



**HAL**  
open science

## Applying two fractal methods to characterise the local and global deviations from scale invariance of built patterns throughout mainland France

François Sémécurbe, Cécile Tannier, Stéphane Roux

### ► To cite this version:

François Sémécurbe, Cécile Tannier, Stéphane Roux. Applying two fractal methods to characterise the local and global deviations from scale invariance of built patterns throughout mainland France. *Journal of Geographical Systems*, 2019, 21 (2), pp.271-293. 10.1007/s10109-018-0286-1 . hal-02066302

**HAL Id: hal-02066302**

**<https://hal.science/hal-02066302>**

Submitted on 16 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Applying two fractal methods to characterise the local and global deviations from scale invariance of built patterns throughout mainland France

François Sémécurbe <sup>1</sup>

Cécile Tannier <sup>1</sup>

Stéphane G. Roux <sup>2</sup>

<sup>1</sup>Laboratoire ThéMA, CNRS-Univ. Bourgogne Franche-Comté. 32, rue Mégevand, F-25 030 Besançon Cedex, France

[françois.semecurbe@univ-fcomte.fr](mailto:françois.semecurbe@univ-fcomte.fr)

[cecile.tannier@univ-fcomte.fr](mailto:cecile.tannier@univ-fcomte.fr)

<sup>2</sup>Laboratoire de Physique UMR 5672, CNRS-Ecole Nationale Supérieure de Lyon. 46, allée d'Italie, F-69364 Lyon Cedex 07, France

[stephane.roux@ens-lyon.fr](mailto:stephane.roux@ens-lyon.fr)

This is a postprint version, the definitive version of this article has been published in SÉMÉCURBE F., TANNIER C., ROUX S. G. Applying two fractal methods to characterise the local and global deviations from scale-invariance of built patterns throughout mainland. *Journal of Geographical Systems*, Springer Verlag, 2018, pp.1-23.

DOI : 10.1007/s10109-018-0286-1

---

**Abstract.**

In the early twentieth century a handful of French geographers and historians famously suggested that mainland France comprised two agrarian systems: enclosed field systems with scattered settlements in the central and western France, and openfield systems with grouped settlements in eastern France. This division between grouped and scattered settlements can still be found on the outskirts of urban areas. The objective of this paper is to determine whether the shape of urban areas varies with the type of built patterns in their periphery. To this end, we identify and characterise the local and global deviations from scale-invariance of built patterns in mainland France. For this, we propose a new method –Geographically Weighted Fractal Analysis – that can characterise built patterns at a fine spatial resolution without making any a priori distinction between urban patterns and suburban or rural patterns. By applying GWFA to the spatial distribution of buildings throughout mainland France we identify six geographically consistent types of built patterns that are distinctive in the way buildings are either concentrated or dispersed across scales. The relationship between the local built textures and the global shape of twenty metropolitan areas is then analysed statistically. It is found that the proportion of dispersed (or concentrated) outer suburban built patterns in metropolitan areas is closely related to the distance threshold that marks the morphological limit of their urban areas.

**Keywords.**

Built textures, fractal analysis, suburban fringes, scale-invariance, mainland France.

## 1. Introduction

If we are to identify the advantages and disadvantages of different city shapes for various planning goals (e. g. preserving ecological connectivity, improving access to urban and rural amenities, ensuring good ventilation of the city centre, etc.) then – among other things – the associated urban built patterns need to be more accurately described and characterized. This need has engendered a wealth of publications describing and characterizing city shapes, including numerous methods of identifying different types of urban patterns along with many spatial indexes for measuring urban sprawl. In this paper, we seek to contribute to this field of research by exploring the multiscale morphological properties of urban built patterns in more depth.

In the early twentieth century, a number of renowned geographers and historians analysed the geography of rural France (Demangeon, 1927; Bloch, 1931; Dion, 1934). Their research led to a classical division of mainland France into two agrarian systems: enclosed field systems associated with scattered settlements in the centre and the West of France, and openfield systems associated with grouped settlements in the East of France. This distinction between grouped and scattered settlements can still be found on the periphery of urban areas. Obviously, the spatial expansion of cities does not take place within a blank, isotropic space, especially when cities belong to centuries-old settlement systems. As outer suburban areas are the space into which cities expand, it is to be expected that pre-existing suburban settlements will influence the shape of the urban development they come to accommodate. Regarding this, the objective of this paper is to determine whether the shape of French urban areas is determined by the shape of extant built-up areas on their outskirts.

Analysis of the relationship between built-up shapes surrounding urban areas and the shape of the urban areas themselves first involves characterizing the built patterns at a fine spatial resolution, without making any a priori distinction between urban patterns and suburban or rural patterns. Yet, built patterns in suburban fringes are somewhat hybrid: they blend inherited features (traditional rural buildings, old village cores) with new urban developments (most often, detached housing estates). This makes them difficult to describe and characterize (Chaudhry & Mackaness 2008). A number of publications have shown the value of fractal dimensions for characterizing irregular shapes of this kind. Following pioneering research supporting the assumption that cities are fractal (Batty & Longley, 1986, 1987; Frankhauser, 1988; Mandelbrot, 1982; White & Engelen, 1993), a string of publications has explored the fractal aspects of built forms (e. g. among the earliest ones, Longley and Batty, 1989; Batty & Xie, 1996; Frankhauser, 1998; Benguigui et al, 2000; Shen, 2002; Benguigui & Czamanski, 2004). Fractal analysis can be used to compare the built patterns of entire urban regions (e. g. Feng & Chen, 2010; Chen & Wang, 2013; Tannier & Thomas, 2013) or smaller spatial units (neighbourhoods, communities) (e. g. De Keersmaecker et al, 2003; Thomas et al, 2008). Fractal dimensions describe the way buildings are spatially distributed at several nested spatial resolutions. The closer the fractal dimension is to 2, the more uniform is the spatial distribution of buildings. In contrast, a fractal dimension close to 0 describes a highly concentrated spatial distribution of buildings at just a few locations across all scales. More often than not, buildings exhibit concentrated spatial configurations at some spatial resolutions and more dispersed

patterns at other resolutions (fractal dimension in-between 1 and 2). In any event, the fractal dimension is not expected to be constant in reality (Goodchild, 1980); most often, it is constant over a limited range of scales but varies somewhat over successive ranges of scales (Lam, 1990; Lam & Quattrochi, 1992; White & Engelen, 1994; Tannier & Pumain, 2005).

In this paper, we propose a new method, Geographically Weighted Fractal Analysis (GWFA). As with Geographically Weighted Regression (Brunsdon, Fotheringham & Charlton, 1996; Fotheringham, Brunsdon & Charlton, 2003), GWFA uses a mathematical kernel that describes the way the neighbourhood of a point is taken into account to estimate the fractal dimension of that point. By applying GWFA to French built patterns, we analyse the local deviations from scale-invariance of the spatial distribution of buildings throughout mainland France. Then the classification of analysis results enables us to map the built textures of France at a spatial resolution of 2000 m. Next, we compare the built textures thus mapped with the global shape of twenty metropolitan areas. For this, we choose to apply the method proposed by Tannier et al. (2011) to detect discontinuities in space across scales. This method can determine a distance threshold specific to each area under study. The statistical comparison of this distance threshold with the local built textures resulting from the classification indicates whether or not the global shape of French metropolitan areas varies with the built shapes in their periphery.

## 2. Data and method

Source data are from the vector data to the nearest 1 m provided by the French cartographic service, BD Topo ® IGN 2011. This represents about 24 million buildings mapped in 2D. It would have been possible to use another data source providing similar kind of data, especially Open Street Map. Yet, building data from BD Topo ® IGN are easier to gather and to use than OSM data when analysing the whole mainland France. In general, OSM data are more accurate than data from BD Topo ® IGN in dense urban areas. Conversely, data from BD Topo ® IGN are more accurate than OSM data in rural areas visited by a small number of individuals. For the analysis, each building is represented by its centroid in order to reduce both the size of the data base and the computation time.

In order to analyze this very detailed data set, a four-steps methodology has been defined. It takes into account three guiding principles: 1) introducing no a priori distinction between urban built patterns and suburban or rural patterns; 2) focusing the analyses on the deviations from scale-invariance; 3) examining and confronting both global and local deviations from scale-invariance.

**Step 1.** *Determining if the spatial distribution of buildings in mainland France is locally scale-invariant or not*

Analysis is based on a 2000 m-spaced grid of 145,178 estimation points  $i$  that covers the whole country. On this basis, GWFA is used to estimate the fractal dimension at each point  $i$  of the study area (see section 3). Fractal dimensions describe the way buildings are

spatially distributed at several nested spatial resolutions. The underlying assumption is that the spatial distribution of buildings is scale invariant, which means that the variation in the built density from one given spatial resolution to another is constant across scales. The scale-invariance is proven when the goodness-of-fit of the estimation of the fractal dimension is satisfactory. This goodness-of-fit is usually very high when estimating fractal dimensions. Therefore,  $R^2$  values between 0.95 and 0.995 do not guarantee scale-invariance for the whole scale range, and  $R^2$  values less than 0.95 are sure signs of a disruption in scale-invariance.

**Step 2.** *Identifying different types of built patterns according to the way the spatial distribution of buildings deviates locally from a scale-invariant distribution.*

Here the spatial distribution of buildings at each point  $i$  is not supposed to be scale invariant and each point  $i$  is characterized by a series of scaling indexes  $Sc$  calculated for successive scale intervals. These  $Sc$  indexes correspond to the  $\alpha(\epsilon)$  indexes used by Thomas et al. (2010) to characterize the scaling behaviour of built patterns of European urban wards.

In order to analyse the profile of  $Sc$  indexes for each point  $i$ , we apply a Principal Component Analysis. Then we classify the 145,178 points  $i$  according to their two coordinates on the PCA axes using the k-means method.

**Step 3.** *For a selection of twenty French metropolitan areas, characterizing the global deviation of the spatial distribution of buildings from a scale-invariant distribution. For this, calculation of a distance threshold that indicates a crucial discontinuity in space across scales.*

For this, we use the free software application named MorphoLim (<https://sourcesup.renater.fr/morpholim/>) that enables us to apply the method proposed by Tannier et al. (2011). By this method a dilation of each individual building is first applied, then a distance threshold is detected on the dilation curve. This distance threshold corresponds to the dilation step at which the curve deviates the most from a straight-line. At this point on the dilation curve, distances separating buildings no longer exhibit the same fractal behaviour. Thus this point indicates the maximum morphological difference between the inlying built patterns (i. e. the morphological agglomeration) and the outlying built patterns (i. e. the surrounding suburban areas). The distance threshold is specific to each area under study: the less an urban area differs morphologically from its surrounding built landscape (in terms of variations in distances separating buildings), the greater the distance threshold; the fewer the buildings in a study area that are connected across scales, the greater the distance threshold.

**Step 4.** *For the twenty selected French metropolitan areas, making a statistical comparison of the distance threshold that characterizes globally the spatial distribution of buildings across scales with the dominant types of local built textures identified at step 2.*

### 3. Geographically Weighted Fractal Analysis (GWFA): detailed description

#### 3.1 Estimation of the fractal dimension at each point $i$ of the study area

The fractal dimension  $D$  of an object is defined through a scaling function  $F_\varepsilon$  that describes an aspect of the spatial distribution of the points  $e$  of the object.

$$F_\varepsilon \sim \varepsilon^{-D} \text{ where } \varepsilon \text{ is the size of the counting window} \quad (1)$$

For theoretical mathematical objects,  $D$  corresponds to the asymptotic behaviour of  $F_\varepsilon$  when  $\varepsilon$  tends to 0. For real world objects,  $D$  is estimated for the range of scales  $\varepsilon$  for which the points of the curve of  $\log(F_\varepsilon)$  with respect to  $-\log(\varepsilon)$  form a straight line.  $D$  is the slope of this straight line.

Each scaling function  $F_\varepsilon$  enables the calculation of a given fractal dimension. The chosen scaling function for Geographically Weighted Fractal Analysis (GWFA) is  $M_r^q$ , where  $r$  corresponds to  $\varepsilon$  and  $M$  corresponds to  $F$ .

$$M_r^q \sim r^{Dq} \quad (2)$$

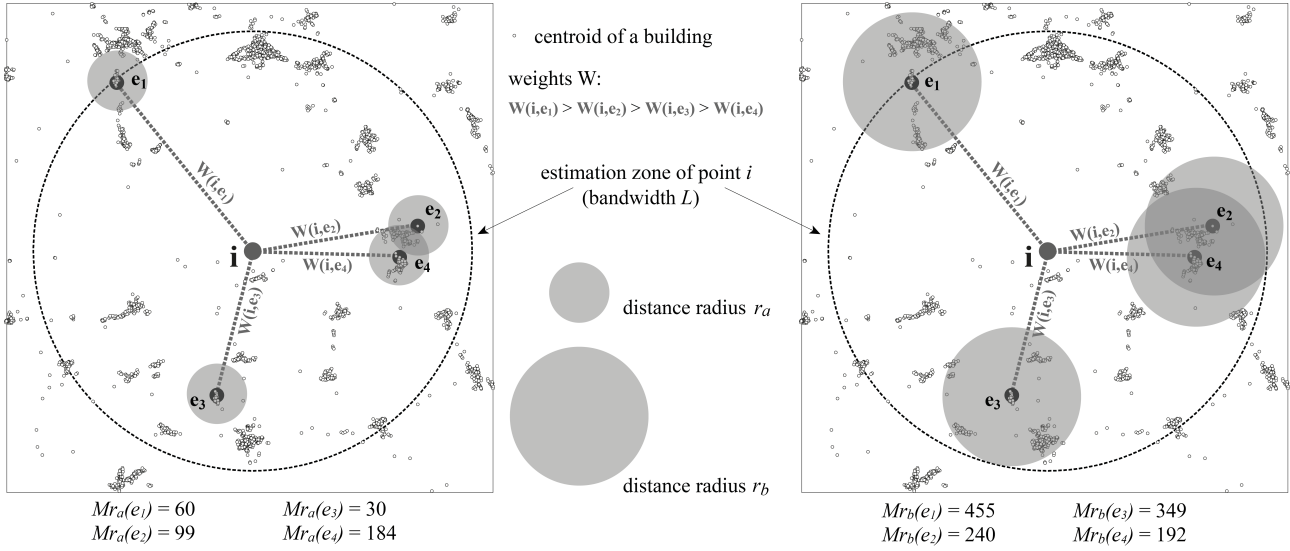
For a given point  $e$ ,  $M_r(e)$  is the number of points within a distance  $r$  around  $e$ .  $M_r^q$  is obtained by aggregating the  $M_r(e)$  of all points  $e$ . When a simple average is used for the aggregation, the fractal dimension obtained is the correlation dimension (Grassberger & Procaccia, 1983). In contrast, when a mathematical norm is used for the aggregation (eq. 3), the fractal dimension obtained is a multifractal generalized dimension  $D_q$ . The corresponding estimation method is the sandbox method (Vicsek, 1990).

$$M_r^q = \left[ \frac{1}{n} \sum_e M_r(e)^{q-1} \right]^{\frac{1}{q-1}} \sim r^{D_q} \quad (3)$$

The parameter  $q$  controls the nature of the aggregation: more importance is given either to high values of  $M_r(e)$ , when  $q$  is high, or to low values of  $M_r(e)$ , when  $q$  is low. In GWFA,  $q$  is set to 0 meaning that high or low values of  $M_r(e)$  all have the same importance. The corresponding scaling function is noted  $M_r(i)$  (eq. 4) as it is calculated for each estimation point  $i$ . The resulting generalized dimension  $D_0$  is closely related to, but more reliable than, the well-known box-counting dimension (Tél, Fulop & Vicsek, 1989; Vicsek, 1990).  $D_0$  is obtained by estimating the slope of the curve of  $\log(M_r(i))$  with respect to  $\log(r)$  for a given set of distance radii  $r$  by OLS (ordinary least squares) regression.

$$M_r(i) = \left[ \sum_e (W_{i,e} M_r(e))^{-1} \right]^{-1} \sim r^{D_0} \quad (4)$$

For counting, a circular neighbourhood of diameter  $L$  is defined around each estimation point  $i$ . Only points  $e$  located within this neighbourhood  $L$  are taken into account but points counted around each  $e$  may be located outside the neighbourhood  $L$  (**Figure 1**). In this way, boundary effects related to points  $e$  located close to the boundary of the neighbourhood  $L$  are avoided.



**Figure 1.** Counting variables

Weights  $W_{i,e}$  are attributed to the  $M_r(e)$  values. Each weight corresponds to the distance of the point  $e$  under consideration to the estimation point  $i$  (**Figure 1**). Thus GWFA introduces distance-based weights assigned to each pair of points  $(i,e)$  in the calculation of the scaling function  $M_r(i)$ .  $M_r(e)$  values for points  $e$  close to  $i$  have a higher weight in the estimation of  $M_r(i)$  than  $M_r(e)$  values for points  $e$  that lie far from  $i$ . Weights  $W_{i,e}$  are calculated by applying a mathematical positive kernel  $K$ :

$$W_{i,e} = K(d_{ie} / L)$$

where  $d_{ie}$  is the distance between the estimation point  $i$  and the observation  $e$ ;  $L$  is the bandwidth of the analysis.

For GWFA, the quartic (bi-weight) kernel is used.

$$K(x) = (1 - x^2)^2, \text{ if } x < 1,$$

and 0, otherwise

The quartic kernel resembles the Gaussian kernel but has a compact support and enables faster computation. In any event, a positive, continuous and decreasing kernel that tends asymptotically to 0 ensures regular and continuous estimations in space, which enables results to be mapped satisfactorily (Brunsdon, Fotheringham & Charlton, 1996).



GWFA can be computed using Python and/or R. The computing code written to calculate  $D_0$  for each estimation point  $i$  is provided in Appendices A and B. The computing time required for all buildings of mainland France (i. e. about 24 million building centroids) is about 5 minutes with a calculating server having a low computing power. Alternatively, when the data set is quite small (less than about 1 million building centroids), an R package entitled GWFA can be used (<https://CRAN.R-project.org/package=gwfa>).

### 3.2 Setting values of parameters

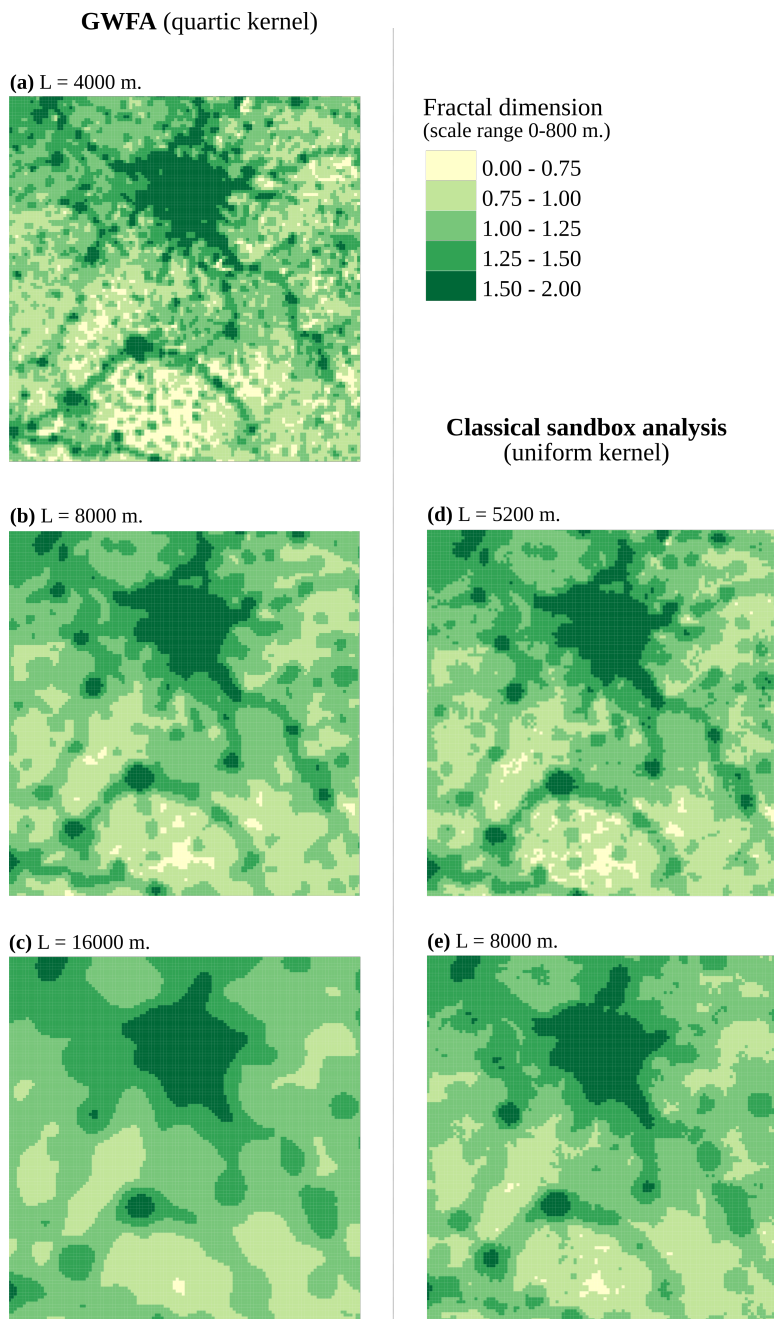
The distance between any two estimation points  $i$  is set at 2000 m. This distance is small enough to analyse local variations in fractal behaviour. At the same time, it is large enough to obtain a reliable estimation of fractal dimension for scales (i. e. distance radii  $r$ ) ranging from 0 to 800 m (see below).

The size of the neighbourhood  $L$  determines the fact that the analysis is local (when  $L$  is small) or global (when  $L$  is large). In any cases, the neighbourhood  $L$  around each estimation point  $i$  has to be larger than the distance between two neighbouring points  $i$  in order to enable the analysis of built textures in an almost continuous way. Accordingly,  $L$  is set at 8000 m. Thus it includes two neighbouring estimation points  $i$  in each direction around a given point  $i$ . On average, each zone of radius  $L$  around each estimation point  $i$  contains 8600 centroids of buildings.

Five distances  $r$  are considered: 50, 100, 200, 400, and 800 m. As the target estimated curve is a straight line, five points are enough to perform a statistical estimation of the fractal dimension that is reliable. Thus introducing additional distances  $r$  between 50 and 800 m. is useless. Of course, standard errors of the estimates may be rather large but if they are, this shows that the spatial distribution of buildings is locally not scale-invariant. The 1:10 ratio between  $r = 800$  and  $L = 8000$  ensures that the statistical estimations of fractal dimensions are robust. In addition, 100 circles of radius  $r = 800$  m can entirely cover each estimation zone of radius  $L$  around  $i$ . Introducing additional distances  $r$  larger than 800 m. would not be relevant because this would create boundary effects. Introducing additional distances  $r$  lower than 50 m. would also not be relevant because each point  $e$  corresponds to the centroid of a building but not to its real spatial footprint.

### 3.3 Sensitivity of estimated fractal dimensions to the chosen values of parameters

First, we have studied the results of GWFA when changing the size  $L$  of the neighbourhood (see **Figures 2a, 2b** and **2c**). It appears that the spatial structures revealed when  $L=4000$  m. exhibit too many unnecessary details. Contrarily, the spatial structures revealed when  $L=16000$  m. are too smoothed and not enough informative. In-between, the neighbourhood size  $L=8000$  m. represents an interesting compromise.



**Figure 2.** Comparison of local fractal dimensions obtained with different kernels and different bandwidths  $L$ . Paris's metropolitan area and its southern periphery.

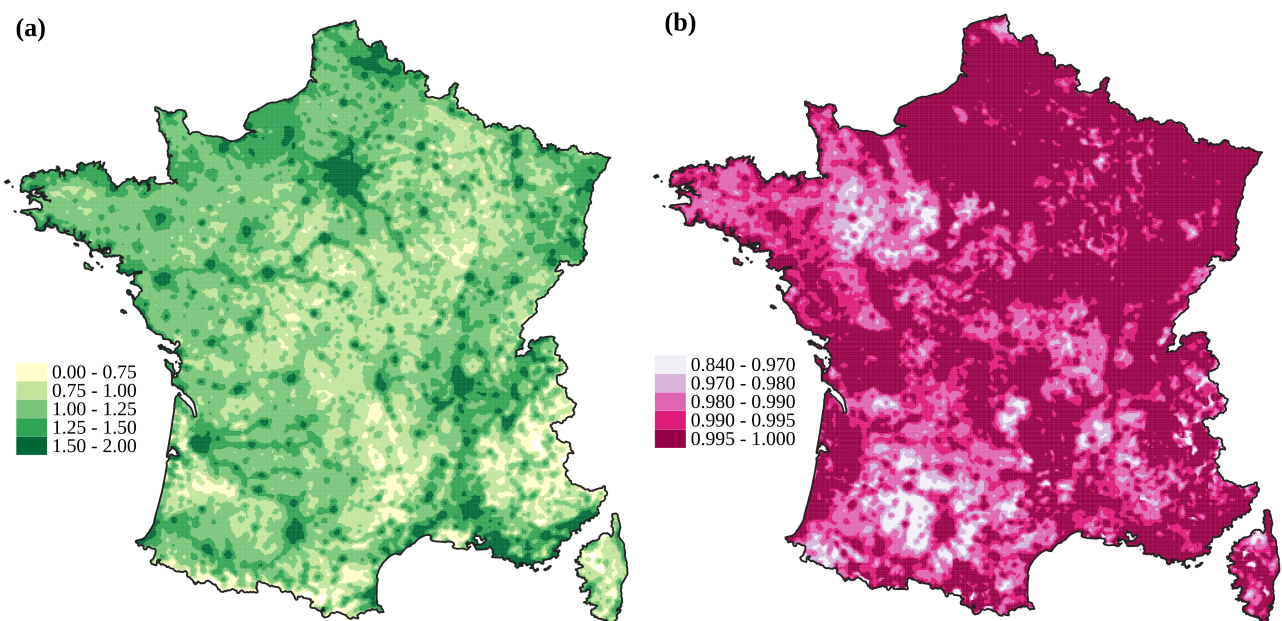
Second, we have compared the results obtained with GWFA and the results obtained with the classical sandbox fractal analysis method, which does not introduce geographically weightings. The classical sandbox method is equivalent to assigning to each point  $e$  a weight  $W_{i,e}$  equal to 1 when  $e$  lies within the bandwidth  $L$  around  $i$  and a weight  $W_{i,e}$  equal to 0 when  $e$  lies outside  $L$ . In other words, a uniform (boxcar) kernel is used for the classical sandbox method whereas a smoothing (quartic) kernel is used for GWFA. In order to ensure the comparability of analysis results obtained with the two methods, we have followed the suggestion of Diggle (1985) that is to calibrate the two kernels by equating their variance. By applying this equivalence rule, a bandwidth  $L = 8000$  m. in

case of a quartic kernel corresponds to a bandwidth  $L = 5200$  m. in case of a uniform kernel. Diggle (1985), quoting (Silverman, 1978), asserted that the successful implementation of smoothing data with a kernel function rests on choosing an appropriate value of the bandwidth  $L$ , whereas the choice of the kernel function is of very secondary importance. The fact is that the two maps on **Figure 2b** (GWFA with  $L=8000$  m.) and **Figure 2d** (classical sandbox analysis with  $L=5200$  m.) look a lot alike. Yet, the spatial structures are more clearly delineated on **Figure 2b**, which shows the interest of applying a smoothing quartic kernel. Additionally, spatial structures displayed on **Figure 2e** seem to us too smooth (not enough detailed) regarding both **Figures 2b** and **2d**.

## 4. Results

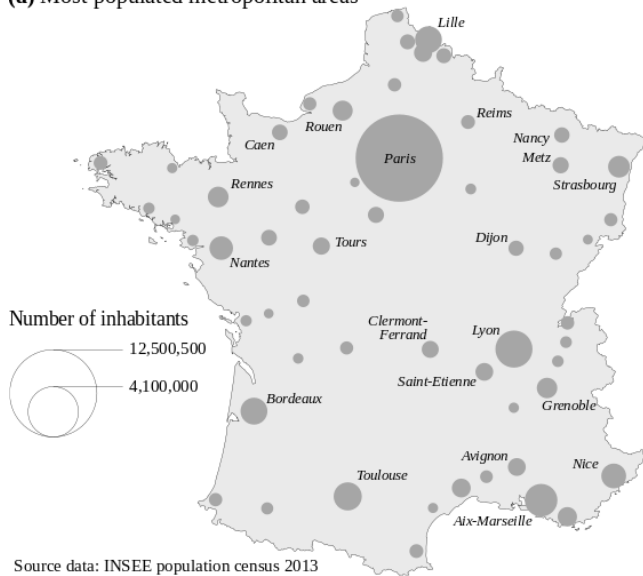
### 4.1 Estimated fractal dimensions show local deviations from scale-invariance, especially in the West and South-West of France

By applying GWFA to the building centroids for the whole of mainland France, local fractal dimensions were computed for the five distance radii  $r$  (50, 100, 200, 400, and 800 m) (**Figure 3a**). **Figure 3b** shows that the goodness-of-fit of the estimation is higher than 0.995 only in eastern France and in the largest urban areas (see **Figure 4** for their location). In contrast, most places in central and western France exhibit  $R^2$  values of less than 0.995. Thus, the spatial distribution of buildings there is probably not scale invariant.

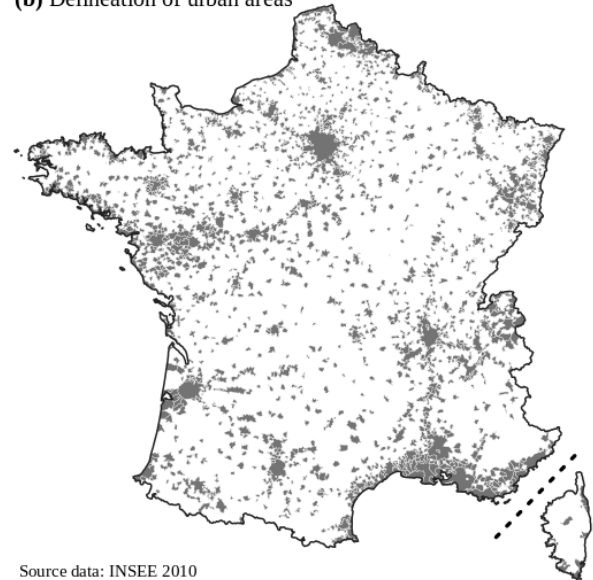


**Figure 3. (a)** Fractal dimensions estimated with GWFA for the scale range 0 to 800 m. **(b)** Goodness-of-fit of the estimation ( $R^2$ ). 145,178 estimation points  $i$ . On the maps, each estimation point corresponds to a 2000 m-wide square spatial unit.

(a) Most populated metropolitan areas



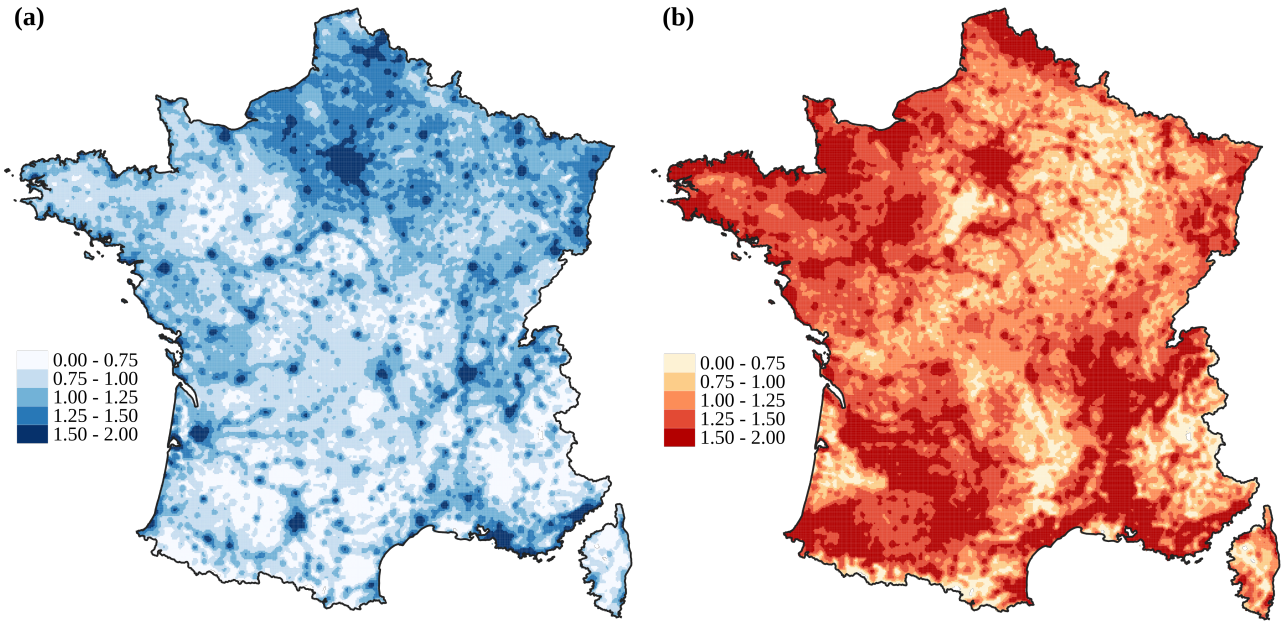
(b) Delineation of urban areas



**Figure 4.** Metropolitan areas and urban areas in mainland France. Note: in France, an urban area (*unité urbaine*) is an isolated *commune* (i. e. the basic administrative unit used for dividing up France) or a group of neighbouring *communes* having more than 2000 inhabitants and containing a contiguous built-up zone that gathers more than half of the population of each *commune*. A built-up zone is made up of adjacent buildings forming a whole such that no building is more than 200 metres apart from the next nearest building (Le Gléau, Pumain & Saint-Julien 1997).

Considering this, we estimated two other fractal dimensions, one for distance radii  $r$  50, 100, and 200 m and another for distance radii  $r$  200, 400, and 800 m. As expected, these two fractal dimensions are clearly different for places located outside the main urban areas in the West and South-West of France: in those places, estimated fractal dimensions are low for  $r$  50, 100, and 200 m, but high for  $r$  200, 400, and 800 m (**Figures 5a** and **5b**).

These results suggest that estimating one fractal dimension for the entire scale range (**Figure 3**) or two fractal dimensions for two smaller scale ranges (**Figure 5**) does not enable us to characterize the spatial distribution of buildings in each zone around  $i$  precisely enough.



**Figure 5.** Fractal dimensions estimated for the distance radii **(a)** 50, 100, and 200 m, and **(b)** 200, 400, and 800 m.

#### 4.2 Typology of local built textures built on scaling indexes $Sc$

$Sc$  indexes are defined as the differences in the logarithms of two scaling functions calculated for two successive scale ranges, e.g.:

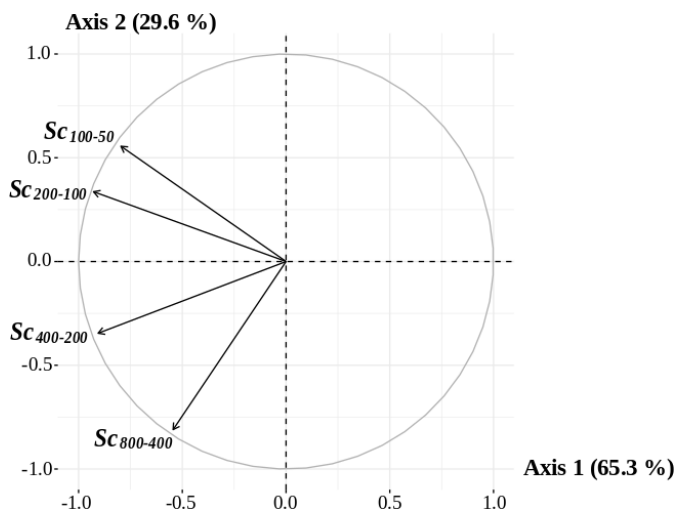
$$Sc_{100-50} = \log(M_{100}(i)) - \log(M_{50}(i)) \quad (5)$$

With five distance radii  $r$ , it is possible to calculate four scaling indexes  $Sc$ . In order to analyse them simultaneously, we applied a Principal Component Analysis. The two first components of the PCA concentrate 90% of the total inertia (**Figure 6**).

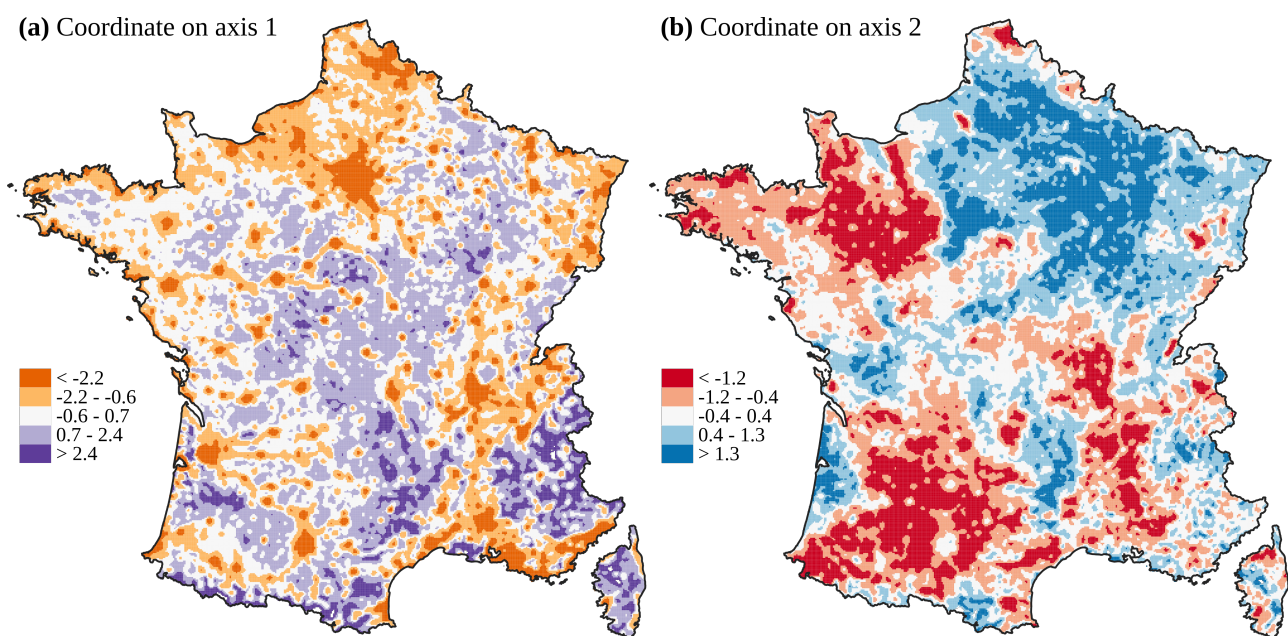
On the circle of correlations, we observe that this first principal component is strongly correlated with the four  $Sc$  indexes, all oriented in the same direction with respect to the first axis. Thus the first component contrasts areas characterized by a low fractal dimension and a low built density (*i. e.* rural areas) with areas characterized by a high fractal dimension and a high built density (*i. e.* urban areas) (**Figure 7a**).

The second component of the PCA reflects the fact that a built pattern may or may not exhibit a scale-invariant distribution across all scale ranges, from 0 to 800 m. It orders the scaling indexes in accordance with their distance radii: on the circle of correlations,  $Sc_{100-50}$  and  $Sc_{200-100}$  lie very close to each other and far from  $Sc_{800-400}$ ;  $Sc_{400-200}$  lies in an intermediate position. In **Figure 7b**, we see that the coordinate of urban areas is close to 0 for this second component. Indeed, their spatial organization is scale-invariant across the whole scale range (from 50 to 800 m) and, consequently, the four indexes  $Sc$  have almost the same values. Contrarily, the values of the scaling indexes vary with the distance radii for peripheral areas in the West of France, which appear in red in **Figure 7b**:

obviously, their fractal dimension is low in **Figure 5a** and high in **Figure 5b**. This stark contrast is explained by the dispersion of clusters of buildings, namely small villages and hamlets: for large distance radii (0–400 and 0–800 m), all points  $e$  have numerous neighbours and the index  $SC_{800-400}$  is high; conversely, for small distance radii, many points  $e$  have almost no neighbours and the indexes  $SC_{100-50}$  and  $SC_{200-100}$  are low. Built patterns in blue in **Figure 7b** exhibit a more complex scaling behaviour because they are locally more concentrated. For distance radii 0–50 m and 0–100 m, each point  $e$  has numerous neighbours located within the same built cluster and the index  $SC_{100-50}$  is high. Then, for distance radii 0–200 m and 0–400 m, there are relatively fewer neighbours of each point  $e$ ; they belong either to the built cluster of point  $e$  or to other built clusters (neighbouring villages) but the probability of finding other villages in the vicinity of the village of point  $e$  is low for such distances. The value of indexes  $SC_{200-100}$  and  $SC_{400-200}$  are accordingly lower. Finally, for large distance radii (0–400 and 0–800), the probability of finding neighbouring villages is much higher and the value of index  $SC_{800-400}$  is also higher.

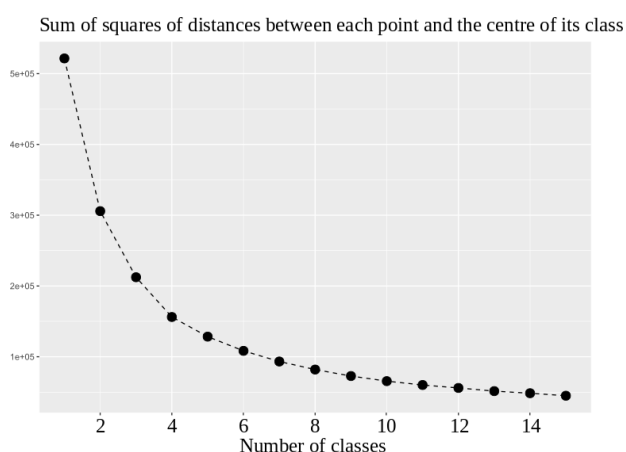


**Figure 6.** Principal Component Analysis: circle of correlations. Four variables:  $SC_{100-50}$ ,  $SC_{200-100}$ ,  $SC_{400-200}$ , and  $SC_{800-400}$ ; 145,178 individuals (2000 m-wide square spatial units).

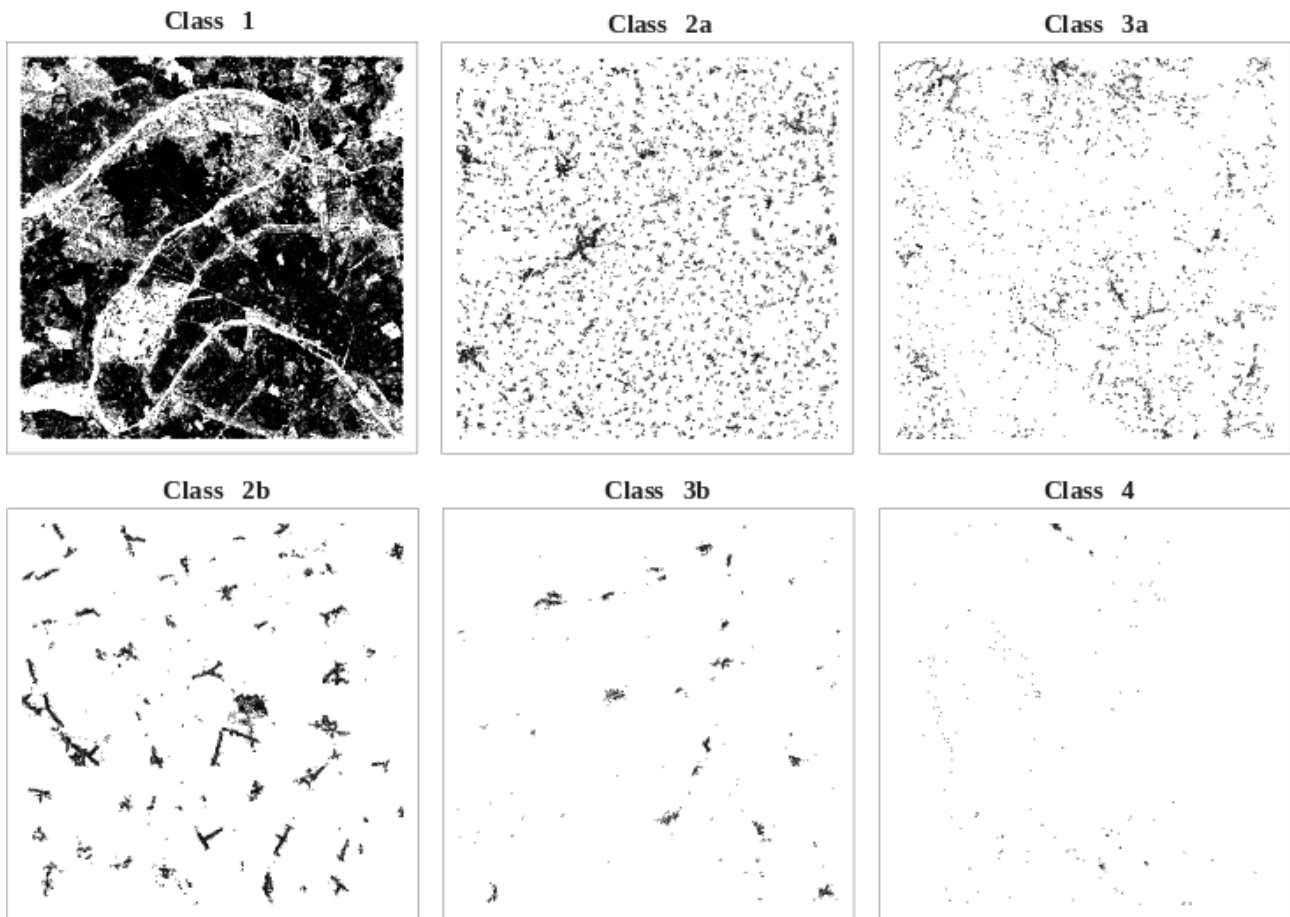


**Figure 7.** Coordinates of each 2000 m-wide zone  $i$  on the two first axes of the PCA.

Finally, we classified the 145,178 spatial units according to their two coordinates on the PCA axes using the k-means method. The scree-plot in **Figure 8** shows the increase in homogeneity within each class with respect to the increase in the number of classes. The point at which the slope of the curve levels off indicates the best compromise between the number of classes (that must be the smallest as possible) and the homogeneity within each class (that must be the highest as possible, i. e. the sum of the squares of the distances from each point to the centre of its class has to be the smallest as possible). Here it seems to be relevant to take on six classes. Classes 1 and 4, which occupy contrasting positions on the first axis of the PCA, are separated in the classification from the other classes (see the PCA scatter plot in **Figure 10**). Classes 2 (a and b) and classes 3 (a and b) are also separated since they occupy contrasting positions on the first axis of the PCA. Simultaneously, classes a (2 and 3) and classes b (2 and 3) are separated since they occupy contrasting positions on the second axis of the PCA.



**Figure 8.** Scree-plot of the k-means clustering.

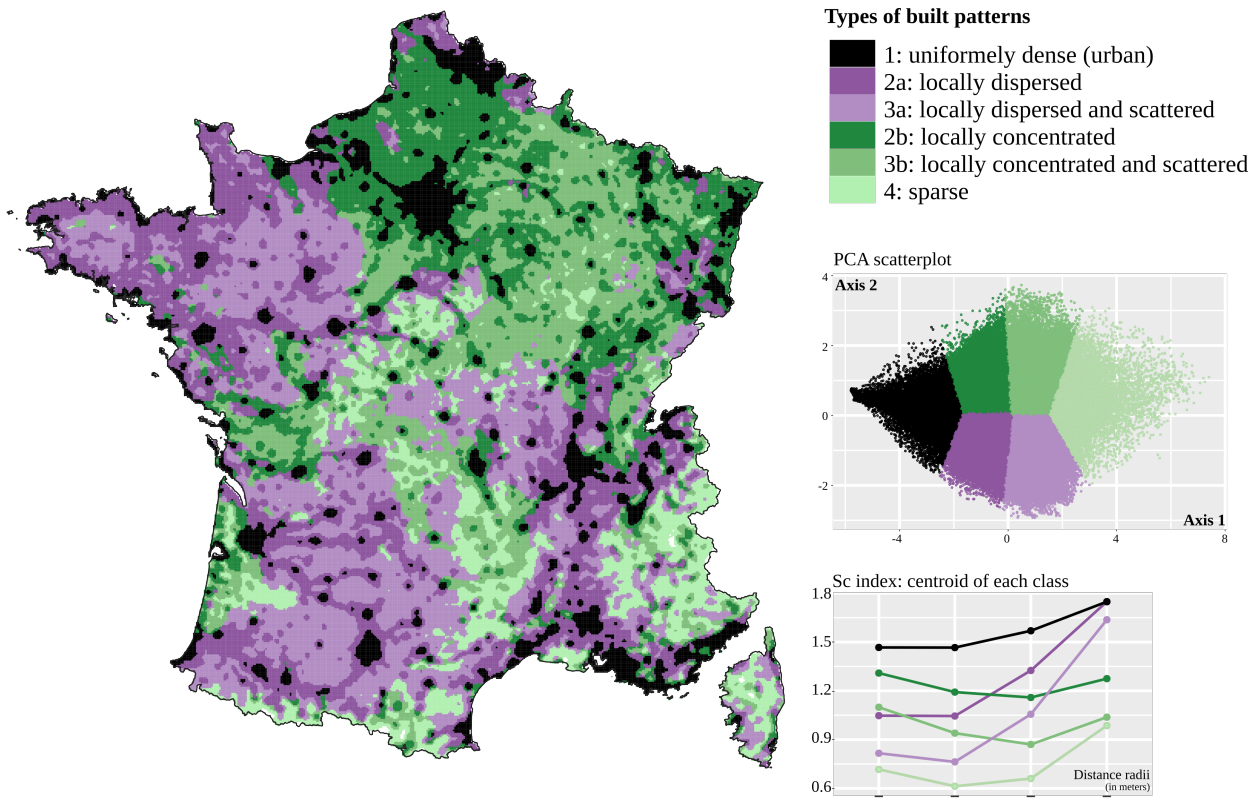


**Figure 9.** Built pattern typical of each class.

**Figure 9** shows that classes 2a and 3a correspond to locally dispersed built patterns. They are more scattered and diffuse in class 3a than in class 2a. Classes 2b and 3b correspond to locally concentrated built patterns; villages in class 2b are larger and more numerous than in class 3b. Class 1 corresponds to urban built patterns. Class 4 corresponds to sparse built patterns.

The resulting typology map is well structured although no contiguity criterion has been introduced into the classification (**Figure 10**). Clusters of spatial units belonging to a same class are very consistent: class 1 spatial units are urban areas; class 4 spatial units are sparsely populated areas; class 2a, 2b, 3a, and 3b spatial units are located in the periphery of urban areas.



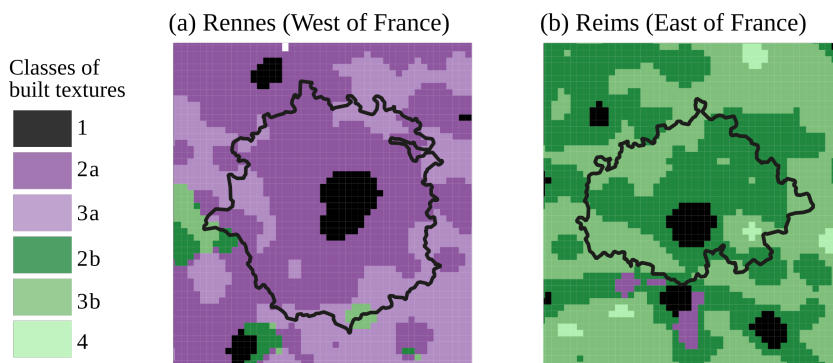


**Figure 10.** Typology of built patterns in mainland France. Colours have been chosen according to the recommendations of (Jenny & Kelso, 2007) using the software application ColorBrewer 2.0 (<http://colorbrewer2.org/>).

#### 4.3 For twenty middle-size metropolitan areas, identification of the distance threshold that characterises globally how the spatial distribution of buildings deviates from scale-invariance

For the purpose of the analysis, we used the delineation of French *aires urbaines* (metropolitan areas) (Figure 11), which encompass a densely built urban area and its commuter belt. We chose to consider neither large metropolitan areas because they encompass several *aires urbaines* that are sometimes spatially separated, nor small metropolitan areas as they do not contain enough 2000 m-wide spatial units. Thus twenty metropolitan areas have been analysed. Their number of inhabitants is comprised between 110 946 (Saint-Quentin) and 897 713 (Nantes) (source: French National Institute of Statistics and Economic Studies 2012).

Within the perimeter of each metropolitan area, a distance threshold that indicates a crucial discontinuity in space across scales was identified using the MorphoLim software application (Tannier et al., 2011). This distance threshold indicates the maximum morphological difference between the built patterns that belong to the urban area (i. e. the morphological agglomeration) and the outlying built patterns (i. e. the surrounding suburban areas). Table 1 shows that this distance threshold varies markedly among cities, from 78 m. for Avignon (South of France) to 529 m. for Troyes (East of France).



**Figure 11.** Two metropolitan areas and their built textures. Black line: limit of the metropolitan area (*aire urbaine*). Colours have been chosen according to the recommendations of (Jenny & Kelso, 2007) using the software application ColorBrewer 2.0 (<http://colorbrewer2.org/>).

**Table 1.** Morphological characteristics of 20 French metropolitan areas in 2012.

Name of the metropolitan area	Distance threshold (m)	Area (km <sup>2</sup> )	% of spatial units exhibiting locally dispersed built patterns (classes 2a and 3a)	% of spatial units exhibiting locally concentrated built patterns (classes 2b and 3b)
Avignon	78	2396	47	3
Angers	127	2744	87	2
Bayonne	143	1440	71	0
Le Mans	143	2364	89	0
Rennes	143	4236	90	2
Vannes	165	940	70	9
Nantes	171	3884	73	10
Caen	181	2168	51	26
Poitiers	207	2648	18	62
Clermont-Ferrand	242	2764	19	60
Niort	245	1652	20	69
Bourges	256	1912	9	79
Nancy	290	2832	2	82
Orléans	318	3572	33	48
Metz	389	1940	1	76
Amiens	395	2424	0	93
Dijon	415	3844	0	87
Saint-Quentin	420	952	0	88
Reims	434	2752	0	93
Troyes	529	2296	0	91

#### 4.4 Comparison of local built patterns with the global shape of metropolitan areas

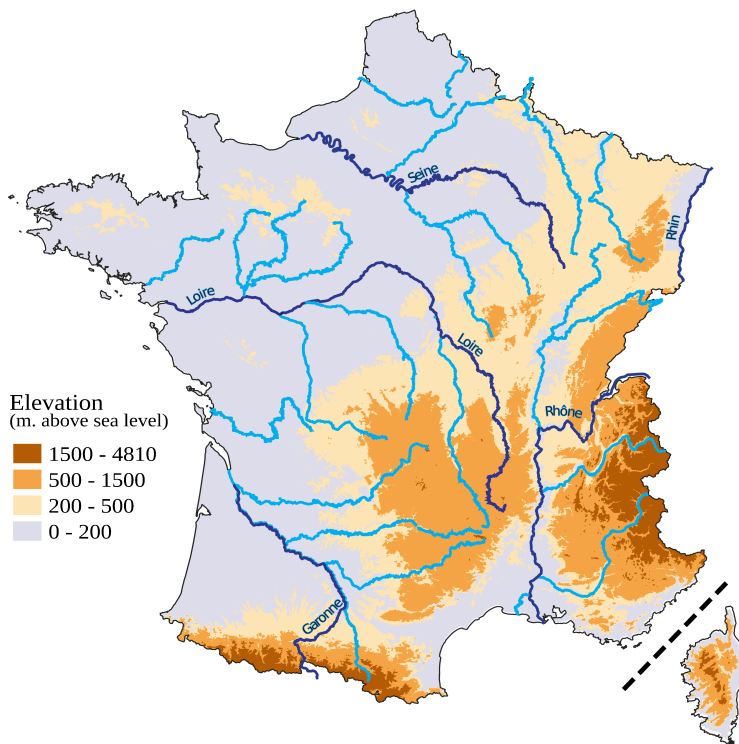
Tables 1 and 2 show that the proportion of dispersed (or alternatively concentrated) suburban built patterns in metropolitan areas is closely related to their global shape, i. e. their distance threshold. The more the suburban built patterns are dispersed into numerous, regularly spaced, small villages and hamlets, the smaller the distance threshold; the more the suburban built patterns are locally concentrated within large villages, the greater the distance threshold. This clearly confirms the hypothesis that the shape of French urban areas strongly depends on the built-up shapes in their periphery. Conversely, neither the proportion of spatial units in classes 1 (urban areas) and 4 (sparsely populated areas) nor the area of the metropolitan areas are correlated with the distance threshold.

**Table 2.** Statistical correlation between the distance threshold characterizing each urban area and other morphological characteristics. 20 metropolitan areas. p-values \*\*\* p<0.001 \*\* p<0.01 \* p<0.05.

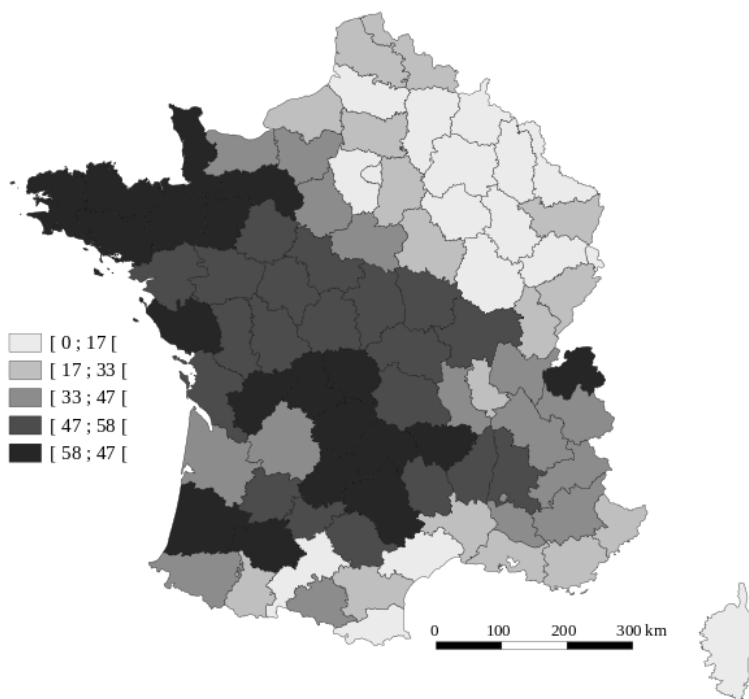
Variable	Pearson correlation coefficient
% of spatial units in classes 2a and 3a (dispersed built pattern)	-0.82***
% of spatial units in class 2a	-0.82***
% of spatial units in class 3a	-0.69***
% of spatial units in classes 2b and 3b (concentrated built pattern)	0.89***
% of spatial units in class 2b	0.81***
% of spatial units in class 3b	0.62**
% of spatial units in class 1	-0.52*
% of spatial units in class 4	0.07
Area of the metropolitan areas	-0.04

## 5. Discussion

On the typology map (**Figure 10**), we can recognize several geographical features of mainland France that can explain the local shape of built patterns. First, the delineations of the main urban areas (see **Figure 4b**) appear clearly on the typology map as well as the mountains and the valleys (see **Figure 12**). Second, the typology map in **Figure 10** exhibits numerous similarities with the map in **Figure 13**. Indeed the typology map shows the classical division of mainland France into enclosed field systems associated with scattered settlements in the centre and the West of France and openfield systems associated with grouped settlements in the East of France. Early twentieth-century historians (Bloch (1931), Dion (1934), Demangeon (1927)) tried to explain local differences in rural settlement patterns in France by adding other explanatory variables to the agrarian system. They showed that mountain and forest landscapes are associated with specific settlement patterns whereas other natural factors such as access to water or the shape of hydrographic networks are not. Analysis results presented in **Figure 10** confirm this. Valleys are either in class 2a (*locally dispersed built patterns*) or in class 2b (*locally concentrated built patterns*). Contrarily, high mountains are characterized by a mix of classes 3b (*moderately concentrated built patterns*) and 4 (*sparse built patterns*) everywhere in France.



**Figure 12.** Main mountains and rivers in mainland France. Sources: BD ALTI ® IGN and BD TOPO ® IGN. Colours have been chosen according to the recommendations of (Jenny & Kelso, 2007) using the software application ColorBrewer 2.0 (<http://colorbrewer2.org/>).



**Figure 13.** Share of sparse population (in %) in the total population of each French sub-regions (*départements*) in 1876. Source: French National Institute of Statistics and Economic Studies (SGF data). Note: the sparse population of each local community is the population located outside the main population cluster of the community, which includes in general the town hall, church, and marketplace.

Finally, the typology map in **Figure 10** enlightens us at a fine spatial resolution about differences in the spatial distribution of buildings that result from shaping factors acting at local or regional scales and being both social and natural.

The classification of scaling indexes obtained with GWFA has enabled the identification of six types of built patterns whose multiscale spatial organization is well differentiated. Sémécurbe, Tannier & Roux (2016) have previously identified similar types of settlement patterns by applying a multifractal analysis on fine-grained population data. That previous study was coarser-grained as mainland France was divided into 992 square spatial units 25 km wide (the present study considered 145,178 estimation points  $i$  2 km apart). The method was also different: the typology (a Hierarchical Ascending Classification) was based on the multifractal spectra of each spatial unit. Interestingly, despite the differences in methods and spatial resolutions, those two studies detected the same phenomena. In particular, the map displaying the value of each point  $i$  on the first axis of the PCA in the present study, which contrasts areas characterized by a low fractal dimension and a low built density (*i. e.* rural areas) and areas characterized by a high fractal dimension and a high built density (*i. e.* urban areas) (**Figure 7a**), resembles the map of generalized fractal dimension of order  $q = 2$  that highlights the strongest spatial singularities (Figure 3d in (Sémécurbe, Tannier & Roux 2016)). It should be recalled here that GWFA calculates a generalized fractal dimension of order  $q = 0$ . Conversely, the map displaying the value of each estimation point  $i$  on the second axis of the PCA in the present study (**Figure 7b** that highlights the deviation with respect to scale-invariance) resembles the map of generalized fractal dimension of order  $q = 0$  for which only the presence or absence of population in each cell is taken into account (Figure 3b in (Sémécurbe, Tannier & Roux, 2016)). Thus, the two studies look at the same phenomena, namely the concentration and dispersion of human settlements at several scales, from two different points of view, namely the departure from scale-invariance in the present paper and the more or less multifractal aspect of the spatial distribution in (Sémécurbe, Tannier & Roux, 2016).

## 6. Conclusion

In this paper, we have proposed a new method for spatial analysis, namely Geographically Weighted Fractal Analysis. We have used it to explore the local deviations from scale-invariance of spatial distributions of buildings in mainland France. On this basis, without making any a priori distinction between urban patterns and suburban or rural patterns and without imposing any contiguity criterion, we have identified six geographically consistent built patterns that differ in the way buildings are concentrated and dispersed across scales. Then, for twenty middle-size metropolitan areas, we have compared the local built textures with the distance threshold that characterises globally the spatial distribution of buildings. This comparison has shown that the global shape of cities is closely related to their surrounding built patterns: the urban sprawl process has affected places differently according to their pre-existing rural built patterns (typically, numerous small dispersed hamlets or fewer large villages), which correspond to centuries-old agrarian systems (typically, enclosed field systems and openfield systems). Finally, the results

obtained from this original four-steps methodological process are more informative than just calculating a local fractal dimension (the comparison of **Figure 3** with **Figure 10** illustrates this very clearly).

One point of interest in the present study is that the relatively small size of elementary spatial units (2000 m-wide) has allowed us to compare and contrast the local built textures with the global shape of cities (i. e. their distance threshold). Because of the use of both a k-means classification and the MorphoLim method, the results obtained may be specific to mainland France: on the one hand, the result of the statistical classification of built textures depends on the diversity of built textures within the study area (in mainland France, the built textures are highly contrasted); on the other hand, the distance threshold characterizing each urban area depends on the definition of metropolitan areas, which may vary among countries. Thus it would be interesting to apply the same methodology as proposed in this paper to analyse built patterns in other countries, with a view to making international comparisons.

The fact that outer suburban patterns are inherited (in part at least) from past settlement patterns is well recognized: like many complex systems, settlement systems are generally characterized by strong path dependency (Andersson, 2008). Nevertheless, in some cases, the urban sprawl process has not preserved features of old agrarian systems but has made a more or less clean sweep of pre-existing rural built patterns. For instance in the Champagne-Ardennes region of France, when land was controlled by cereal farmers and winegrowers, residential developments occurred mainly as extensions to the larger villages. Conversely, housing estates with stereotyped architecture were constructed where successive land reparation made sites available or where farmers were behind the projects so as to cash in on their landholdings or prepare for their retirement (Mancebo & Salles, 2014). Such spatial variation in the intensity of path dependency could be explored quantitatively by applying the method and tools proposed in this paper.

#### **Acknowledgment**

The authors would like to thank Armelle Couillet, cartographer at the research laboratory IDEES (Rouen, France), for her help in making maps readable by people with red-green colour blindness.

#### **References**

- Andersson, C. (2008). Ontogeny and ontology in complex systems modeling. In S. Albeverio, D. Andrey, P. Giordano & A. Vancheri, eds, *The Dynamics of Complex Urban Systems: An Interdisciplinary Approach*, Physica-Verlag, 43–58.
- Batty, M. & Longley, P. (1986). The Fractal Simulation of Urban Structure. *Environment and Planning A*, 18, 1143–79.
- Batty, M. & Longley, P. A. (1987). Urban shapes as fractals. *Area*, 19(3), 215-221.
- Batty, M. & Xie, Y. (1996). Preliminary Evidence for a Theory of the Fractal City. *Environment and Planning A*, 28, 1745–62.
- Benguigui, L., Czamanski, D., Marinov, M. & Portugali, Y. (2000). When and Where Is a City Fractal? *Environment and Planning B: Planning and Design*, 27, 507-519.

- Benguigui, L. & Czamanski, D. (2004). Simulation Analysis of the Fractality of Cities. *Geographical Analysis*, 36(1), 69–84.
- Bloch, M. (1931). *Les caractères originaux de l'histoire rurale française*. Oslo: Instituttet for sammenlignende kulturforskning.
- Brunsdon, C., Fotheringham, A. S. & Charlton, M. (1996). Geographically Weighted Regression: A method for exploring spatial nonstationarity. *Geographical Analysis*, 28(4), 281-298.
- Chaudhry, O., Mackaness, W. (2008). Automatic Identification of Urban Settlement Boundaries for Multiple Representation Databases. *Computers, Environment and Urban Systems*, 32, 95–109.
- Chen, Y., Wang, J. (2013). Multifractal Characterization of Urban Form and Growth: The Case of Beijing. *Environment and Planning B: Planning and Design*, 40(5), 884-904.
- De Keersmaecker, M.-L., Thomas, I., Frankhauser, P. (2003). Using Fractal Dimensions for Characterizing Intra-urban Diversity: The Example of Brussels. *Geographical Analysis*, 35(4), 310-328.
- Demangeon, A. (1927). La géographie de l'habitat rural. *Annales de Géographie*, 36, 1–23.
- Diggle, P. (1985). A Kernel Method for Smoothing Point Process Data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 34(2), 138-147.
- Dion, R. (1934). *Essai sur la formation du paysage rural français*. Tours, France: Arrault.
- Feng, J. & Chen, Y. (2010). Spatiotemporal evolution of urban form and land-use structure in Hangzhou, China: Evidence from fractals. *Environment and Planning B: Planning and Design*, 37(5), 838–856.
- Fotheringham, A. S., Brunsdon, C. & Charlton, M. (2003). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, John Wiley & Sons.
- Frankhauser, P. (1988). Fractal Aspects of Urban Systems. *Sonderforschungsbereich 230 "Natürliche Konstruktionen" 1*, 67–76.
- Frankhauser, P. (1998). The fractal approach: a new tool for the spatial analysis of urban agglomerations. *Population*, 4, pp. 205-240.
- Grassberger, P. & Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica D*, 9, 189–208.
- Goodchild, M. (1980). Fractals and the Accuracy of Geographical Measures. *Mathematical Geology*, 12(2), 85-98.
- Jenny, B. & Kelso, N. V. (2007). Designing Maps for the Colour-Vision Impaired. *Bulletin of the Society of Cartographers*, 41, 9-12.
- Lam, N. (1990). Description and Measurement of Landsat TM Images Using Fractals. *Photogrammetric Engineering and Remote Sensing* 56(2), 187–95.
- Lam, N. & Quattrochi, D. (1992). On the Issues of Scale, Resolution, and Fractal Analysis in the Mapping Sciences. *Professional Geographer*, 44(1), 88–98.
- Longley, P. & Batty, M. (1989). On the Fractal Measurement of Geographical Boundaries. *Geographical Analysis*, 21(1), 47–67.
- Le Gléau, J., Pumain, D. & Saint-Julien, T. (1997). Towns of Europe: to each country its definition. *INSEE Studies*, 6, 1–14.
- Mancebo, F. & Salles, S. (2014). De l'autre côté du miroir. Un périurbain pensé par le rural, pour une périurbanisation modelée par les usages. *Premier plan Le journal d'informations du PUCA (plan urbanisme construction architecture)*, 30, 4–6.
- Mandelbrot, B. (1982). *The Fractal Geometry of Nature*. San Francisco: Freeman.
- Sémécurbe, F., Tannier, C. & Roux, S. G. (2016). Spatial distribution of human population in France: exploring the Modifiable Areal Unit Problem using multifractal analysis. *Geographical Analysis*, 48(3), 292–313.

- Shen, G. (2002). Fractal dimension and fractal growth of urbanized areas. *International Journal of Geographical Information Science*, 16(5), 437–519.
- Silverman, B. W. (1978). Choosing the window width when estimating a density. *Biometrika*, 65, 1-11.
- Tannier C. & Pumain D. (2005). Fractals in Urban Geography: A Theoretical Outline and an Empirical Example. *Cybergeo: European Journal of Geography*, 307 [Online].
- Tannier, C., Thomas, I., Vuidel, G. & Frankhauser P. (2011). A fractal approach to identifying urban boundaries. *Geographical Analysis*, 43(2), 211–227.
- Tannier, C. & Thomas, I. (2013). Defining and characterizing urban boundaries: A fractal analysis of theoretical cities and Belgian cities. *Computers, Environment and Urban Systems*, 41, 234-248.
- Tél, T., Fülöp, Á. & Vicsek, T. (1989). Determination of fractal dimensions for geometrical multifractals. *Physica A: Statistical Mechanics and its Applications*, 159(2), 155–166.
- Thomas, I., Frankhauser, P. & Biernacki, C. (2008). The morphology of built-up landscapes in Wallonia, Belgium: a classification using fractal indices. *Landscape and Urban Planning*, 84, 99–115
- Thomas, I., Frankhauser, P., Frenay, B. & Verleysen, M. (2010). Clustering patterns of urban built-up areas with curves of fractal scaling behaviour, *Environment and Planning B: Planning and Design*, 37, 942–954.
- Vicsek, T. (1990). Mass multifractals. *Physica A: Statistical Mechanics and its Applications*, 168(1), 490–497.
- White, R. & Engelen, G. (1993). Cellular Automata and Fractal Urban Form: A Cellular Modelling Approach to the Evolution of Urban Land-Use Patterns. *Environment and Planning A*, 25(8), 1175–99.
- White, R. & Engelen, G. (1994). Urban Systems Dynamics and Cellular Automata: Fractal Structures between Order and Chaos. *Chaos, Solitons and Fractals*, 4(4), pp. 563-583.



## Appendix A. Computing GWFA using R

```
library(sf) # to use shapefiles
library(dbSCAN) # to identify the neighbours of each point e
library(btb) # to smooth the data with a quartic kernel (with the function msf using a bandwidth of size 2000 m. to 8000 m.)
library(dplyr) # to manage data.frame
library(cartography) # to create maps

radius = c(50, 100, 200, 400, 800) # Distance radii used to perform GWFA

## Transform a shapefile that represents buildings in two dimensions (polygons) into a table that contains building centroids e with their xy coordinates. ##

buildings = read_sf("filename.shp")

# Transform the initial cartographic projection into Lambert 93
buildings = st_transform(buildings, 2154)

# Calculate the coordinates of the centroid of each building
points = st_coordinates(st_centroid(buildings))
points = data.frame(points)
names(points) = c("x", "y")

## Count the number of neighbours of each point (i.e. building centroid) e within each radius (50, 100, 200, 400, 800). ##

for (r in radius)
{
  # Identify the neighbours of each point e
  nn = frNN(points[,1:2], r)

  # Count the number of neighbours of each point e
  points[, paste0("M_", r)] = sapply(adjacencylist(nn),length) + 1
}

## Apply GWFA to the building centroids e for five distance radii (50, 100, 200, 400, and 800 m). ##

# Calculate the indexes  $M_r(i)$  for each radius r and each estimation point i (cf. eq 4)
neighbors = points %>% mutate_at(paste0("M_", radius), function(x){x^-1})
msf = kernelSmoothing(neighbors, 2154, 2000, 8000)
msf = msf %>% mutate_at(paste0("M_", radius), function(x){x^-1})

# Compute the fractal dimension  $D_0$ 
logradius = log2(radius)
X = (logradius - mean(logradius)) / var(logradius) / (length(logradius) - 1)
M_r = msf %>% select(paste0("M_", radius)) %>% st_set_geometry(NULL)
msf$fd_50_800 = as.matrix(log2(M_r)) %*% matrix(X, length(logradius), 1)

## Map the fractal dimension  $D_0$  of each estimation point i ##

choroLayer(msf, var = "fd_50_800")
```

## Appendix B. Computing GWFA using both Python and R

---

### 1. Computing code in Python

---

```
import pandas as pd
import geopandas as gpd
from shapely.geometry import Polygon
from sklearn.neighbors import KDTree
import numpy as np
```

**## Transform a shapefile that represents buildings in two dimensions (polygons) into a table that contains building centroids  $e$  with their XY coordinates ##**

```
building=gpd.read_file("gis.osm_buildings_a_free_1.shp")
```

*# Transform the initial cartographic projection into Lambert 93*

```
building=building.to_crs({'init': 'epsg:2154'})
```

```
centroid = building['geometry'].centroid
```

```
points=np.column_stack([centroid.x,centroid.y])
```

**## Count the number of neighbours of each point  $e$  within each radius  $r$  (50, 100, 200, 400, 800) ##**

```
radius=np.array([50,100,200,400,800])
```

```
neighbors=np.zeros((points.shape[0],radius.size),dtype=np.float)
tree = KDTree(points)
```

```
for i in range(radius.size):
    neighbors[:,i]=tree.query_radius(points,r= radius[i],count_only=True,return_distance=False)
```

```
output=np.column_stack([points,neighbors])
names=np.hstack(['x','y', np.core.defchararray.add('r_',np.char.mod('%d', radius))])
```

```
output=pd.DataFrame(output,columns=names)
output.to_csv("neighbors.csv", float_format='%0.6f')
```

---

### 2. Computing code in R

---

**## Apply GWFA to the building centroids  $e$  for five distance radii  $r$  (50, 100, 200, 400, and 800 m) ##**

```
library(btb)
library(sf)
```

```
neighbors=read.csv("neighbors.csv")
neighbors=neighbors[,-1]
```

```
carreauxSF <- function(df, iCellSize, sEPSG)
```

```
{
```

```
  r <- iCellSize / 2
```

```
  df$geom <- sprintf("POLYGON ((%i %i, %i %i, %i %i, %i %i, %i %i))", df$x-r, df$y+r, df$x+r, df$y+r,
```

```
df$x+r, df$y-r, df$x-r, df$y-r, df$x-r, df$y+r)
```

```
  sdf <- st_as_sf(df, wkt = "geom", crs = as.integer(sEPSG))
```

```
  return(sdf)
```

```
}
```

```

neighbors[,c(-1,-2)]=neighbors[,c(-1,-2)]^-1

liss=kernelSmoothing(dfObservations = neighbors,cellSize = 2000,bandwidth = 8000)
msf=data.frame(liss@.Data)
names(msf)=c("x", "y", "M_50","M_100", "M_200", "M_400","M_800")

msf[,c(-1,-2)]=log2(msf[,c(-1,-2)])
msf$df_20_5=-((msf$M_200-msf$M_50)/2)
msf$df_80_20=-((msf$M_800-msf$M_200)/2)
scale=c(1,2,3,4,5)
x=-((scale-mean(scale))/(var(scale)*4))
msf$df80_5=(as.matrix(msf[,3:7])%*%matrix(x,5,1))

output_8000=carreauxSF(msf,2000,2154)

st_write(output_8000,"analyse_8000.shp")

```