

# Building size modelization

Arlette Antoni and Thierry Dhorne

Laboratoire Lab-STICC DECIDE - UMR 2685  
Vannes

OGRS Perugia - Oct 2016



- 1 Framework
- 2 Approach
- 3 Applications
- 4 Conclusion

it is a fact that

- physical models (noise, temperature ...) need more and more integration of building structures specially when dealing with cities
  - it is very difficult to provide deterministic modelization of building structures
- there is a need for stochastic modelizations of such structures

"... but hardly any work exists on the properties of spatial distributions within individual cities,..."

in

"Scaling and allometry in the building geometries of Greater London" - 2008

by

Michael Batty , Rui Carvalho , Andy Hudson-Smith , Richard Milton , Duncan Smith and Philip Steadman

- a possible way would be to provide stochastic models from expert knowledge
  - an other way is to use general stochastic (spatial or spatio-temporal) models statistically adjusted
  - indeed statistics appear to be more and more used in interaction with GIS
- it can be suited for the analysis of urban data (classification, clustering, ...)  
**but there is a lack of operational interaction (platforms)**

- provide a stochastic spatial modelization of urban structures enabling statistical manipulation like
  - parameter estimation (characterization)
  - clustering
  - simulation (inclusion in other models)

# Modelization of structures

- Through point (or set) marked processes
- modelization of marks = multivariate features
- modelization of spatial interaction ( between points=buildings) = spatial process

# First step: modelization of features

- from a statistical point of view for efficiency
  - to avoid redundancy
  - to achieve completeness
- ≠ experts indicators
- which are generally functional transformations of very few initial data (2 or 3 variables)
  - high multilinear correlations



# 1.1 Step: modelization of size variables

- as suggested by Batty & al
- e.g area see graphic beside



GIS buildings layer

# Notice

- high skewness
  - already noted by Batty & al
- generally log corrected
  - should be done as default when any statistical calculation is done (even a mean)
- suggests that expert indicators cannot be used because made on the initial variables

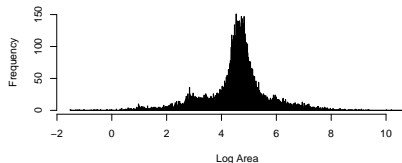
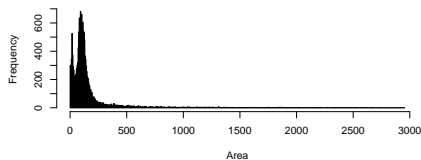
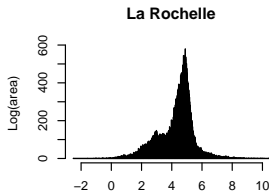
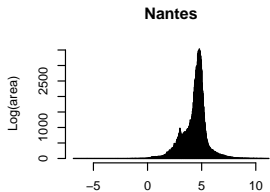
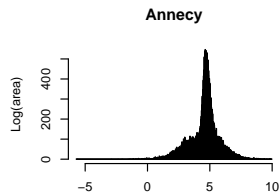
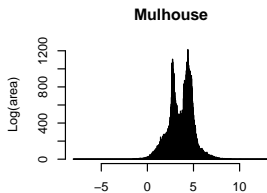
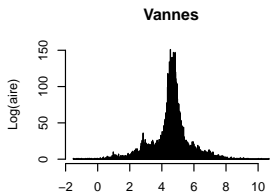


Figure: Log transformation

# Different distributions for cities



Histograms of log-transformed data

# Noticing (following)

- multi modality
  - not mentioned by Batty & Al
  
- 2 to 7 sub populations
  - not ignorable
    - to less or to much
- not gaussian
- not balanced (varying proportions)
- interesting ?
  - for structuration and clustering

Different laws suggested in literature

- non gaussian components
  - as noted by Batty et al. (distribution for allometry)
  - log Laplace
  - log hyperbolic

# Distribution

All laws have an exponential form,

- Log hyperbolic
  - Bagnold and Barndorff-Nielsen (Bagnold & al , 1980)
  - to model the particle-size distribution of naturally occurring sediments.
  -

$$\frac{1}{2\delta\sqrt{1+\pi^2 K_1(\zeta)}} \exp\left[-\zeta\sqrt{1+\pi^2}\sqrt{1+\left(\frac{x-\mu}{\delta}\right)^2} - \pi\left(\frac{x-\mu}{\delta}\right)\right]$$

where  $K_1()$  is the modified Bessel function of the third kind and order 1.

- Log skew Laplace model
  - proposed by Fieller & al
  - **simpler alternative** to the log-hyperbolic distribution. Its density (on a logarithmic scale) is:

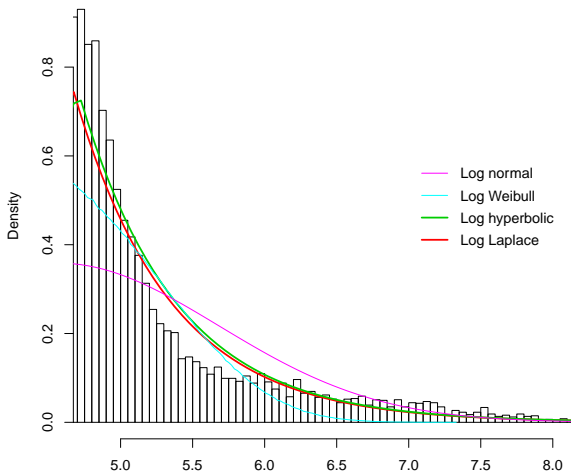
$$f(x) = \begin{cases} \left(\frac{1}{\alpha+\beta}\right) \exp\left(\frac{\mu-x}{\alpha}\right), & x \leq \mu \\ \left(\frac{1}{\alpha+\beta}\right) \exp\left(\frac{x-\mu}{\beta}\right), & x > \mu \end{cases}$$

with  $\alpha$  and  $\beta$  the left and right slopes, in the symmetric case  $\alpha = \beta$



# Adjustement

On the right part of distribution  $Area > 110m^2$ , for one sub population



# Simple clustering on area Building

We suggest 5 clusters

Classes on $\log(\text{Area})$	mean(area)	counts
$< 1.8$	3.28	346
$1.8 \leq \log(\text{Area}V) < 2.6$	9.91	471
$2.6 \leq \log(\text{Area}V) < 3.6$	23.33	1631
$3.6 \leq \log(\text{Area}V) < 5.7$	116.06	12260
$5.7 \leq \log(\text{Area}V) < 6.8$	501.23	1199
$\log(\text{Area}V) \geq 6.8$	2052.2	541

- from a manual (non reproducible at the moment !) maximum likelihood maximization

→ frequency with strong disproportions makes problems (estimation problems)



# Clustering illustration

- by Voronoï on centroid's building
- buildings' areas reveal zone

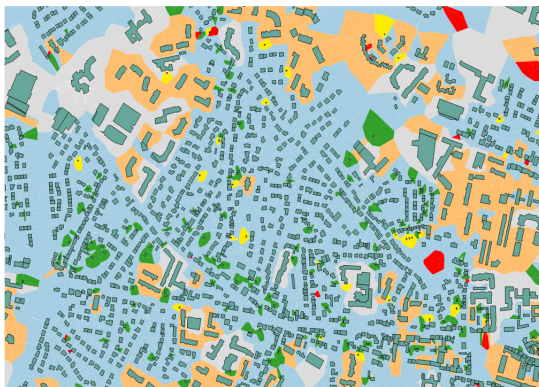


Figure: Voronoï : 5 clusters

# Spatial distribution of clusters

- not uniformly distributed

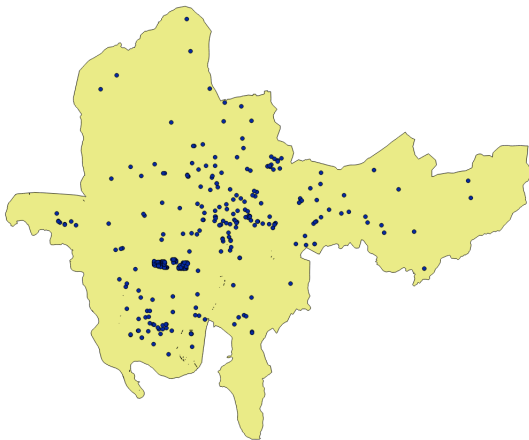


Figure: small size cluster location

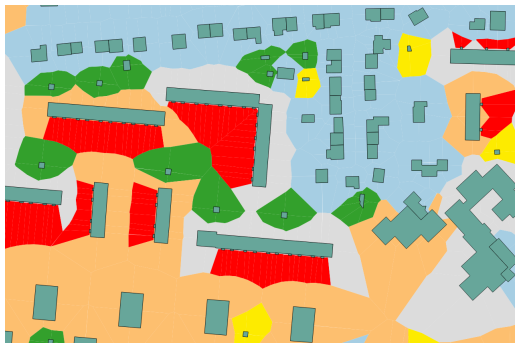


Figure: small building's influence

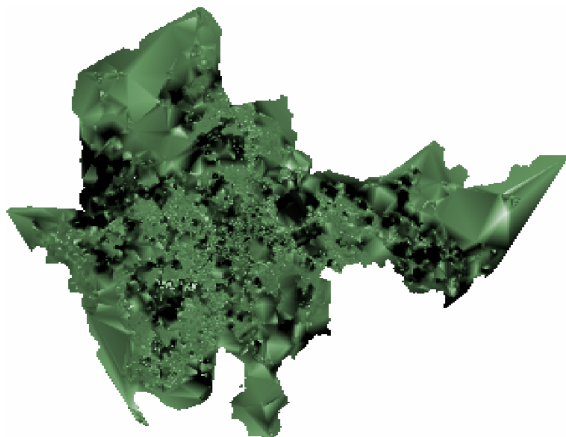


Figure: City structuration

# Conclusions

- this is not reproducible research
  - but improvable research
- because
- of the lack (in France of course!) of reliable "official" data bases (how to improve them?)
  - of the difficulty to interact between GIS and statistical tools (thanks to our colleagues this bug is fixed)
  - we have to work hard on suitable models and corresponding estimators
- thanks to QGIS, R, knitr,  $\LaTeX$ (and other open... things)