A Fractal Approach to Identifying Urban Boundaries

Cécile Tannier\textsuperscript{1}, Isabelle Thomas\textsuperscript{2}, Gilles Vuidel\textsuperscript{1}, Pierre Frankhauser\textsuperscript{1}

\textsuperscript{1}ThéMA, CNRS—University of Franche-Comté, Besançon, France, \textsuperscript{2}Fonds de la recherche scientifique (FRS-FNRS), Center for Operations Research and Econometrics (CORE), Department of Geography, Catholic University of Louvain-la-Neuve, Belgium

Fractal geometry can be used for determining the morphological boundaries of metropolitan areas. A two-step method is proposed here: (1) Minkowski’s dilation is applied to detect any multiscale spatial discontinuity and (2) a distance threshold is located on the dilation curve corresponding to a major change in its behavior. We therefore measure the maximum curvature of the dilation curve. The method is tested on theoretical urban patterns and on several European cities to identify their morphological boundaries and to track boundary changes over space and time. Results obtained show that cities characterized by comparable global densities may exhibit different distance thresholds. The less the distances separating buildings differ between an urban agglomeration and its surrounding built landscape, the greater the distance threshold. The fewer the buildings that are connected across scales, the greater the distance threshold.

Introduction

Defining city boundaries is invariably an enthralling challenge for urban geographers and planners. Where do cities start and end? Is there a continuum of rural and urban land uses or a sharp divide between them? While some urban analyses (Dujardin, Thomas, and Tulkens 2007; Parr 2007) also include functional (e.g., socioeconomic, attractiveness) and operational (e.g., administrative, political) considerations, the focus here is on morphological (physical) criteria alone.

A general consensus does not appear to exist about the definition of the urban morphological agglomeration, or what Parr (2007) calls the “built city.” Working definitions often result from classifications or typologies of built patterns, and are based on applications of satellite and geographic information system technologies.
(e.g., Treitz, Howarth, and Gong 1992; Donnay, Barnsley, and Longley 2001; Abed and Kaysi 2003). Besides classifications, predefined morphological thresholds are commonly used (e.g., Donnay, Barnsley, and Longley 2001; Weber 2001). After classifying elementary spatial units, so that similar ones can be grouped together, the criterion of contiguity—generally involving a distance threshold—is added. However, there is no consensus about the choice of contiguity criteria (Le Gléau, Pumain, and Saint-Julien 1997; Dujardin, Thomas, and Tulkens 2007), and the relevance of a distance threshold becomes questionable when analyzing urban fringes. European cities, in particular, exhibit irregular patterns where recent detached housing estates and traditional rural buildings are mixed. Consequently, the spacing of neighboring buildings within sprawling patterns varies considerably, making application of a single distance threshold very difficult (Chaudhry and Mackaness 2008).

We suggest a method for identifying the morphological boundary of urban agglomerations that avoids the need for any predefined maximum distance threshold between buildings. Given that crucial spatial discontinuities can be identified by analyzing phenomena over a range of scales (Lam and Quattrochi 1992), we propose to use the conceptual and methodological tools of fractal geometry for defining the morphological boundaries of urban agglomerations.

Fractal geometry has been widely applied in geography for some 30 years now. Pioneering works argue that local and global spatial structures are interrelated at several scales (Goodchild 1980; Arlinghaus 1985, 1987; Batty and Longley 1986; Goodchild and Mark 1987; Frankhauser 1988). A string of publications specifically about urban geography shows that urban growth engenders a multiscale spatial organization (Fotheringham, Batty, and Longley 1989; Batty and Xie 1996, 1999; Benguigui et al. 2000; Shen 2002; Thomas, Frankhauser, and Biernacki 2008). Other contributions explore the fractal aspects of area–perimeter relations and the fractal boundary of built forms (Longley and Batty 1989; Arlinghaus and Nystuen 1990; Frankhauser 1994). Yet, others discuss the scaling relationship between the number of built clusters and their size (White and Engelen 1993; De Keersmaecker, Frankhauser, and Thomas 2003; Benguigui and Czamanski 2004; Benguigui, Blumenfeld-Lieberthal, and Czamanski 2006).

The main characteristic shared by fractals is their intrinsic multiscale spatial organization. They may, however, exhibit different kinds of multiscale behavior. For example, the elements of the familiar Sierpinski carpet are interconnected and actually form a single cluster (Mandelbrot 1982). By contrast, elements in a Fournier dust are unconnected and form clusters separated by empty lanes of different sizes: a Fournier dust contains just a few wide lanes and an increasing number of ever narrower ones. The lane widths (i.e., the distances separating the clusters) comply with a strict hierarchy (Mandelbrot 1982).

The hierarchy of empty lanes characterizing Fournier dusts can be explored by the dilation method for determining fractal dimensions proposed by Minkowski (1903) and further developed by Bouligand (1929). In dilation, each built element
of a study area is surrounded by a buffer zone that widens at each step of the pro-
cess. As dilation progresses, what were initially separate buildings become merged,
and clusters form. As these clusters expand, their number decreases. Dilation stops
when just one big cluster remains. In fractals like Fournier dusts, the number of
clusters \( N \) is related to the size of the dilation buffer \( \varepsilon \) according to a power-law
function (equation [1]).

\[
N = a\varepsilon^D,
\]

where \( a \) is a constant of proportionality. The scaling exponent of this law is the
fractal dimension \( D \). If we take the logarithm of both sides, this expression has the
form of a linear relationship with slope \( D \) (equation [2]).

\[
\log(N) = D \log(\varepsilon) + \log(a).
\]

In geography, dilation is used for analyzing and segmenting various types of
images (e.g., remotely sensed images, aerial photographs). Some researchers (Pes-
aresi and Benediktsson 2000; Benediktsson, Pesaresi, and Arman 2003) apply
mathematical morphology (that uses dilation and erosion processes) to detect the
boundaries of urban objects in remotely sensed images. With their method, local
discontinuities emerge from an iterative transformation of the pattern itself, and
predetermined thresholds are eschewed. However, few applications concern dila-
tion for delimiting built agglomerations, and those applications that exist introduce
predefined distance thresholds to identify morphological urban boundaries (e.g.,

Here we use dilation for identifying the morphological boundaries of urban
agglomerations without introducing predefined distance thresholds. We begin with
the log–log plot that counts the number of built clusters at each dilation step. If the
built area of a city is fractal, then the number of clusters is related to their size by a
power-law function corresponding to a linear relation visible on a log–log plot. We
posit that the point at which the dilation curve deviates most from a straight line is a
crucial distance threshold, corresponding to the morphological multiscale bound-
ary of an urban agglomeration. This boundary, referred to here as the \textit{urban enve-
lopes}, is a spatial discontinuity across scales (Frankhauser and Tannier 2005). It
separates two morphological spatial subsets that are distinct in fractal terms and
 corresponds to a break point in the multiscale organization of a built pattern. Be-
yond this dilation threshold, the pattern is no longer self-similar, or its self-similarity
changes dramatically.

Below a dilation threshold, the self-similarity of a pattern may be strictly fractal.
However, the fractal dimension \( D \) is not expected to be constant in reality (Good-
child 1980); most often, it is constant over a limited range of scales but varies
somewhat over successive ranges of scales (Lam 1990; Lam and Quattrochi 1992;
White and Engelen 1994; Tannier and Pumain 2005). Hence, the fractality of a
pattern below its dilation threshold is most probably multifractal.\(^1\)
Our aim is to identify an urban boundary, yet the possible existence of a rural–urban continuum cannot be ignored (Champion and Hugo 2004). Consequently, we do not assume that an urban envelope always exists—it exists only if two spatial organizations can be distinguished from a multiscale perspective. If this is the case, then a distance threshold can be identified; otherwise, in the case of a rural–urban continuum, detecting any discontinuity between the urban agglomeration and the outlying area is impossible.

**Identifying an urban envelope**

The data used are vector maps representing buildings in two dimensions (polygons) (Fig. 1). The spatial extent of a study area is necessarily large because it comprises an (either monopolar or multipolar) urban agglomeration and its hinterland (i.e., a suburban or rural area that is under the influence of its urban core).

A two-step method is adopted for identifying an urban envelope: (1) Minkowski’s (1903) dilation is applied to detect the existence of any multiscale spatial discontinuity and (2) a distance threshold is located on the dilation curve corresponding to a major shift in its behavior. We therefore measure the maximum curvature of the dilation curve (Lowe 1989).

**Minkowski dilation**

A vector version of the Minkowski dilation is applied to each built polygon of a map. This consists in surrounding each built polygon with a buffer of increasing width. The buffer width increases according to a geometric logic, which corresponds to the fractal logic. The number of clusters is counted after each dilation step. The results are portrayed as a log–log plot, where the X-axis represents the

![Figure 1](image.png)

**Figure 1.** Sample of data used (excerpt from the 1957 map of Basel; see Fig. 7).
width of the dilation buffer and the Y-axis represents the corresponding number of built clusters (Fig. 2a).

Identification of a significant threshold on a dilation curve
A significant threshold on a dilation curve is defined as the point characterized by the main value of curvature. The curvature function measures how a curve deviates to a greater or lesser extent from being straight at a given point (Lowe 1989). For each point on a curve, this value may be calculated as

$$\kappa = \left(\frac{y''}{(1 + y'^2)^{3/2}}\right),$$

(3)

where $y'$ (first derivative) measures the speed at which the number of built clusters decreases in the course of dilation and $y''$ (second derivative) measures the acceleration of this decrease. The formula defines the curvature as a ratio of speed to acceleration. For a straight line, the speed is constant and the acceleration is zero. Hence, the main curvature point corresponds to the maximum deviation of a curve.

Figure 2. Identification of a significant threshold on the dilation curve (the metropolitan area of Besançon, France, for example).
with regard to a linear decrease in the number of built clusters in the course of
dilation.

To identify the point of main curvature, a dilation curve is approximated by a
series of polynomials of increasing degrees. We then select the polynomial of the
lowest degree that best represents the original curve (Fig. 2b). This choice is based
on Schwarz’s (1978) Bayesian information criterion (BIC) for evaluating the trade-off
between goodness of fit and complexity for statistical models. The BIC is stricter
than Akaike’s (1974) information criterion (AIC) in penalizing the loss of degrees of
freedom by having more parameters in a fitted model. The polynomial degree
chosen here corresponds to the point at which the BIC bottoms out or begins to
increase. In any case, the adjusted correlation coefficient ($R^2$) between the real and
the estimated curves must be $> 0.9$.

Once the estimated curve is obtained, points of maximum curvature are iden-
tified (Fig. 2c): they are characterized by a zero derivation value of the curvature
function. To avoid points of maximum curvature arising from estimation artifacts,
the two extremities of an estimated curve are removed before calculating the cur-
vature function (Fig. 2d). This adjustment implies that the distance threshold de-
termining an urban envelope is neither a few meters (too small) nor several
kilometers (too large). The point of main curvature has the highest absolute value
of curvature among the points of maximum curvature. Finally, the point of main
curvature is located on the estimated curve, which gives the distance threshold for
drawing an urban envelope (Fig. 2d).

The step-by-step vector dilation and cluster count were performed for a com-
puter application developed with the Java Topology Suite and Geotools library.
The polynomial estimations of the dilation curve, the BIC, and the curvature func-
tion were computed using Octave, an open source version of Matlab. The
two applications are integrated in the software MorphoLim (http://www.spatial-
modelling.info/MorphoLim-Identifying-city).

Application to theoretical urban patterns

Three theoretical patterns are studied here. Figure 3a portrays a regular fractal pattern
constructed in accordance with a strict iterative logic and characterized by strict self-
similarity. This hybrid Sierpinski carpet combines a Sierpinski carpet consisting of a
single cluster and a Fournier dust of unconnected elements at each step of iteration
(Frankhauser and Tannier 2005). As with Fournier dusts, the hierarchy of distances
between built clusters in the hybrid Sierpinski carpet obeys a power-law function. The
hybrid carpet of Fig. 3a comprises 28,500 identical square buildings.

Figure 3b portrays a random fractal pattern characterized by quasi-self-similarity.
As with the hybrid carpet, its elements are partly connected and partly unconnected.
A single scaling law defines the fractal organization of the pattern, but the respective
locations of the elements vary randomly. As with a real city, elements are not all of
the same size.
Figure 3. The envelopes of the three theoretical cities.
Figure 3c portrays a regular fractal pattern in a nonfractal environment. It is the same hybrid carpet as in Fig. 3a, but this time it is surrounded by randomly and fairly uniformly distributed elements. These elements are like noise around the fractal pattern. Figure 3c can be seen as a metaphor of a city surrounded by rural settlements.

The multiscale envelope of each pattern has been identified by the method proposed previously. For the regular fractal city, the dilation curve exhibits some plateaus but is essentially linear: any shift along the $X$-axis has a corresponding similar shift along the $Y$-axis (Fig. 3a). This synchronized shifting is because a fractal relationship associates a geometric (hence multiplicative) change in the number of elements with a geometric change in their size (or in the distance between them). A log–log plot changes a geometric progression into a linear progression. The slope of the curve is the fractal dimension $D$, which is constant for a simple regular fractal pattern regardless of the scale. Hence, the dilation curve is suitably estimated by a first-degree polynomial (i.e., a linear function), as shown in Fig. 4. In this case, the second derivative is always zero; consequently, the curvature function is zero, too. No point of maximum curvature can be identified, and thus no significant threshold exists in the dilation curve. Therefore, an envelope is impossible to identify for a subset of elements of the image: the envelope here bounds the entire image.

In the case of the random fractal city, the dilation curve is smoother. The randomness of the pattern means that no plateau appears (Fig. 3b). However, the curve is linear (see Fig. 4—highly adjusted $R^2$ for a first-degree polynomial). No main curvature can be identified and the envelope bounds the entire pattern. By contrast, for the hybrid carpet in a nonfractal environment, the dilation curve is no longer linear (Figs. 3c and 4). A clear threshold appears. This is an important feature arising from the spatial organization of the pattern under study: it reveals a shift in the behavior of the pattern in the course of dilation. For simple theoretical fractal patterns, spatial relations between elements remain the same regardless of the scale of analysis, whereas for more complex patterns, like the hybrid carpet in a nonfractal environment, the spatial organization between two scales may vary. Such a variation can be interpreted as a fundamental morphological discontinuity indicative of

![Figure 4](image_url). Adjusted $R^2$ for the series of polynomial estimations of the dilation curve.
a sharp divide between two subsets of elements. Such a discontinuity corresponds to a change in the fractal dimension between two scales.

**Application to European cities**

**Six French and Belgian urban agglomerations**

The usefulness and operationality of the preceding method can be illustrated with some small- and medium-sized European urban agglomerations. Here we study three French cities (Besançon, Belfort, and Montbéliard) near Switzerland, and three Belgian cities (Namur, Liège, and Charleroi) in the former industrial belt in Wallonia. Their populations range from about 80,000 inhabitants for Belfort to 500,000 inhabitants for Liège.

The study areas considered for each city are the French and Belgian *aires urbaines* (metropolitan areas) (Luyten and Van Hecke 2007; INSEE 2009). In France, an *aire urbaine* encompasses a densely built *unité urbaine*—which is akin to a U.S. urban area—and its commuter belt. The Belgian definition of *aire urbaine* is similar to the French one. The spatial extent of the six agglomerations studied varies from 3.7 km² (Belfort) to 28.5 km² (Liège).

Data used for the French agglomerations come from the “BD Topo” vectorial database of the Institut Géographique National (IGN 2009). All buildings with footprints of more than about 1 m² are mapped regardless of their function (e.g., residential, commercial, industrial). For the Belgian agglomerations, data are from the Plan de Localisation Informatique (PLI) developed by the Ministère de la Région Wallonne, Direction générale opérationnelle—Aménagement du territoire, Logement, Patrimoine et Energie (2004). The PLI is based on digitized topographic maps at a scale of 1 : 10,000 (IGN) and the land register (*cadastre*). Polygons representing buildings are stored in a vector data layer.

The shapes of the dilation curves obtained for each metropolitan area are very similar (Fig. 5), but the curves for the Belgian metropolitan areas are a little straighter and have a lower main curvature value. This is understandable because Belgian metropolitan areas are very prone to urban sprawl (Caruso 2001; Champion 2001). Belgian suburban (sprawling) patterns, however, are mainly fractal and exhibit multiscale spatial contrasts (Thomas, Frankhauser, and Biernacki 2008). By comparison, French urban agglomerations typically are individual cities that stand clearly apart from their rural surroundings. Belgium is a very small and highly urbanized country (approximately 30,000 km² for 10 million inhabitants) compared to France (657,000 km² for 65 million inhabitants); the structures of the two countries’ urban systems are also different.

Besides the shape of the dilation curves, distance thresholds are clearly different, ranging from 465 m (Namur) to 170 m (Belfort). Study areas characterized by comparable global densities may exhibit different dilation thresholds. This tendency is consistent with the contention that a same density corresponds to various fractal dimensions. The information captured by fractal dimensions measured
For areas is fundamentally different from the information given by density (for mathematical explorations of the differences and relationship between density and fractal dimension, see Batty and Kim 1992; Batty and Xie 1996; Thomas, Frankhauser, and De Keersmaecker 2007). The less an urban agglomeration differs from its surrounding built landscape (in terms of variations of distances separating buildings), the greater the distance threshold. Moreover, the fewer the buildings in a study area that are connected across scales, the greater the distance threshold. The example of two fractal forms, the Fournier dust and the Sierpinski carpet (Mandelbrot 1982), is useful for illustrating this idea. The Fournier dust is made up of unconnected elements across scales, whereas the Sierpinski carpet is made up of connected elements across scales. European urban areas usually combine features of both Fournier dusts and Sierpinski carpets with regard to connectivity between buildings (Thomas, Frankhauser, and De Keersmaecker 2007). If an urban agglomeration is more like a Sierpinski carpet than a Fournier dust and at the same time clearly different from its surrounding built landscape, then it exhibits a low dilation threshold (e.g., Belfort and Liège). In contrast, an urban agglomeration more like a Fournier dust than a Sierpinski carpet, and whose spatial organization is similar to that of its surrounding built pattern, exhibits a high dilation threshold (e.g., Montbéliard).

**Diachronic analysis of the Basel urban agglomeration**

We worked here with three vector maps of the built pattern of the Basel agglomeration and its hinterland for 1882, 1957, and 1994. The spatial extent of the study area is approximately 530 km². It extends over three European countries: Switzerland, France, and Germany. The urban agglomeration population was about 170,000 inhabitants in 1999. The original data used for making the three maps are topographic maps from the Bundesamt für Landestopographie, 3, 084 Wabern, Germany. They are at different scales (from 1:25,000 to 1:50,000) and were drawn using different cartographic generalization methods (varying by date and...
country). Because of these differences, cartographic adjustments were required to develop a consistent set of vector-built maps. Each polygon displayed with one of the finalized vector maps provides a fairly accurate representation of a building or part of a building. Small buildings, such as huts or sheds, are not mapped.

Results for the urban envelope of Basel at three points in time are presented in Fig. 6. The dilation curve for Basel in 1882 has two different sections. The first is characterized by a rather slow decline in the number of built clusters from one dilation step to the next, whereas the second exhibits a faster rate of decline. This trend is reminiscent of the spatial organization of the regular fractal city in a non-fractal environment (Fig. 3c). The two portions of the curve correspond to two spatial organizations of the elements on this map: a clearly delimited urban center and small, well-separated, built clusters. The distance threshold identified corresponds to the spatial boundary between these two spatial organizations.

The spatial organization of Basel in 1957 and 1994 is markedly different from that in 1882. The number of built clusters declines much faster from one dilation step to the next. Distance thresholds are more difficult to identify because the spatial organization is more hierarchic. The metropolitan areas for both dates are close to a strict fractal hierarchy of the built clusters, as in the random fractal city in Fig. 3b. Distance thresholds differ for each of the three dates: 300 m for 1882, 279 m for 1957, and 173 m for 1994. Hence, the morphological boundary of an urban agglomeration based on a single predefined distance threshold is not consistent with the morphological history of the urban area of Basel. Its distance threshold has decreased over time. Two phenomena explain this decrease: (1) distances between buildings were greater in 1882 than in 1957 or 1994 and (2) the hierarchy of distances between buildings remained the same both within the urban agglomeration and outside it over a wide range of distances in 1882, but not in 1957 or 1994.

The shape of the urban envelope changed over time (Fig. 7). The urban agglomeration expanded significantly between 1882 and 1957, especially in
Figure 7. Basel’s boundaries in 1882 (a), 1957 (b), and 1994 (c). The boundaries of each urban agglomeration result from the successive application of a positive and a negative buffer the width of which is the distance threshold of Fig. 6. Data source: Bundesamt für Landestopographie, 3084 Wabern, Germany. Digitalized by research team “Image, Ville, Environnement,” Strasbourg, France.
Switzerland, where population growth was particularly high in the 1950s (Office fédéral de la statistique 2002). The urban agglomeration was located entirely within Switzerland in 1882 but became transnational in the first half of the 20th century. This change may be explained by the development of the tram network since 1895, serving both the French and the German zones of the metropolitan area. A ring of hamlets around the urban agglomeration persisted in all three countries.

Between 1957 and 1994, the urban agglomeration underwent fewer changes, although the morphological evolutions are still noteworthy. The dominant trend is the expansion of the urban agglomeration in France and its contraction in Germany and Switzerland. Earlier gaps in the built pattern were filled in the French part of the agglomeration (near Hegenheim and Huningue). Conversely, gaps widened or formed in the German and Swiss part of the agglomeration (e.g., in the south of Basel and in Weil am Rhein); built clusters that belonged to the urban agglomeration in 1957 became separate by 1994 (e.g., Pratteln in the east of the Swiss part of the agglomeration). Differences in the evolution of the urban agglomeration between France, on the one hand, and Germany and Switzerland, on the other, can be attributed to differences in planning and economic policies in the three countries. However, the ring of hamlets around the urban agglomeration in the three countries remained.

Conclusions drawn on the basis of the identification of the envelope of Basel at these three dates are consistent with existing knowledge about the evolution of this metropolitan area. In particular, Reitel (2006) shows the late but rapid growth of the metropolitan area after 1880 to be a consequence of the development of its chemical industry: at that time, the metropolitan area spread in both France (this part of France being German at that time) and Germany because of the establishment of industrial outlets for goods for the German market. The same study shows that more recently, in spite of the creation of a transnational framework, the spatial organization of the urban agglomeration remains clearly different in the three countries (Reitel 2006).

**Conclusion**

A fractal-based method is proposed for identifying the morphological boundaries of urban agglomerations by detecting multiscale changes in their built morphologies. A dilation of each individual building is first applied; then a distance threshold is detected for the dilation curve, leading to a characterization of its urban envelope. This distance threshold corresponds to the dilation step at which distances separating buildings no longer exhibit the same fractal (or multifractal) behavior. Thus, the traditional morphological criterion (the maximum distance threshold between neighboring buildings) is superseded by a multiscale definition of morphological connectivity. Theoretical and real-world examples illustrate the operationality of this method.

While the proposed method can be applied to any metropolitan area regardless of its size or characteristics, significant thresholds may or may not be detected. If
built elements are uniformly distributed in space, no urban agglomeration is distinguishable because a dilation threshold can be identified only if local variations in distances separating buildings change dramatically at some scale. In contrast, if local variation of the distances separating buildings is always the same across scales, or if distances separating buildings do not vary, no threshold appears.

A threshold identified for an agglomeration may vary with the geographic extent of a study area. Any such variation is interesting because it may reveal hidden features of the spatial organization of that study area. This contention is consistent with Openshaw’s (1984) idea that the modifiable areal unit problem is not one that has to be solved using sophisticated means, but it is a useful tool for exploring the multiscalar structure of a phenomenon (Lam and Quattrochi 1992). Moreover, besides identifying an urban boundary for just one (either monocentric or polycentric) agglomeration, the method could be applied to larger spaces to search for morphological thresholds common to several metropolitan areas or to identify differences between the thresholds found for a set of metropolitan areas and the threshold found for a wider study area encompassing that set of individual metropolitan areas. This approach could bridge the gap between studies of individual cities and of systems of cities. The method proposed could further our understanding of urban growth and the changes affecting urban settlements. This potential now needs to be tested more thoroughly over a wider range of case studies.

Note
1 Mandelbrot defines multifractality differently. He builds multifractal patterns by introducing several reduction factors into the same generator (Mandelbrot 1982). Such multifractal patterns are characterized by several fractal dimensions that do not change regardless of the range of scales. The two points of view (Mandelbrot’s and that expressed in this article), however, are not mutually exclusive.

References


