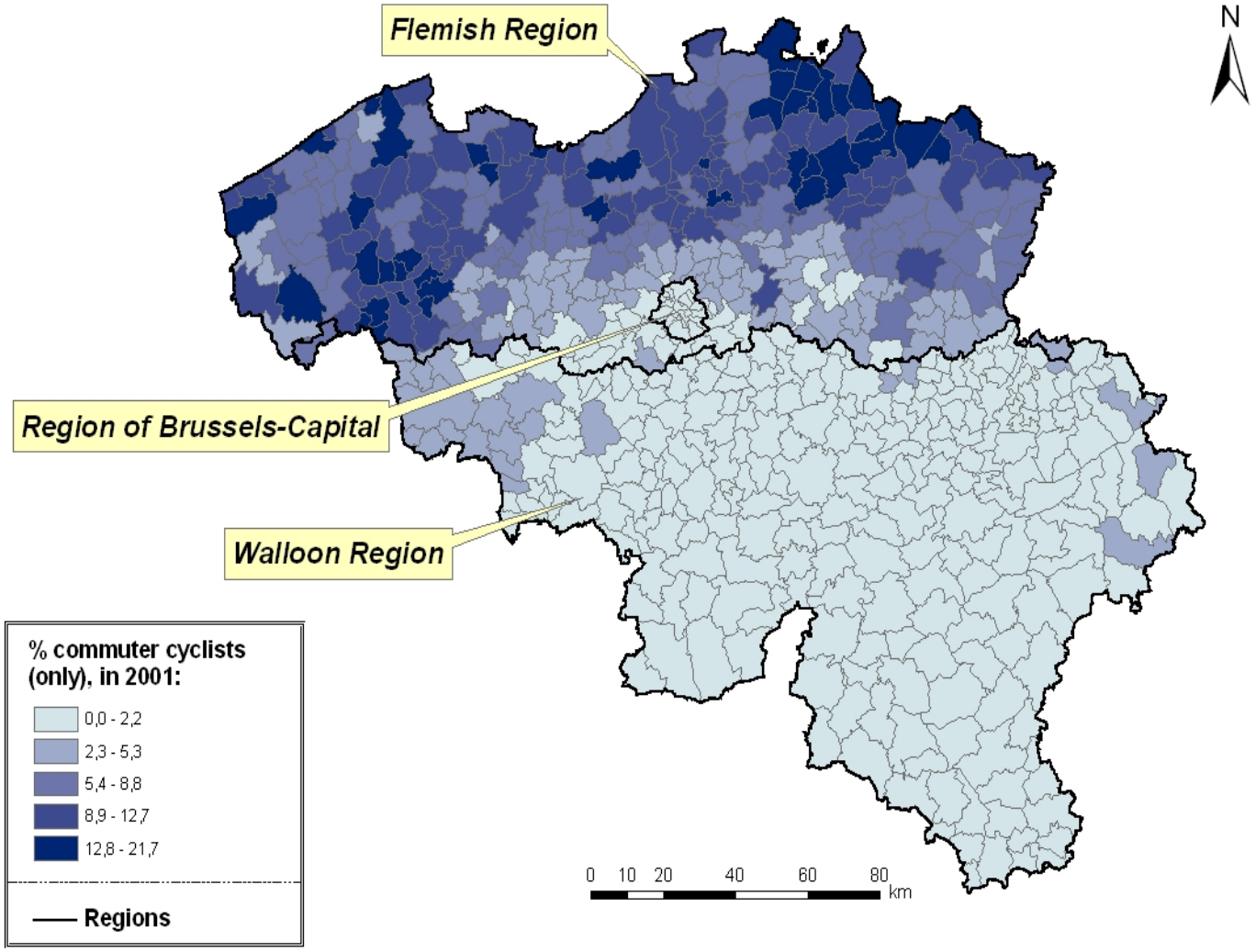
+++/–––significant at 99.9%; ++/–– significant at 99%; +/– significant at 95%).



Inspired from : IBM website

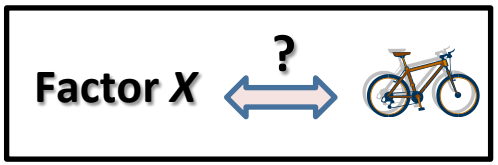
Objective: explain variations in *Y*

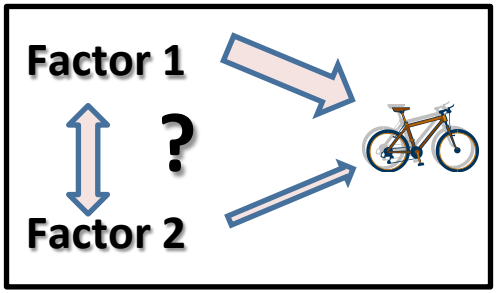
Controlling spatial biases



Source : Vandenbulcke et al, 2011

2 steps





EXPLORATORY

Identify potential **explanatory factors**

Statistical tools:
- Graphics, (basic statistics)
- Cluster analyses, (PCA)
- Correlations (*x*,*y*)

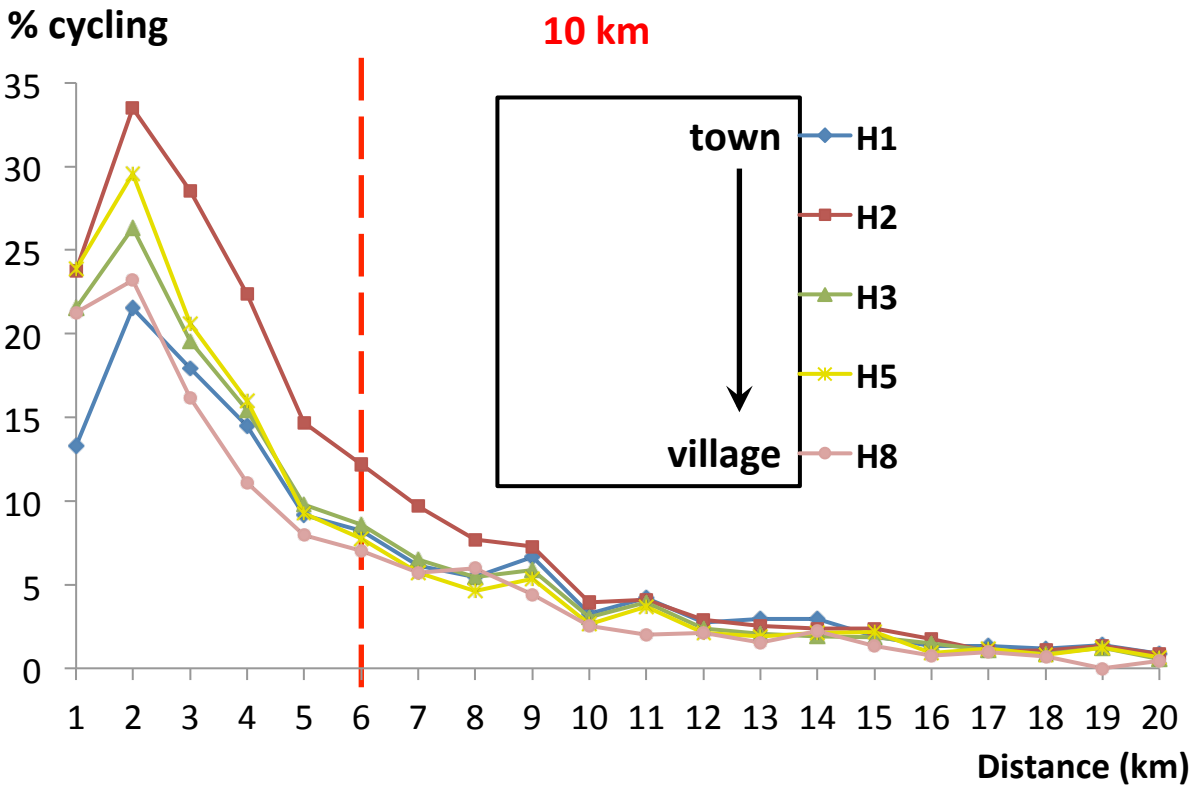
STATISTICAL MODELLING

Relative **importance** of variables

Statistical tools
- Statistical models
- Corrections for multicollinearity & spatial effects

EXPLORATORY

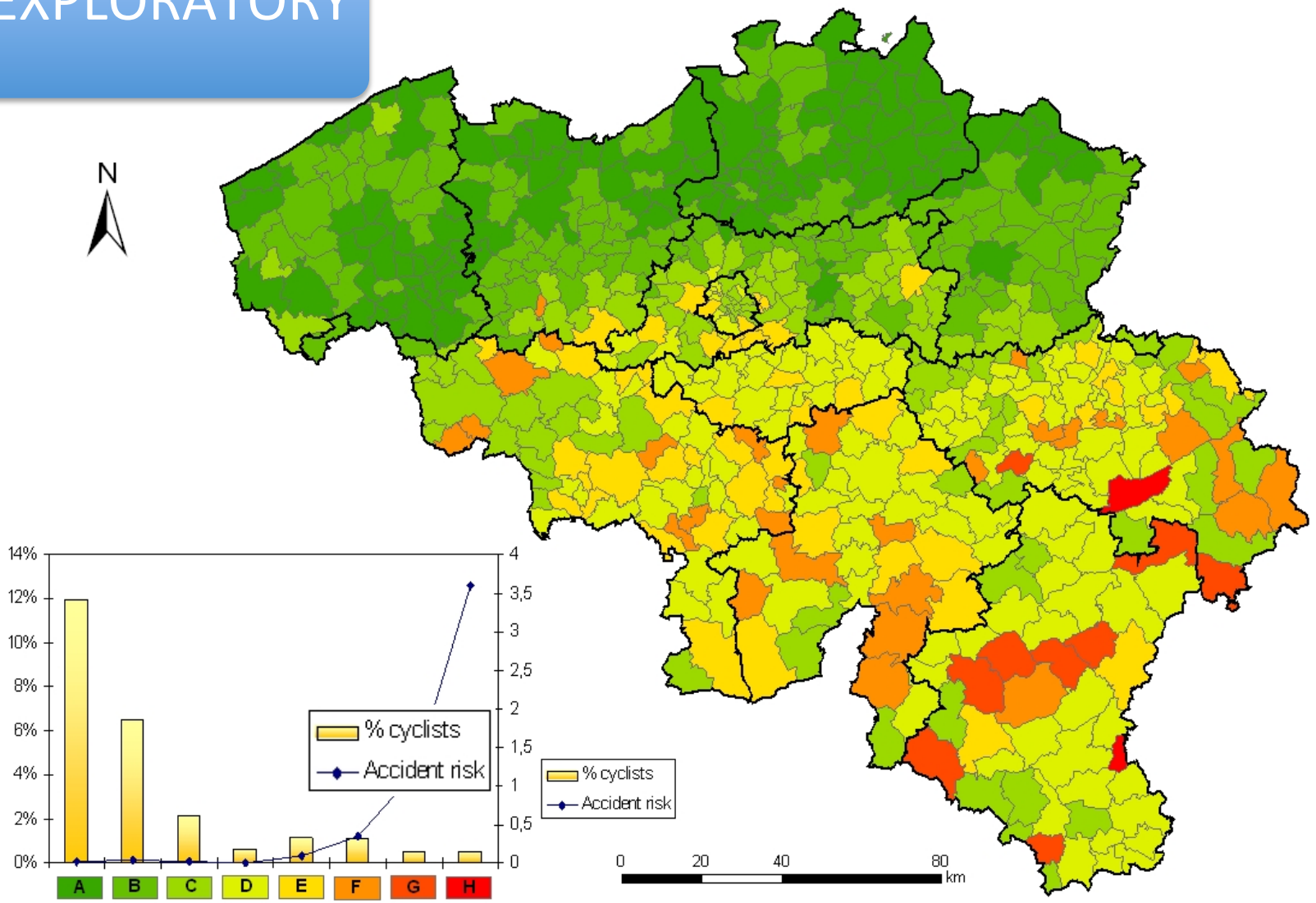
- Commuting **distances** (< 10 km)
- **Town size**: regional towns > large towns
- **Regional** differences (culture + …)



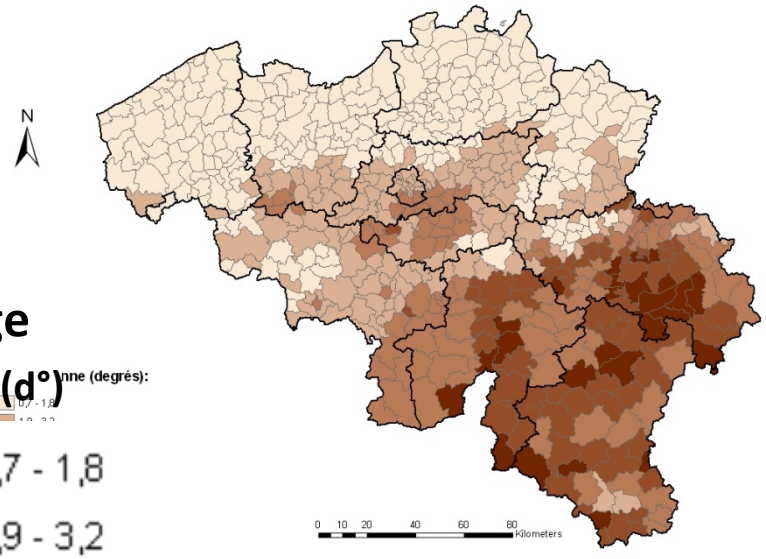
Source : Vandenbulcke et al, 2011

3.3 Model for *i* = communes

EXPLORATORY

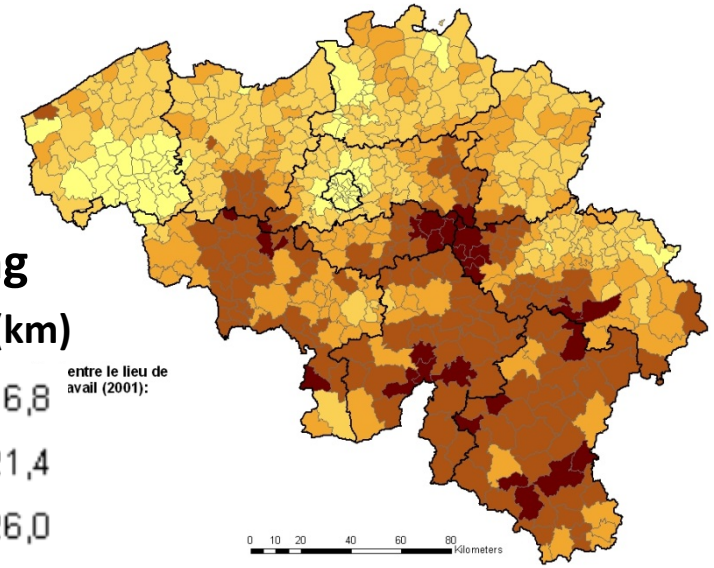


Source : Vandenbulcke et al, 2011



Average slopes (d°)

- 0,7 - 1,8
- 1,9 - 3,2
- 3,3 - 4,9
- 5,0 - 7,1
- 7,2 - 10,8



Commuting distances (km)

- 10,2 - 16,8
- 16,9 - 21,4
- 21,5 - 26,0
- 26,1 - 31,9
- 32,0 - 42,7

ρ_{xy} = 1

(correlation)

Active people < 25 years: 0.54

Job density: 0.38

No child, town size: 0.23

ρ_{xy} = 0

Accident risk: – 0.32

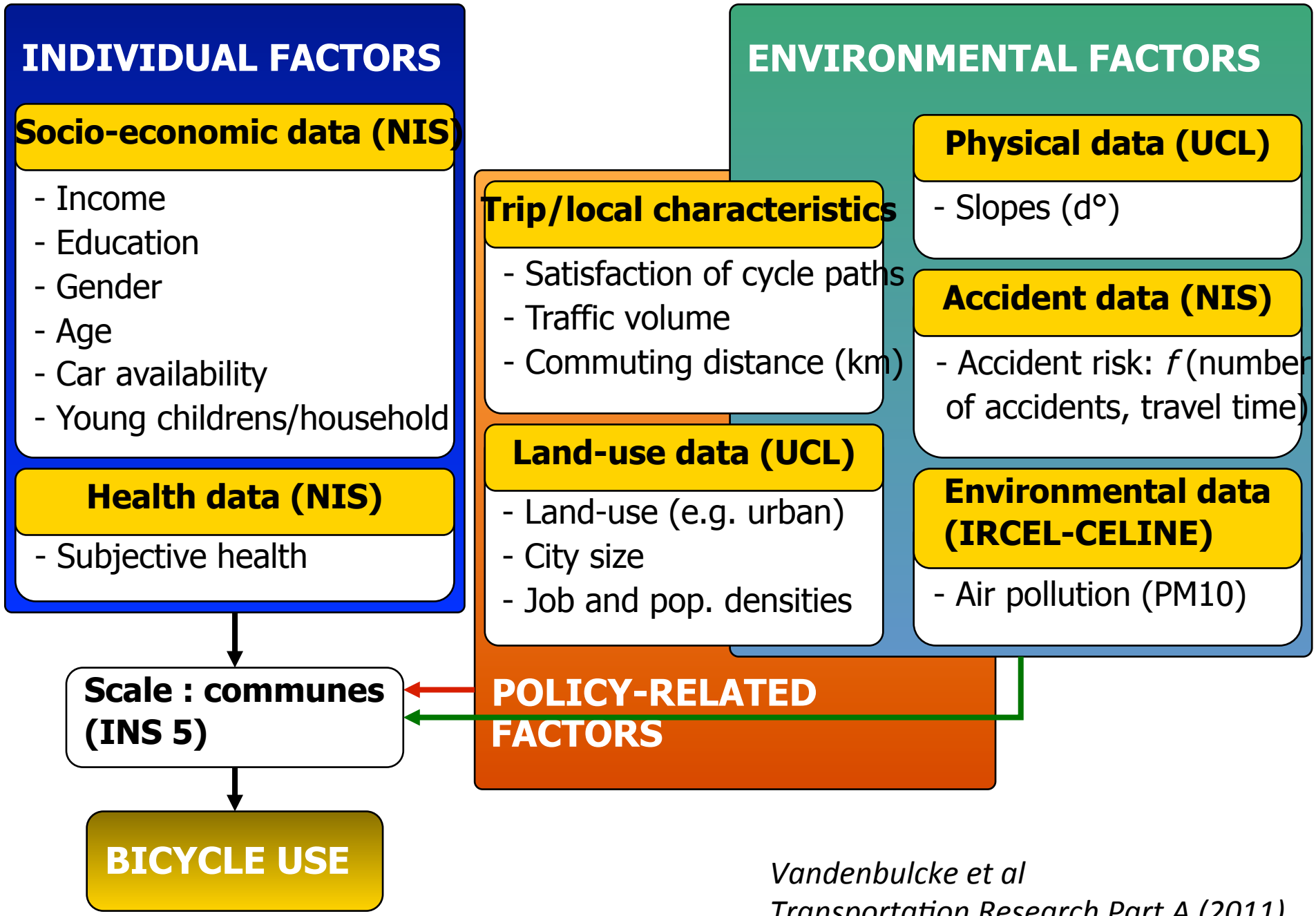
Commuting distances: – 0.54

Poor health: – 0.58

Slopes: –0.77

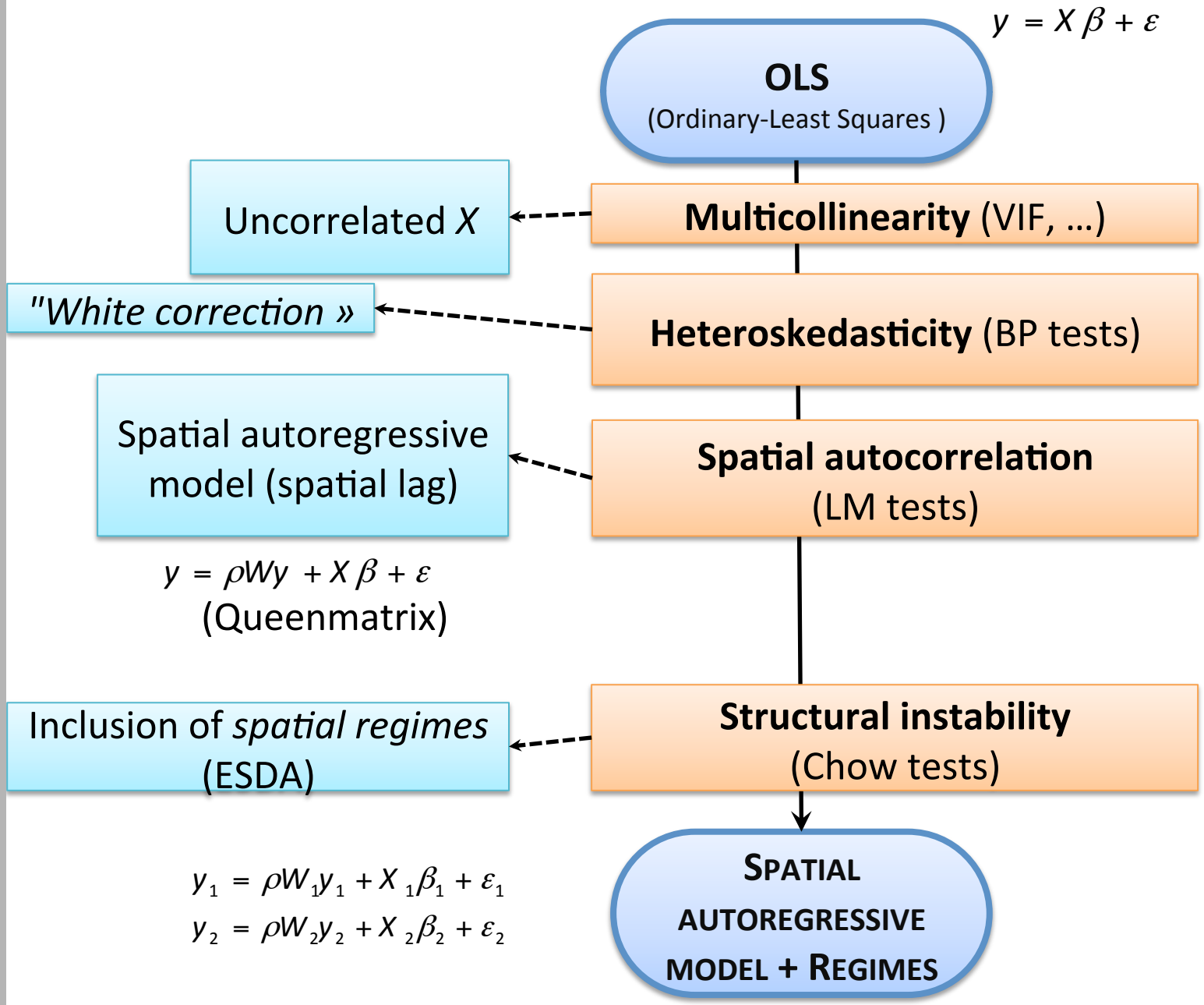
Unsatisfaction of cycleways: –0.82

ρ_{xy} = –1

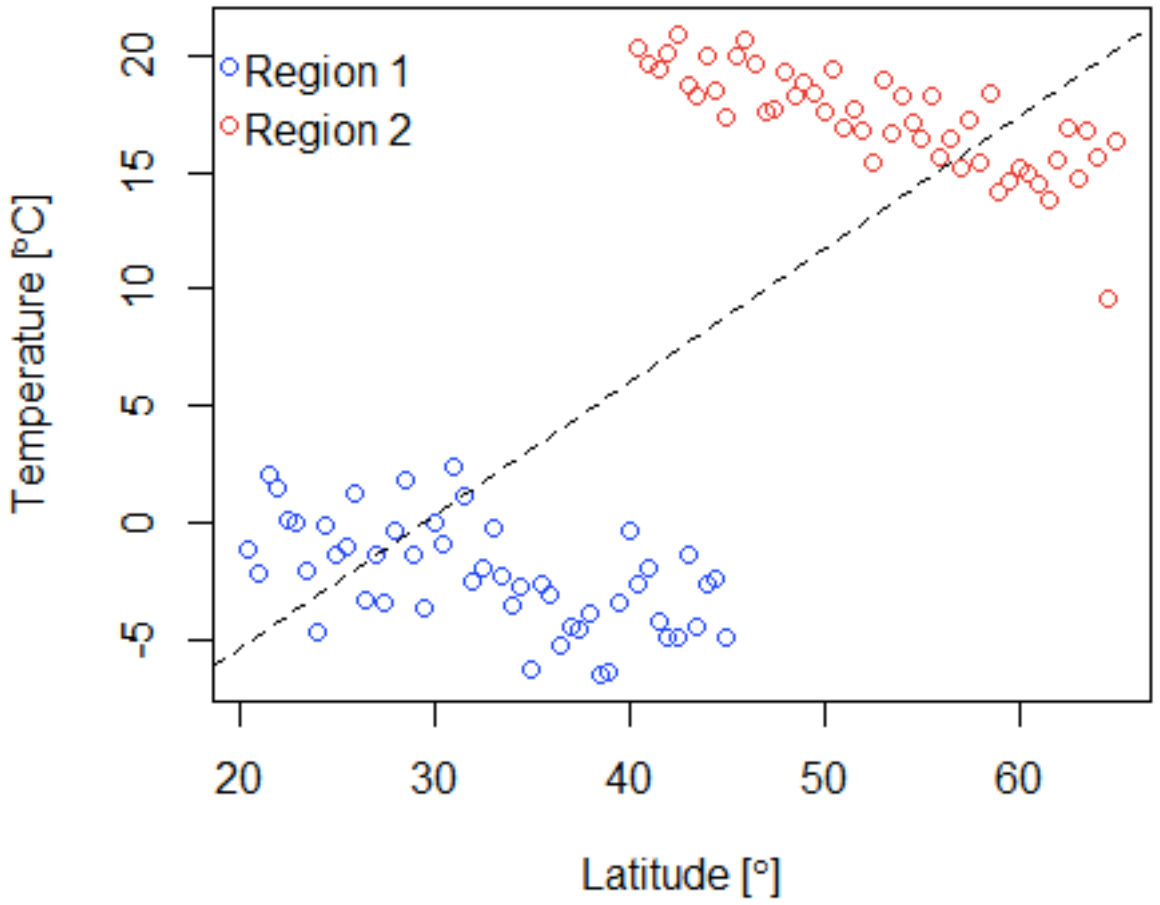


Vandenbulcke et al
Transportation Research Part A (2011)

3.3 Model for *i* = communes



Simpson's paradox



Spatial LAG model + Regimes N-S

Y *= % commuter cyclists in commune i*

North = Flanders
South = Wallonia & Brussels

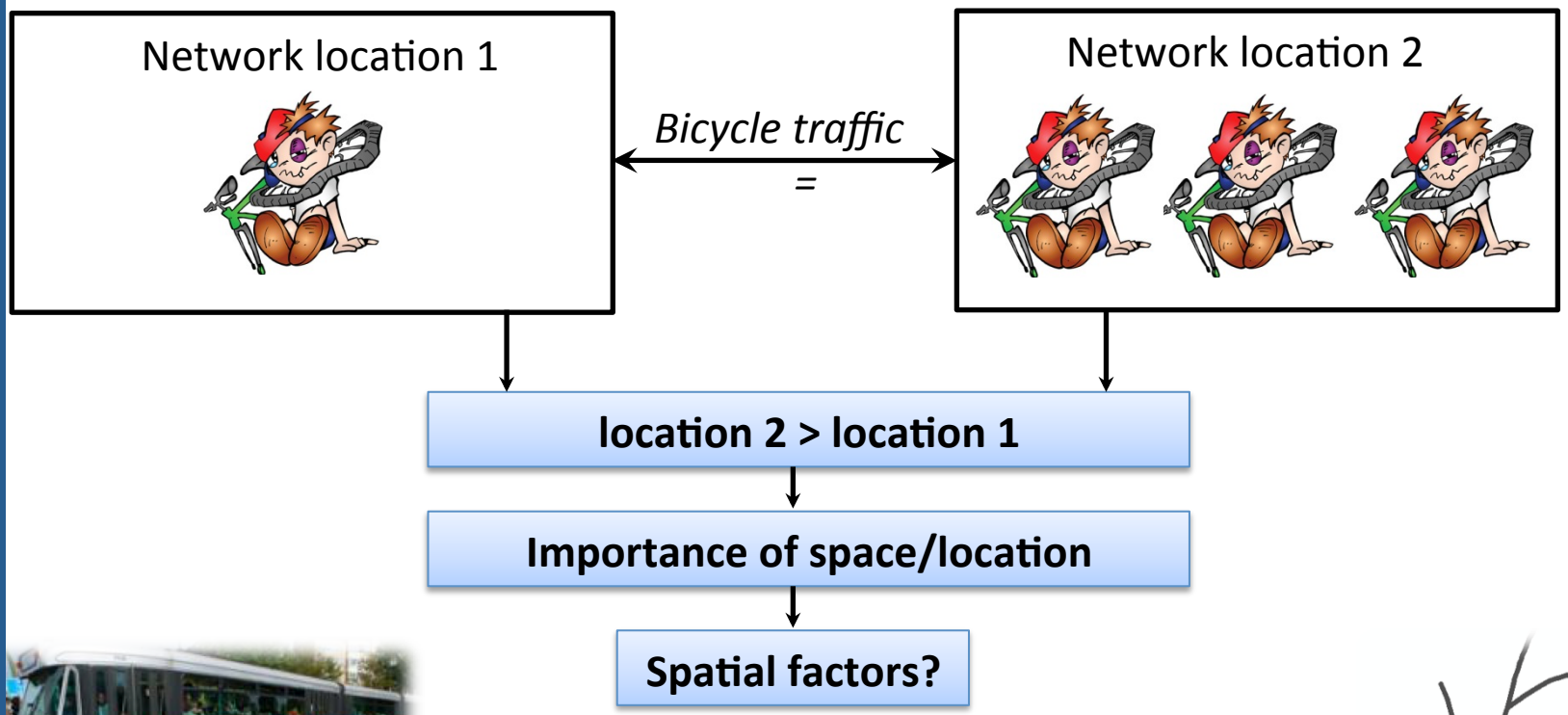
Demographic factors
Socio-economic
Environmental factors

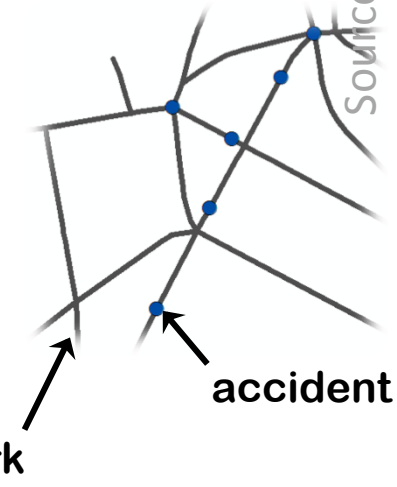
- Dissatisfaction with cycle facilities
+ Town size
- Accident risk
- Traffic volume

	North	***South***
Intercept	2,3084*	4,30951****
Median income	0,0311*	-0,0027
Active men	0,0296**	0,0008
Age 2 (45-54 years)	-0,0417**	-0,0205***
Young children	-0,0365***	-0,0247***
Cycleways unsatisfaction	-0,0052***	-0,0045***
Commuting distance	-0,0165***	-0,0047*
Air quality	0,01384****	-0,0054
City size	-0,11459****	-0,03615****
Bad health	-0,0098	-0,0146**
Accident risk	-0,76319****	-0,14892****
Traffic volume 2 (municipal network)	-0,2357	-0,4521**
Age 3 (> 54 years)	-0,1074	-0,0680
Education 3 (university degree)	-0,0968	-0,3132***
Slopes	-0,1931**	-0,1972****
Lag coefficient (ρ)	0,5362****	
N	589 (N_{North} = 308; N_{South} = 281)	
Log Likelihood	93,923	

Source : Vandenbulcke et al, 2011

3.4 Model for *i* = addresses





Source : Vandenbulcke, IntPanis, Thomas 2014

3.4 Model for *i* = addresses

- $Y_i = 0,1$ ⇨ logistic specification
- **Corrections for**
  - Multicollinearity
  - Heteroskedasticity
  - Residual spatial autocorrelation

    ⇨ omitted variables? ⇨ *spatial models*
- **Bayesian framework**

Source : Vandenbulcke, IntPanis, Thomas 2014

Models based on accident-only data

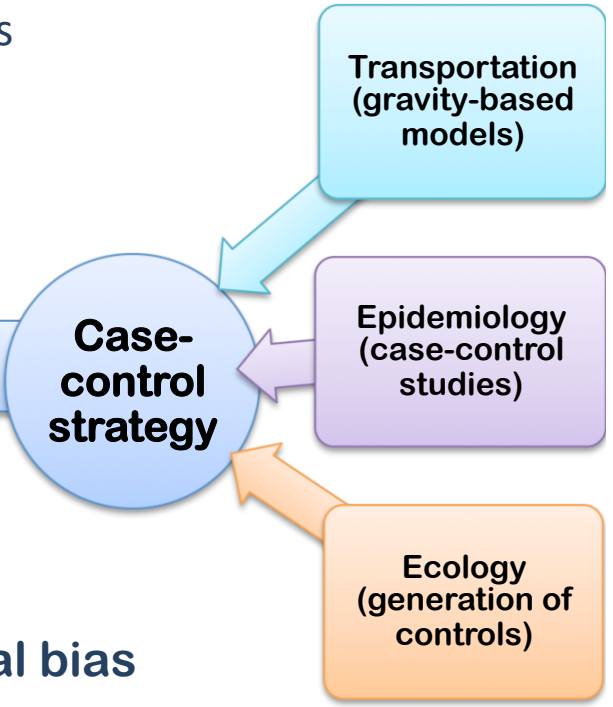
- Regression methods (e.g. multinomial logit models)
- Issues: over-/under-dispersion, underreporting, etc.

Models based on surveys, road trajectories

- Regression methods (e.g. logistic models)
- Main issue: bias in the selection of road trajectories

Models based on case-controls?

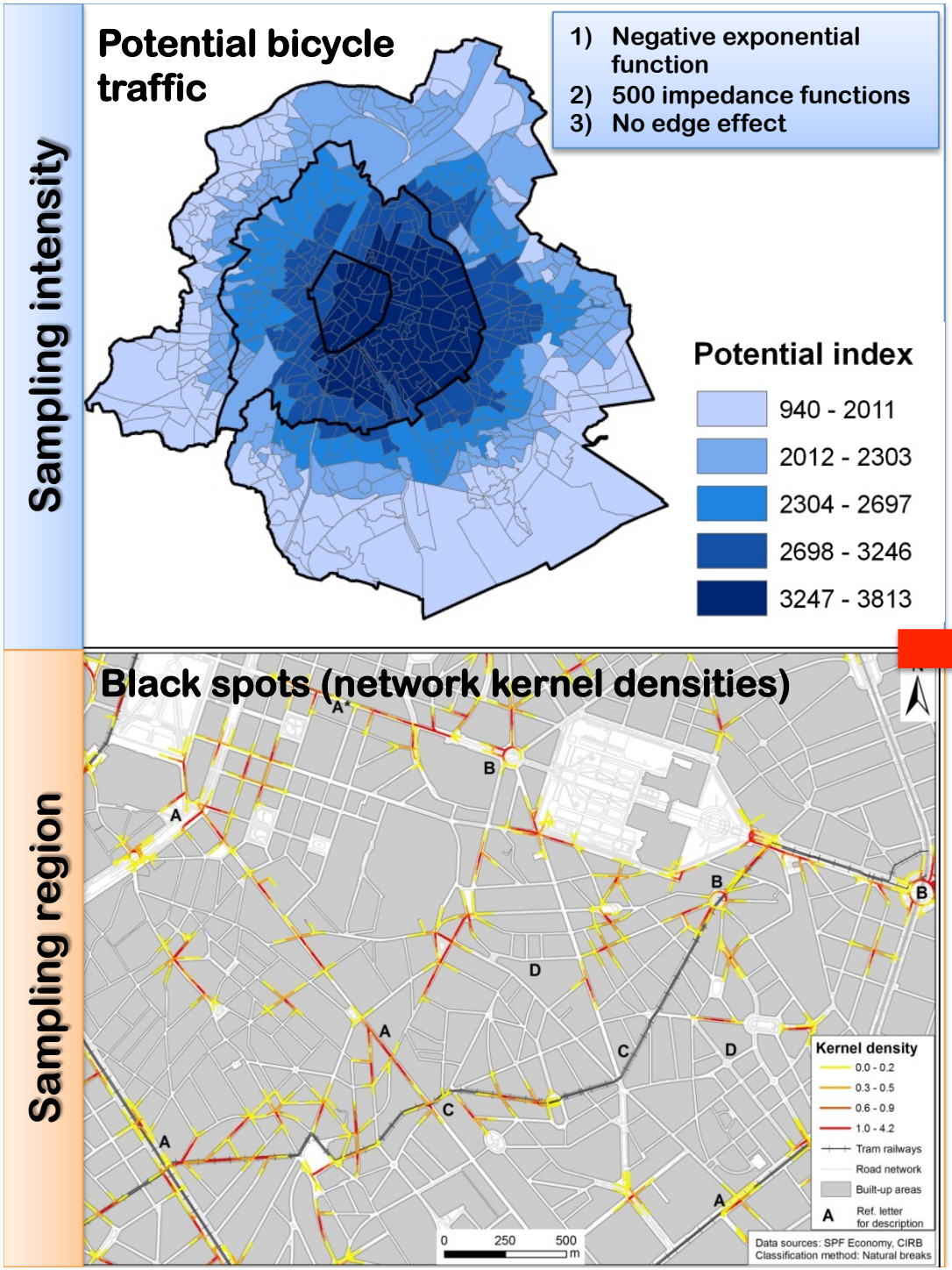
- Cases = accidents
  + Controls = generated absences ⇨ y_i= (0,1)
- Regression methods (e.g. logistic models)
- **Advantage**: estimation of risk, reduced statistical bias
- **Issues**: no vehicle & human factors, selection of controls



Data collection

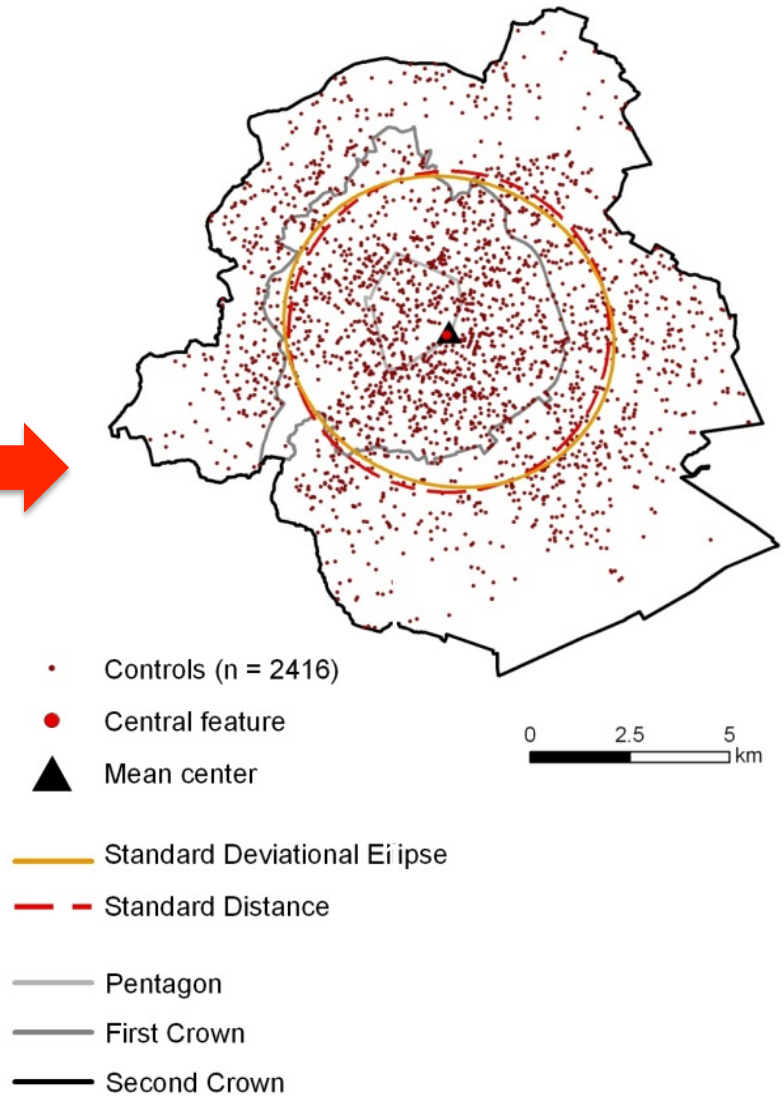
- **Accident risk = time-consuming process**
  - **Accidents (cases)** ⇨ to be geocoded/located
  - **'Absences' (controls)** ⇨ to be generated
  - **Road network** ⇨ exclude 'unbikeable' links
  - **Risk factors** ⇨ to be collected…

Source : Vandenbulcke, IntPanis, Thomas 2014



Stratified random sampling

$N_{controls} = 4 * N_{accidents}$



Collecting risk factors

Infrastructure factors

- Cycling facilities & contraflow cycling
- Discontinuities
- Parking areas & garages
- Bridge & funnels
- Crossroads & complexity
- Tram railways
- Traffic-calming areas
- Major roads
- Proximity city centre
- Distance to specific points of interest (e.g. schools, bus stops, etc.)

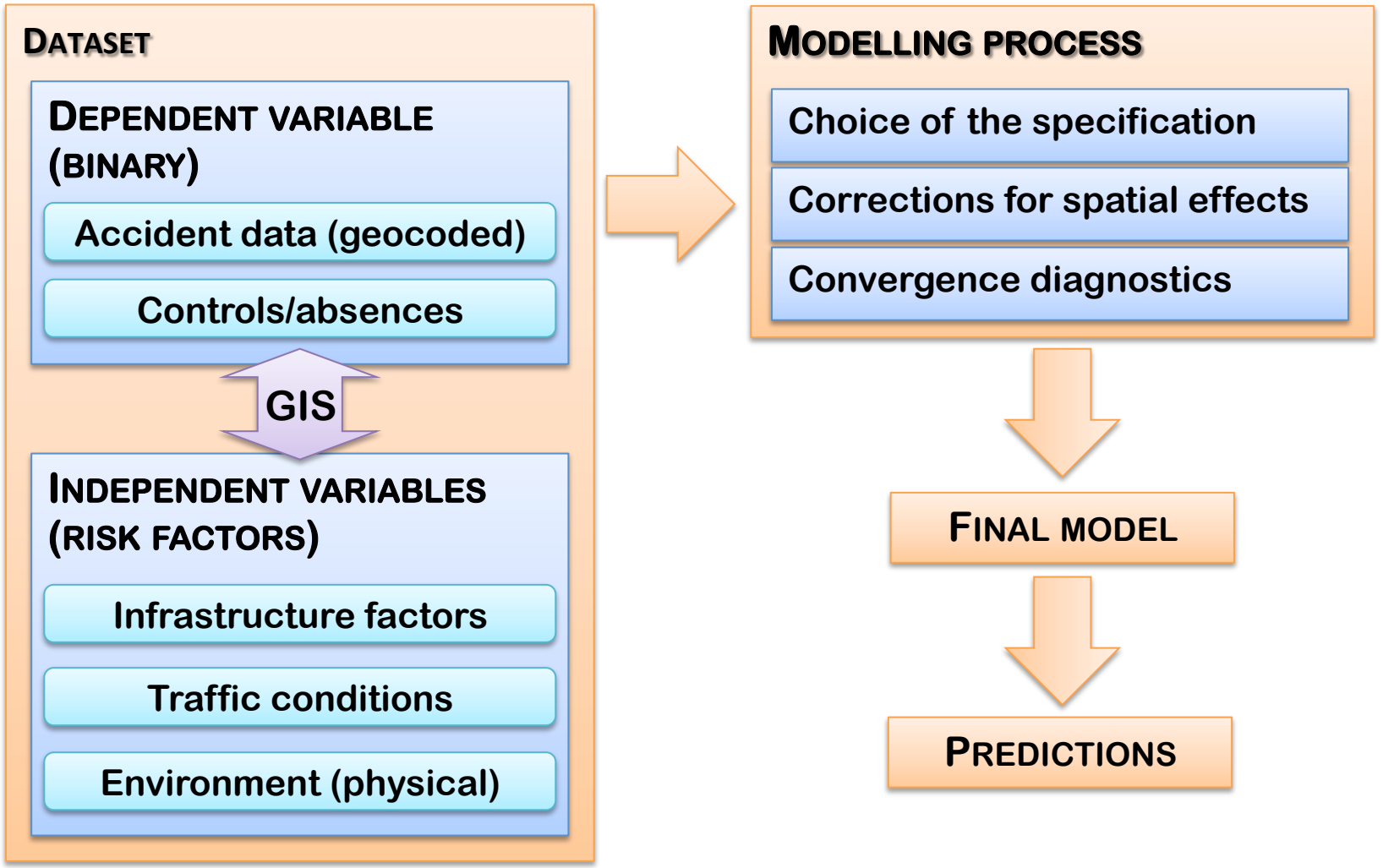
Traffic conditions

- Cars
- Trucks/lorries & buses
- Vans

Environmental factors

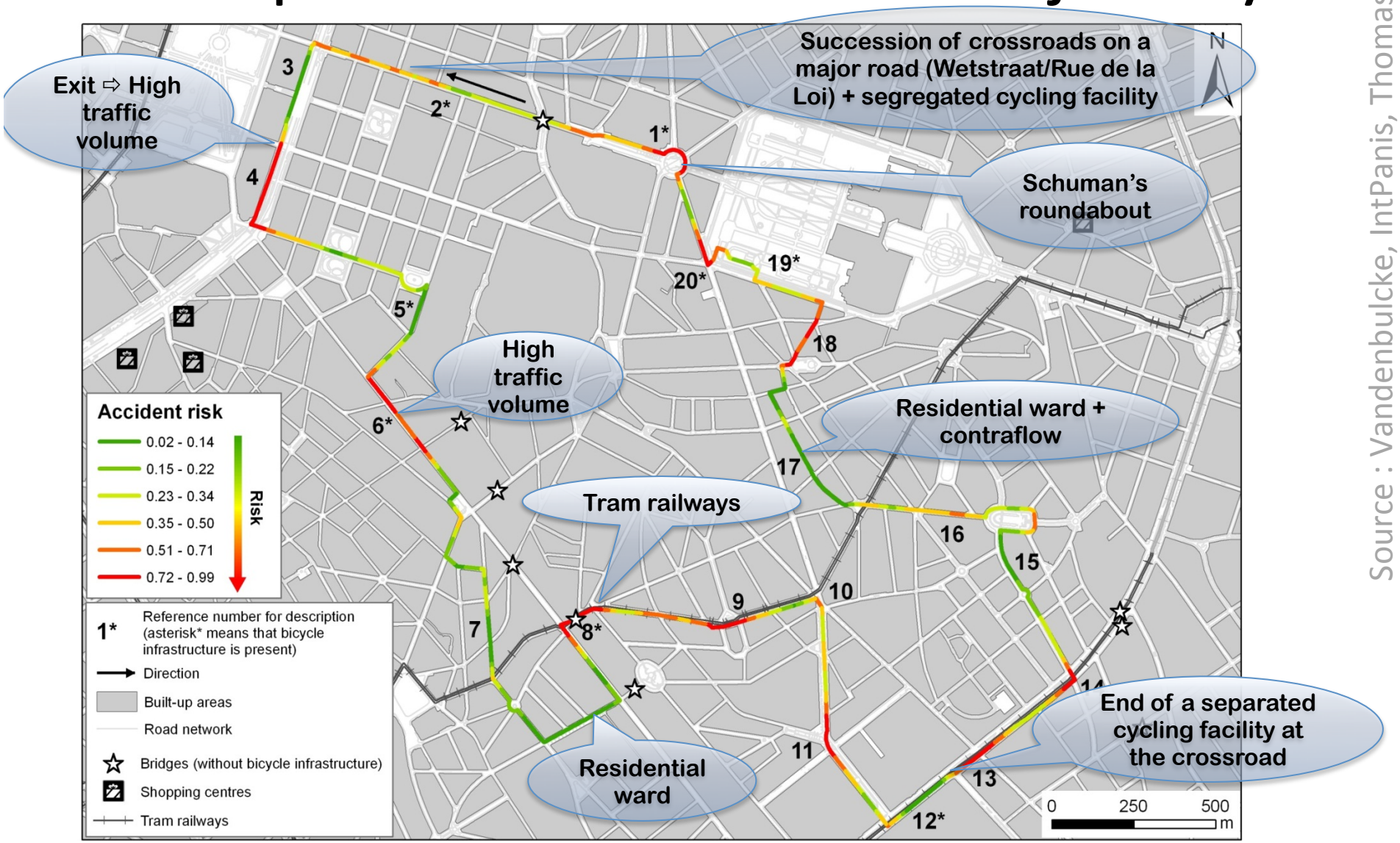
- Gradients
- Green blocks (parks, etc.)

Source : Vandenbulcke, IntPanis, Thomas 2014

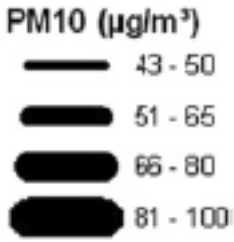


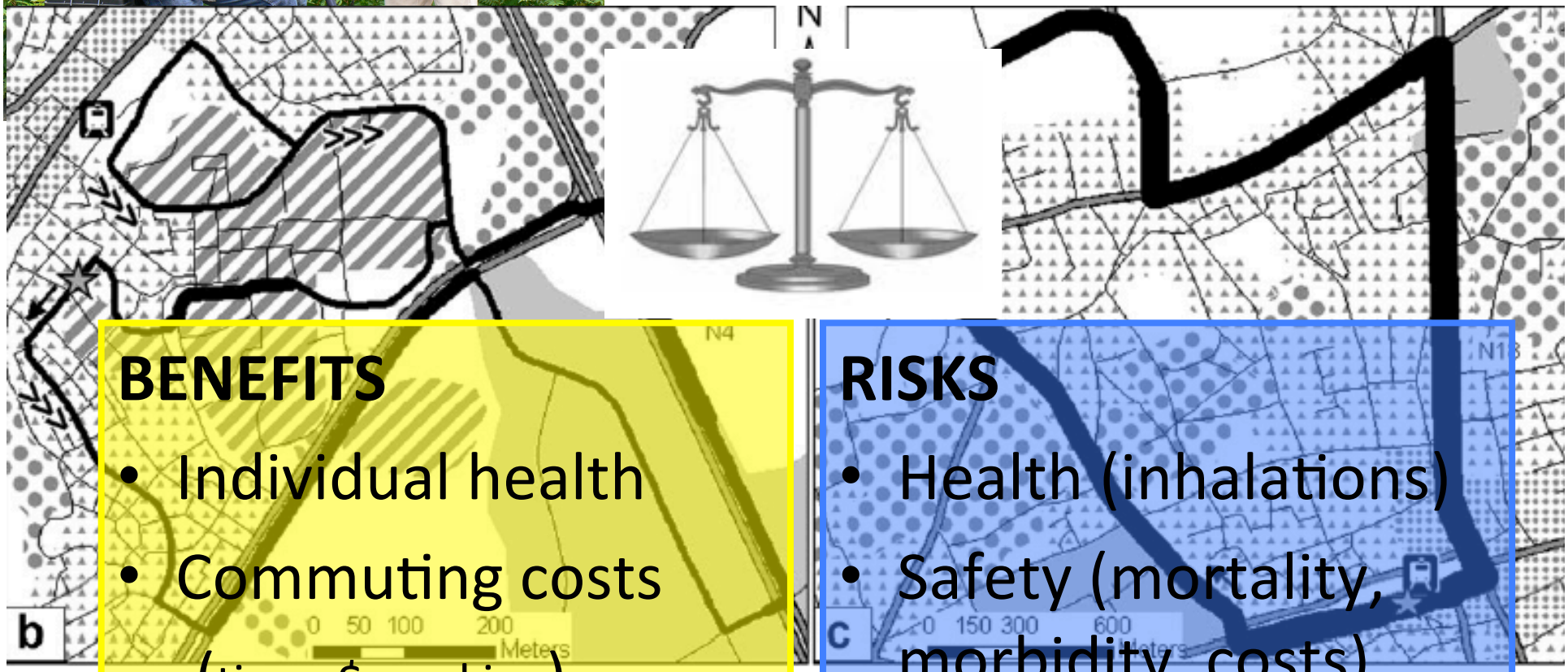
Source : Vandenbulcke, IntPanis, Thomas 2014

Output: Predictions for a trajectory



Source : Vandenbulcke, IntPanis, Thomas 2014





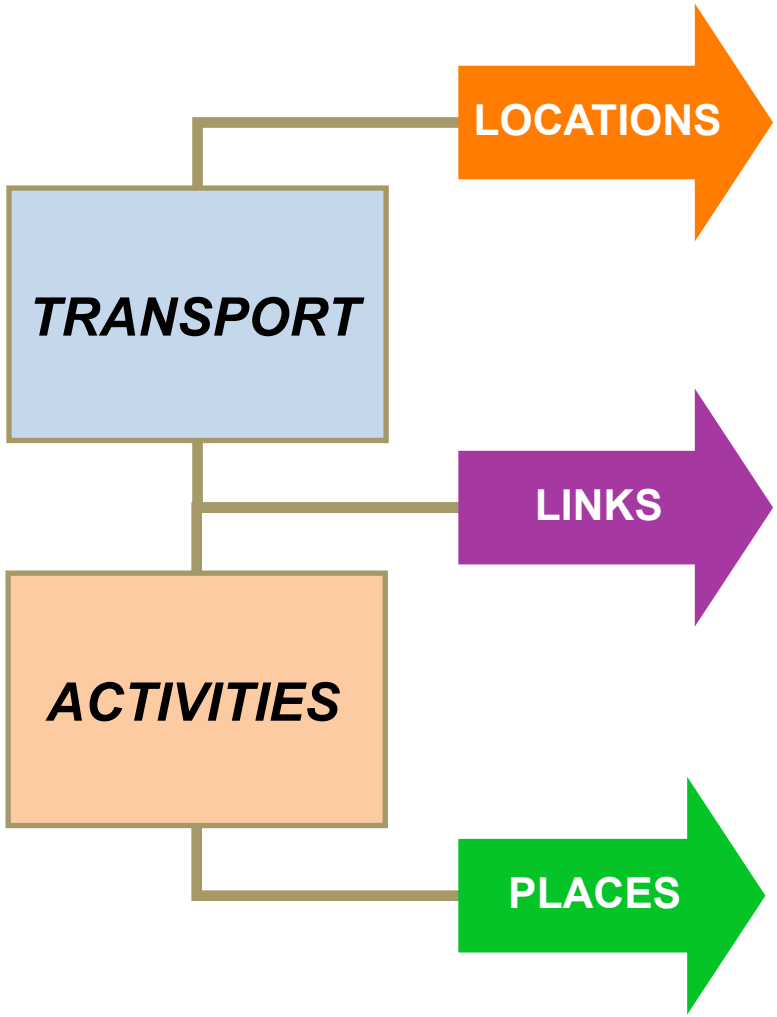
BENEFITS

- Individual health
- Commuting costs (time, $, parking)
- « For the Planet »

RISKS

- Health (inhalations)
- Safety (mortality, morbidity, costs)

COMPLEX SYSTEMS



Nested scales
(Non) Linearity
Exo-/endo-geneous
(Sub-)optimal
Static – dynamic
Open systems
Emergence
Stochastic
Self-organisation
…

Inspired from : Rodrigue, 2014

1. Spatial is special.

- **Location**(s) and **distance**(s)
- **Scale** (nested and interdependant scales)
- **COMPLEXITY** of spatial processes
- **UNCERTAINTY**
- MAUP, heterogeneity, border …

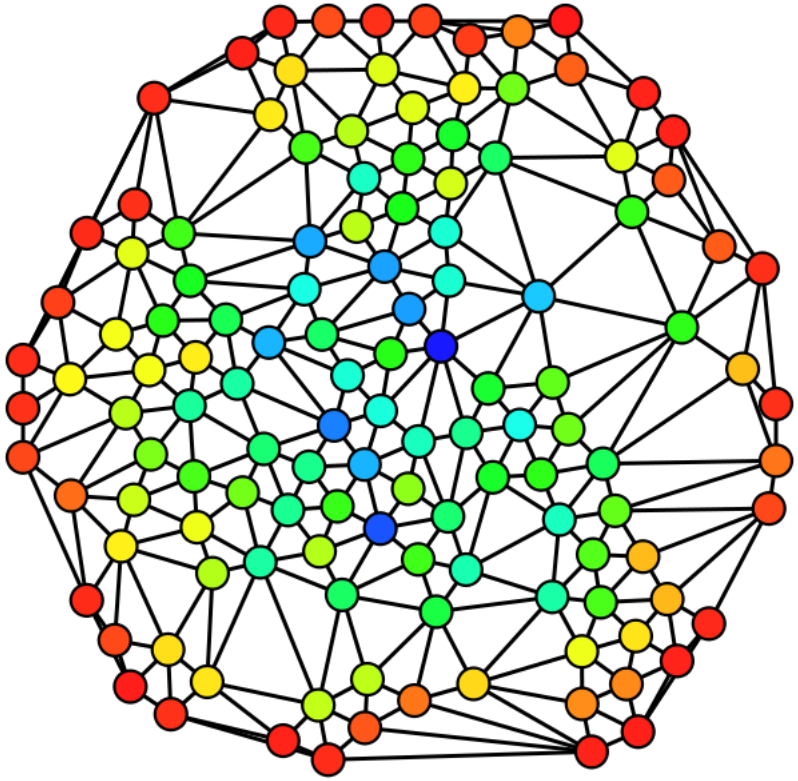
Econometrics, spatial analysis, GIS

2. New ICT « big/soft » data

BUT
- we need to capture the meaning of data, not just the data itself – epistemological implications of the big data revolution (rapid changes)
- we need to develop and understand methods and link them with existing spatial (urban) theories

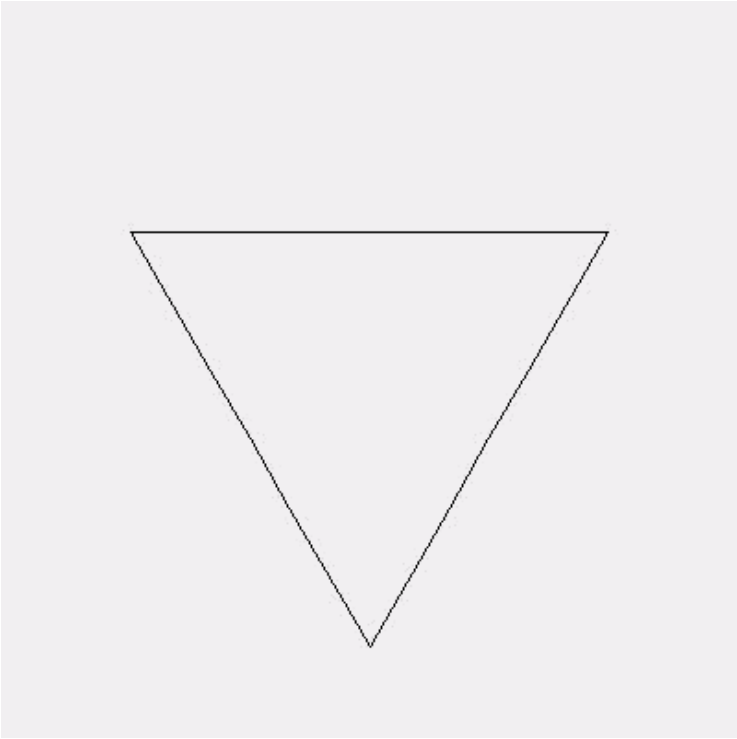
Networks, transport geography

Networks
(People centered communities)



- ***Can represent relationships at a variety of scales at once.***
- Structural properties of networks provide means of understanding how they work > Rd Acc.
  - Nodes and links, direction
  - Degree centrality and betweenness

Fractals
(Place based morphologies)



- The same pattern appears across all scales. Scale invariant.
- The relationship between size of box and pattern in it is constant.
- Fractals follow their own power law relating how number of boxes needed to cover a shape change in relation to their size.

3. We cannot do without models, whatever they are.

« The need for theory is of even greater significance that it ever was and as data volumes grow the need to approach such bigness with clear theory has never been more important »

M. Batty, 2008

.

Full details about the examples are to be found in

- Thomas I. (1996), Spatial Data Aggregation. Exploratory Analysis of Road Accidents. *AAP*, 28:2, 251-264
- SteenberghenT. *et al*. (2004) Intra-urban location of road accidents blackzones: a Belgian example. *IJGIS*: 18,2, 169-181.
- Vandenbulcke G.,. *et al.* (2011) Bicycle commuting in Belgium: Spatial determinants and re-cycling strategies, *TR – A* 45 118–137
- Thomas I., Frankhauser P. (2013) Fractal dimensions of the built-up footprint: buildings versus roads. Fractal evidence from Antwerp (Belgium). *Environment and Planning B*, 40, 310-329.
- Vandenbulcke G., Thomas I., IntPanis L. (2014), Predicting cycling accident risk in Brussels: an innovative spatial case-control approach. *AAP*, 62, 341-357