

Building size modelization

Arlette Antoni¹ and Thierry Dhone¹

¹Lab-STICC - CNRS UMR 6285, Université de Bretagne Sud, 56000 Vannes France

ABSTRACT

New challenges in the efficient management of cities depend on a deep knowledge of their inner structures. It is therefore very important to have access to reliable models of cities characteristics and organization. This paper aims at providing and validating a stochastic modelization based on statistical data of buildings parameters which can be useful as an entry for many other models considered in a wide range of fields where buildings structure is a main factor of a thorough modelization of cities.

Keywords: Building size distributions, building geometry, statistical modelization, R language

INTRODUCTION

As quoted by Michael Batty in 2008 ([2]) but still up to now “hardly any work exists on the properties of spatial distributions within individual cities” and particularly in what concerns buildings. A better knowledge of analytical properties of buildings is necessary in order to manage the key problems of to-day cities, in the sense that buildings support an important part of the general problems concerned by sustainable development and urban planning.

The usual way to manage building information in physical models is rather analytic and indeed not many work has been done on stochastics approach in this field.

The aim of the present talk is to provide elements of stochastic modelization of building characteristics with a special emphasize on size (area, height,...).

The level of modelization is the individual level, i.e. the building is considered in itself. Of course, a higher level, specially a spatial level should also be considered to cope with spatial organization structures, but this level is highly depending on the first level analysis coped with in this paper.

This type of modelization appears to be necessary (and relatively efficient) for :

- quality of initial data (corrections, normalizations,...),
- enrichment of data,
- classification (supervised or unsupervised) of city (districts),

the last point probably requiring a spatial modelization.

The study is concentrated on french urban areas but could be extended to other European countries by means of suitable data thanks to the french ANR project MAPUCE : <http://www.cnrm-game-meteo.fr/spip.php?article787&lang=fr>.

DATA

Though the general framework of this study concerns french urban areas, the data studied are restricted to some specific french cities. The initial data treatments are indeed rather heavy due to:

- the number of buildings,
- the necessity to control even roughly data quality,
- the need for an expert knowledge
- the specificity of the calculations.

39 Data come from french topographic data base: BD-TOPO. In this data base, each building is repre-
40 sented as a geometrical object and more precisely as a polygon (figure 1).

41 Classical spatial operators are then applied to these raw data in order to extract spatial attributes (or
variables) from perimeter, area, up to convexity index and fractal dimension,...



Figure 1. Representation of the GIS buildings layer

42

43 The study has been focused on five mid to big french cities:

City	Area km2	Inhabitants nr (2013)	Buildings nr
Anncy	13.65	52 029	37 271
La Rochelle	28.43	74 344	78 896
Mulhouse	22.18	112 063	126702
Nantes	65.19	292 718	262643
Vannes	32.30	53 032	16448

44

45 DATA ANALYSIS

46 The whole data analytics is supported by the R language for statistical analysis. The area is the main
47 variable studied. The initial data analysis reveals a high skew-symmetry of this variable when narrow
48 binned histograms are considered.

49 Such skew-symmetries suggest the use of a logarithmic transformation which better reveals the essence
50 of the phenomenon involved as represented in figure 2.

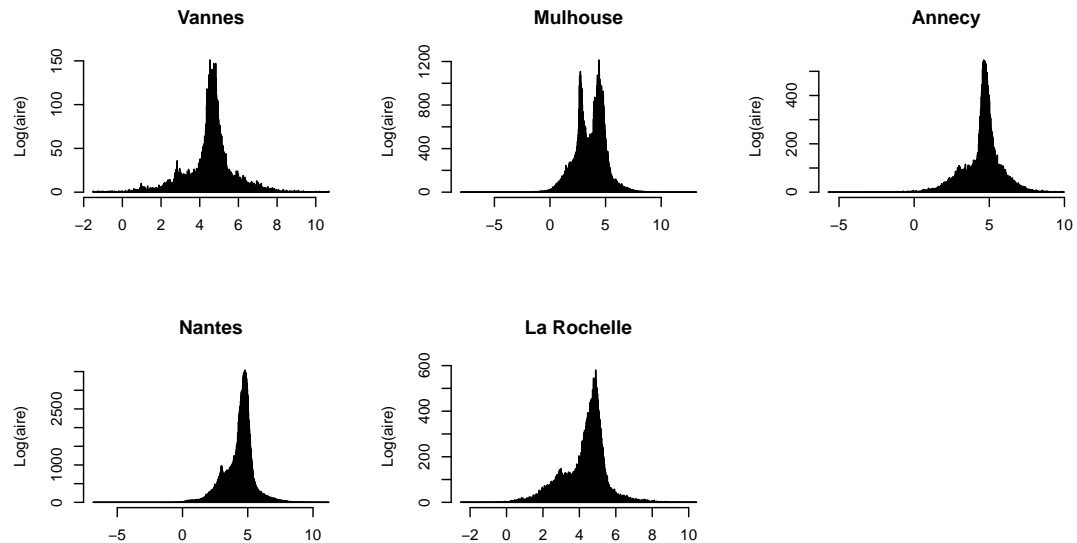


Figure 2. Histograms of log-transformed data

51 These histograms (figure 2) let appear a relative but meaningful multi-modality which can be important
 52 in some cities (for instance Mulhouse). The rather sharp distribution suggest that a gaussian model, usually
 53 considered by non specialists, is unsuited. Moreover, it seems that the slopes of left and right parts of the
 54 distributions are somewhat different. A thorough study of the data is then necessary.

55 LITTERATURE MODELS

56 Size models have been widely used in the literature but as quoted before not for within cities studies and
 57 particularly for buildings studies.

58 First attempt to modelize size variables have been proposed for particles and then extended to other
 59 topics. The main models considered are the followings:

- Weibull or Rosin Rammler distribution is a useful distribution for representing particle size distributions generated by crushing operations. Its density is:

$$f(x, \lambda, k) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} \exp - \left(\frac{x}{\lambda}\right)^k$$

60 with K , λ shape and scale parameters.

- Log Normal distribution is often used to approximate the particle size distribution. Its density (on a logarithmic scale) is:

$$f(x) = \frac{1}{\sigma * \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2\right)$$

- 61 • Log Hyperbolic distribution was proposed by Bagnold and Barndorff-Nielsen ([1]) to model the
 62 particle-size distribution of naturally occurring sediments.

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

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

- 64 • Log skew Laplace model was proposed by Fieller, Gilbertson and Olbricht ([3]) as a simpler
65 alternative to the log-hyperbolic distribution. Its density (on a logarithmic scale) is:

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

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

- Log double Pareto distribution appears also to be suitable as it is easier to manage than hyperbolic. Its density (on a logarithmic scale) is:

$$f(x) = \begin{cases} = \frac{\theta}{2\beta} \left(\frac{x}{\beta}\right)^{\theta-1}, & 0 < x \leq \beta \\ \frac{\theta}{2\beta} \left(\frac{\beta}{x}\right)^{\theta+1}, & x > \beta \end{cases}$$

67 Distributions functions are either power-law or exponential. The model complexity comes from the
68 number of parameters. Four out of these five models are based on logarithmic transformation.

69 COMPONENT MODELIZATION

70 The search for a suitable model for individual components identified in the population is rather difficult in
71 presence of other mixed components. As seen on the histograms presented above, there is fortunately a
72 major component corresponding to the bigger sizes. It is therefore possible to isolate such components in
73 each city and to perform an analysis of the type of decreasing frequency distribution in order to choose
74 the best model. Among modelization tested, Weibull and Log Normal appear to be poorly efficient. On
75 the other hand, Log Hyperbolic is not suited for modelling tails. Attention is then restricted to the two
76 remaining distributions.

77 MIXTURE PARAMETER ESTIMATION

78 Once selected the appropriate model for any of the components, it is possible to modelize the data by a
79 mixture of such components. The remaining problem consists in estimating the individual parameters of
80 each component and the proportion (or conditional probability) of each (minus 1 due to the normalization
81 constraint).

82 This is done by maximum likelihood estimation. An algorithm based on expectation-maximization
83 has been developed and applied to the five cities data. A set of R functions is available for further use of
84 these techniques.

85 REFERENCES

- 86 [1] Bagnold, R. and Barndorff-Nielsen, O. (1980). The pattern of natural distribution. *Sedimentology*, 27
87 (2):199:207.
88 [2] Batty, M., Carvalho, R., and Hudson-Smith, A. (2008). Scaling and allometry in the building
89 geometries of greater london. *Physics of Condensed Matter*, 63 (3):303:314.
90 [3] Fieller, N.R.J; Gilbertson, D. and Olbricht, W. (1984). A new method of environmental analysis of
91 particle size distribution data from shoreline sediments. *Nature*, 311 (5987):648:651.