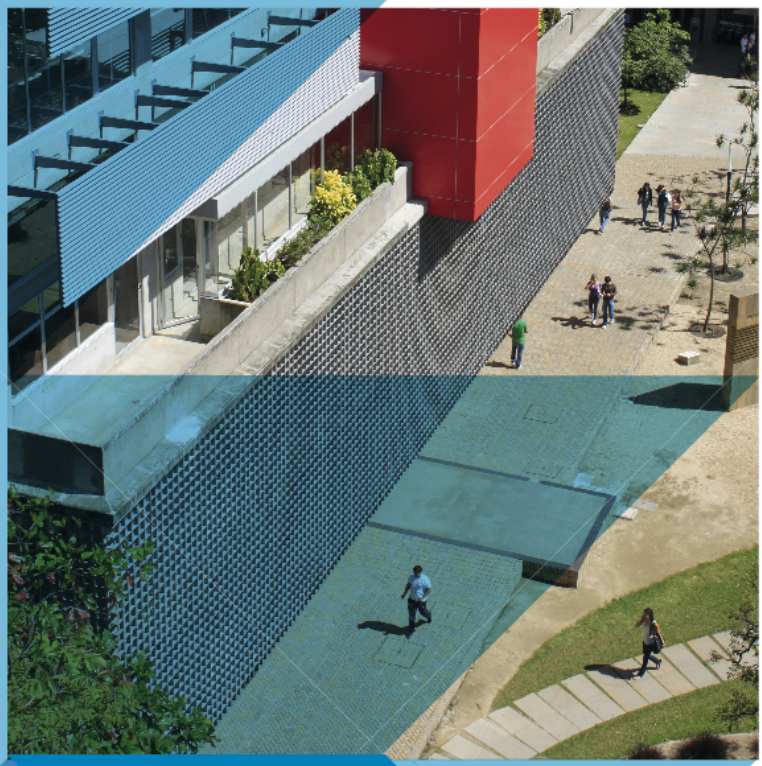


UNIVERSIDAD EAFIT
Abierta al mundo

Escuela de Economía y Finanzas

Documentos de trabajo

Economía y Finanzas

Centro de Investigación
Económicas y Financieras

No. 13-20
2013

An algorithmic approach for simulating realistic irregular lattices

Duque, Juan C.; Betancourt, Alejandro; Marín, Freddy

cief
Centro de Investigaciones
Económicas y Financieras
Universidad EAFIT

An algorithmic approach for simulating realistic irregular lattices*

Juan C. Duque[1]
Alejandro Betancourt[2]
Freddy Marin[3]

July 15 2013

Resumen/abstract
There is a wide variety of computational experiments, or statistical simulations, in which regional scientists require regular and irregular lattices with a predefined number of polygons. While most commercial and free GIS software offer the possibility of generating regular lattices of any size, the generation of instances of irregular lattices is not a straightforward task. The most common strategy in this case is to find a real map that matches as closely as possible the required number of polygons. This practice is usually conducted without considering whether the topological characteristics of the selected map are close to those for an "average" map sampled in different parts of the world.
In this paper, we propose an algorithm, *RI-Maps*, that combines fractal theory, stochastic calculus and computational geometry for simulating realistic irregular lattices with a predefined number of polygons. The irregular lattices generated with *RI-Maps* have guaranteed consistency in their topological characteristics, which reduces the potential distortions in the computational or statistical results due to an inappropriate selection of the lattices.

Palabras claves/Keywords
RI-Maps, MR-Polygons, Regional Science, Lattices, Computation, Experiment

JEL Clasification
C02, C63

[1] Research in Spatial Economics (RISE-group), Department of Economics,
EAFIT University,
e-mail: jduquec1@eafit.edu.co
Phone: (574) 261-9354, Fax: (574) 261-9294

[2] Research in Spatial Economics (RISE-group), Department of Economics,
EAFIT University

[3] Department of Fundamental Sciences,
EAFIT University

* The authors wish to thank Colciencias (Departamento Administrativo de Ciencia y Tecnología e Inovación) for their financial support under the program "Jovenes Investigadores." The authors also thank the Cyberinfrastructure Service for High Performance Computing, "Apolo", at EAFIT University, for allowing us to run our computational experiments in their supercomputer. The usual disclaimer applies.

1 Introduction

The complexity of computational experimentation in regional science has drastically increased in recent decades. Regional scientists are constantly developing more efficient methods, taking advantage of modern computational resources and geocomputational tools, to solve larger problem instances, generate faster solutions or approach asymptotics. The literature has many examples: The first formulation of the p-median problem provides a numerical example that required 1.51 minutes to optimally locate four facilities in a 10-node network (ReVelle and Swain, 1970); three decades later, Church (2008) located five facilities in a 500-node network in 1.68 minutes. As noted by Anselin et al. (2004), spatial econometrics has also benefitted from computational advances; the computation of the determinant required for maximum likelihood estimation of a spatial autoregressive model proposed by Ord (1975) was feasible to apply for data sets not larger than 1,000 observations. Later, Pace and LeSage (2004) proposed a procedure that was capable of computing this determinant for over a million observations. According to Blommestein and Koper (2006), one of the first algorithms for constructing higher-order spatial lag operators, which was devised by Ross and Harary (1952), required 8,000 seconds (approximate computation time) to calculate the sixth-order contiguity matrix in a 100x100 regular lattice. Anselin and Smirnov (1996) proposes new algorithms that are capable of computing a sixth-order contiguity matrix for the 3,111 U.S. contiguous counties in less than a second.

An important aspect when conducting computational experiments in regional science is the selection of the way that the spatial phenomena are represented or conceptualized. This aspect is of special relevance when using a discrete representation of continuous space, such as polygons (Haining, 2010). This representation can be accomplished in two ways: regular or irregular lattices; the use of one or the other could cause important differences in the computational times, solution qualities or statistical properties. We suggest four examples, as follows: (1) The method proposed by Duque et al. (2011a) for running the AMOEBA algorithm (Aldstadt and Getis, 2006) requires an average time of 109 seconds to delimit four spatial clusters on a regular lattice with 1,849 polygons. This time rises to 229 seconds on an irregular lattice with the same number of polygons. (2) For the location set covering problem, Murray and O'Kelly (2002) concluded that the spatial configuration, number of needed facilities, computational requirements and coverage error all varied significantly as the spatial representation was modified. (3) Elhorst (2003) warns that the parameters of the random effects spatial error and spatial lag model might not be an appropriate specification when the observations are taken from irregular lattices.[1] (4) Anselin and Moreno (2003) finds that the use of regular or irregular lattice affects the performance of test statistics against alternatives of the spatial error components form.

However, returning to the tendency toward the design of computational experiments with large instances, there is an important difference between generating large instances of regular and irregular lattices. On the one hand, regular lattices are easy to generate, and there is no restriction on the maximum number of polygons. On the other hand, instances of irregular lattices are usually made by sampling real maps. Table .1 shows some examples of this practice.

The generation of large instances of irregular lattices has several complications that are of special interest in this paper. First, the size is limited to the sizes of the available real lattices. Second, the possibility of generating a large number of different instances of a given size is also limited (e.g., generate 1,000 instances of irregular lattices with 3,000 polygons). Third, as shown in Fig. 1, the topological characteristics of irregular lattices built from real maps change drastically, depending on the region from where they are sampled, which could bias the results of the computational experiments.[2]

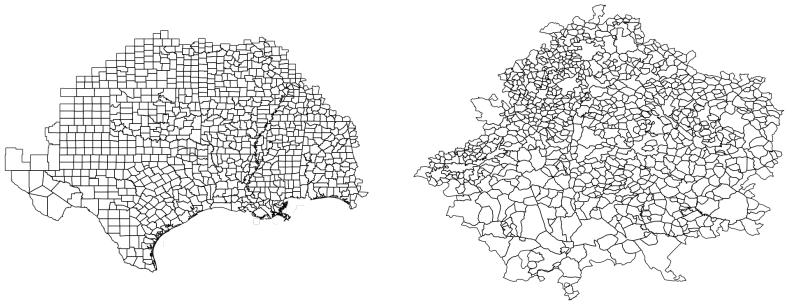
This paper seeks to contribute to the field of computational experiment design in regional science by proposing a scalable recursive algorithm (*RI-Maps*), which combines concepts from

[1] See also Anselin (1988), pg 51

[2] Later in this paper, we show that the topological characteristics of Voronoi diagrams are far from those for an "average" map sampled in different parts of the world

Table .1 Annotated chronological listing of studies that use irregular lattices generated by sampling real maps.

Study	Purpose	Source of irregular lattices
Mur Lacambra (1992)	Compares different methods to detect spatial autocorrelation.	Spain provinces in 1985 (sizes 14 and 48 polygons)
Anselin et al. (1996)	Performance of a diagnostic test for spatial dependence.	COROP and economic geographic regions in The Netherlands (sizes 40 and 81, respectively)
Smirnov and Anselin (2001)	Performance of a new method for evaluating the Jacobian term.	921 counties (Kreise) for Germany; 3,107 U.S. continental counties; 3,140 U.S. counties, and 29,762 U.S. postal zip codes.
Anselin and Moreno (2003)	Extend the knowledge about the properties of spatial correlation tests, especially in empirical applications.	Spatial grouping of Western U.S. counties for dimensions 46, 80, 124, 264, 413 and 1,013
Duque et al. (2012)	Performance of an algorithm for spatial clustering (the max-p-regions model)	Sacramento census tracks (403), Colombian municipalities (1,068) and U.S. census tracks (3,085).



(a) United States (b) Spain

Fig. .1 Examples of two instances of 900 irregular polygons.

stochastic calculus (mean reversing processes), fractal theory and computational geometry to generate instances of irregular lattices that have an unlimited size and topological characteristics that are a good representation of the irregular lattices sampled from around the world. The use of instances obtained from *RI-Maps* will guarantee that differences in the results in the computational experiment will not come from differences in the topological characteristics of the used lattices and will also make it easier to generate the unlimited number of large instances.[3]

The remainder of this paper is organized as follows: Section 2 introduces the basic definitions of the polygons and lattices and proposes a consensus taxonomy of the lattices. Section 3 presents a set of indicators that are used to characterize the topological characteristics of a lattice and shows the topological differences between regular and irregular lattices. Section 4 presents the algorithm

[3] This algorithm as well as all of the instances of *RI-Maps* used in this article will be publicly available for the academic community after its publication (Duque et al., 2011b).

for generating irregular lattices. Section 5 evaluates the capacity of the algorithm to generate realistic irregular lattices. Finally, Section 6 presents the conclusions.[4]

2 Conceptualizing polygons and lattices

A polygon is a plane figure enclosed by a set of finite straight line segments. Polygons can be categorized according to their boundaries, convexity and symmetry properties, as follows:

i) Boundary: A Polygons is *Simple* when it is formed by a single plain figure with no holes, and it is *Complex* when it contains holes or multiple parts.[5]
ii) Convexity: In a *convex* polygon, every pair of points can be connected by a straight line without crossing its boundary. A *Concave* polygon is simple and non-convex.
iii) Symmetry: A *regular* polygon has all of its angles of equal magnitude and all of its sides of equal length. A non-regular polygon is also called *irregular* (Johnson, 2001; Coxeter, 1974).

A lattice is a set of polygons of any type, with no gaps and no overlaps, that covers a subspace or the entire space. Next, a more formal definition: A lattice is the division of a subspace $S \subseteq R^n$ into k subsets $i \subseteq S$ such that $\cup_i = S$ and $\cap_i = \phi$, where ϕ is the empty set of R^n (Grunbaum and Shephard, 2011)[6]. There exist different taxonomies of lattices depending on the field of study. In an attempt to unify these taxonomies, a consensus lattice taxonomy is presented in Fig. .2. This taxonomy classifies lattices according to the shapes of their polygons, the spatial relationships between them, and the use, or not, of symmetric relationships to construct the lattice:[7]

i) According to the variety of the shapes of the polygons that form the lattice: **homomorphisms** are lattices that are formed by polygons that have the same shape, and **polymorphisms** are lattices that are formed by polygons that have different shapes.
ii) According to the regularity of the polygons that form the lattice and the way in which they intersect, each vertex:[8] **Regular**, lattices formed by regular polygons in which all of the vertexes join the same arrangement of polygons (Tilley, 2006); **Semi-regular**, when the polygons are regular but there are different configurations of vertexes; and **Irregular** otherwise (Ghyka, 2004).
iii) According to the existence of symmetric relationships within the lattice:[9] **Symmetric**, when the lattice implies the presence of at least one symmetric relationship; and **Asymmetric** otherwise.
iv) According to the symmetric relationship of translation: A lattice is **periodic** if and only if it implies the use of translation without rotation or reflection; it is **Aperiodic** otherwise (Tilley, 2006).

Table .2 shows an example of each category of this consensus taxonomy.

---

[4] A dataset of 700 RI-Maps is available at "http://www.rise-group.org/section/Research/Publication/AnAlgorithmicAproach/"

[5] Complex polygons do not refer to polygons that exist in the Hilbert plane (Coxeter, 1974).

[6] This paper focuses exclusively on bidimensional lattices (i.e., $n = 2$)

[7] An alternative category is proposed for lattices formed by fractal polygons that are informally defined by Mandelbrot (1982) as rough fragmented geometric shapes that could be infinitely divided into scalable parts.

[8] considering the vertexes to be all of the points of the lattice that intersect three or more polygons)

[9] There are three types of symmetrical relationships: *Translation*, when the lattice is formed by translating a subset of polygons; *reflection*, when there are axes of reflection in the lattice; and *rotation*, when it is possible to obtain the same lattice after a rotation process of less than 2π (Radin, 1993).

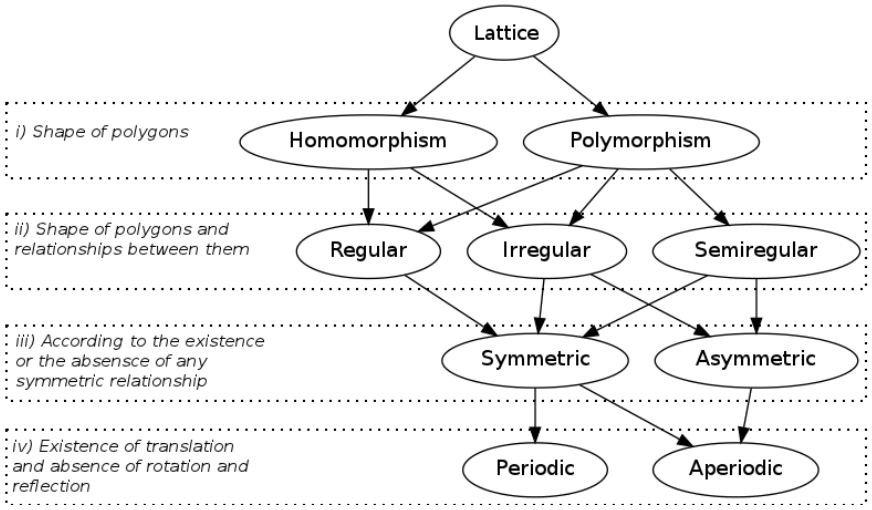
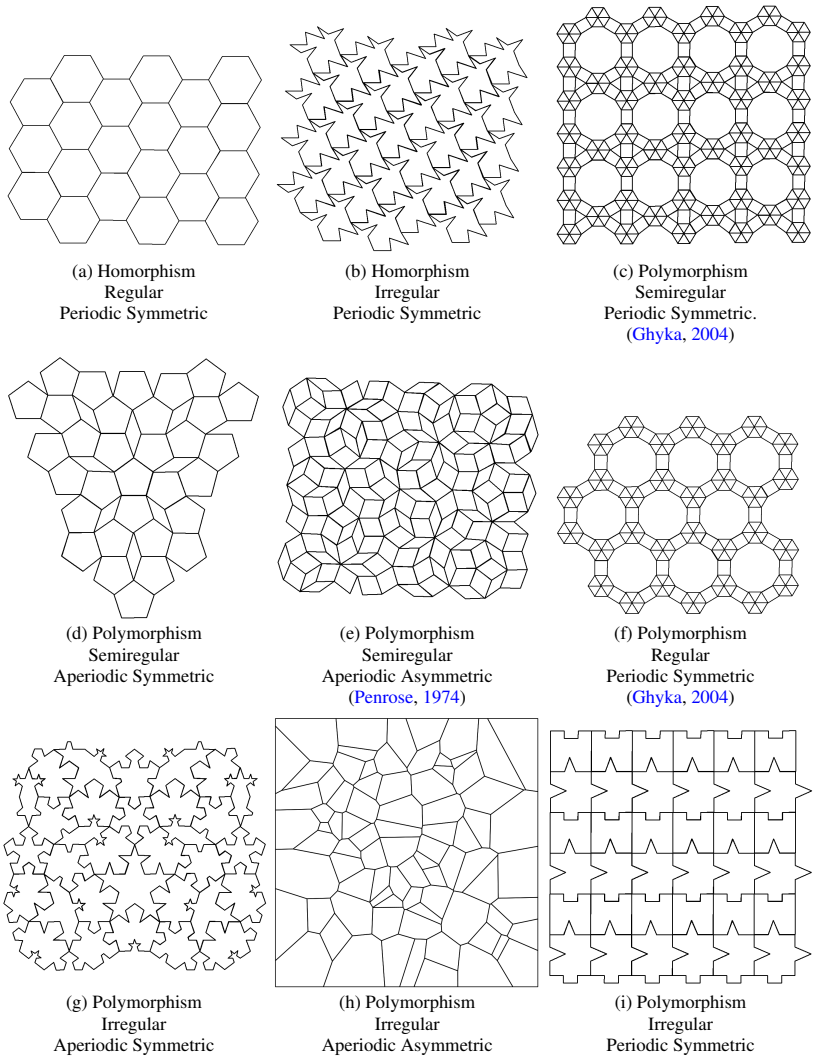


Fig. .2 Consensus taxonomy of lattices

The topological characteristics of lattices are usually summarized through the properties of the sparse matrix that represent the neighboring relationships between the polygons in the map, the so-called W matrix (Gould, 1967; Boots, 1982; Le Caer and Delannay, 1995; Aste et al., 1996; Peshkin et al., 1991).[10] This paper uses six indicators; the first three indicators are self-explained: the maximum (M_n), minimum (m_n) and average (μ_1) number of neighbors per polygon. The fourth indicator, Sparseness *(S)*, is the percentage of non-zero elements of the W matrix. The fifth indicator is the first eigenvalue of the W matrix (λ_1), which is an algebraic tool that is frequently used in graph theory (Garrison, 1964; Tinkler, 1972) and regional science (Gould, 1967; Boots, 1982, 1984, 1985) to summarize different aspects of the W matrix. The last indicator is the variance of the number of neighbors per polygon (μ_2), which measures the spatial disorder of a lattice.

Within the field of regional science, lattices are frequently used with two purposes: First, real lattices can be used to study real phenomena, e.g., to analyze spatial patterns, confirm spatial relationships between variables, and detect spatio-temporal regimes within a spatial panel, among others. Second, lattices can be used to evaluate the behavior of statistical tests (Anselin and Moreno, 2003; Mur Lacambra, 1992); algorithms (Duque et al., 2011a); and topological characteristics of lattices (Aste et al., 1996; Le Caer and Delannay, 1995, 1993). In these cases, it is necessary to use sets of lattices that satisfy some requirements imposed by the regional scientist, e.g., the number of polygons, regularity or irregularity of the polygons and the number of instances. To accomplish this goal, it is a common approach to use a geographical base for real or simulated data *polymorphism irregular aperiodic asymmetric* (e.g., real lattices and Voronoi diagrams) or *homomorphism regular periodic symmetric* (e.g., regular lattices). The next sections focus on the second use of lattices.

[10] See Anselin (1988) for more information about this matrix.



(a) Homorphism
Regular
Periodic Symmetric

(b) Homorphism
Irregular
Periodic Symmetric

(c) Polymorphism
Semiregular
Periodic Symmetric.
(Ghyka, 2004)

(d) Polymorphism
Semiregular
Aperiodic Symmetric

(e) Polymorphism
Semiregular
Aperiodic Asymmetric
(Penrose, 1974)

(f) Polymorphism
Regular
Periodic Symmetric
(Ghyka, 2004)

(g) Polymorphism
Irregular
Aperiodic Symmetric

(h) Polymorphism
Irregular
Aperiodic Asymmetric

(i) Polymorphism
Irregular
Periodic Symmetric

Table .2 Example Lattices

3 Topological characteristics of regular and irregular lattices

As stated above, regional scientists have the option of using regular or irregular lattices in their computational experiments. However, this section will show that there are important topological differences between these types of lattices.

Real lattices have topological characteristics that vary substantially from location to location. As an example, Fig. .3 presents the topological characteristics of lattices of different sizes (100, 400 and 900 polygons) sampled in Spain and the United States. Each box-plot summarizes 1,000 instances. Important differences emerge between these two places: Spanish polygons tend to have

more neighbors, are more disordered, and their first eigenvalues are higher in mean and variance. These differences in the topological characteristics have direct repercussions on the performance of algorithms whose complexity depends on the neighboring structure (Aldstadt and Getis, 2006; Duque et al., 2011a).

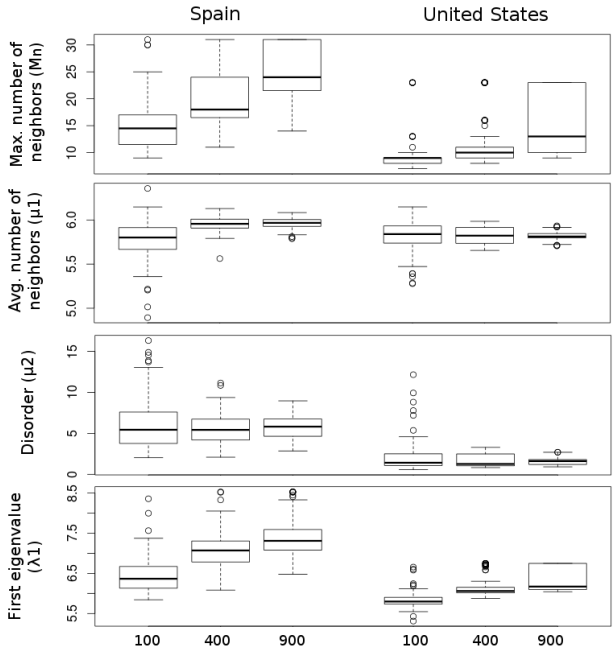


Fig. .3 Topological differences of lattices from Spain and the United States

Regular lattices and Voronoi diagrams are also commonly used for computational experiments because they are easy to generate, there is no restriction on the size of the instances (the number of polygons in the map), and their over-simplified structure allows for some mathematical simplifications or reductions (Whittle, 1954; Bartlett, 1975; Griffith, 1987). However, the topological characteristics of these lattices are substantially different from real, irregular lattices. These differences can lead to biased results in theoretical and empirical experiments, e.g., spatial stationarity in STARMA models (Hooper and Hewings, 1981), improper conclusions about the properties of the power and sample sizes in hypothesis testing (Anselin and Moreno, 2003; Mur Lacambra, 1992), and the over-qualification of the computational efficiency of the algorithms (Duque et al., 2011a; Aldstadt and Getis, 2006), among others. Table .3 shows the topological differences between real maps, two types of regular lattices and Voronoi diagrams.

To illustrate the magnitude of these differences, we calculated the topological indicators (M_n, m_n, μ_1, μ_2, S and λ_1) for six thousand lattices of different sizes (1,000 instances each of 100, 400, 900, 1,600, 2,500, and 3,600 polygons) that were sampled around the world at the smallest administrative division available in Hijmans et al. (2011). As an example, Fig. .4 shows seven of those instances. These real instances are then compared to regular lattices that have square and hexagonal polygons and Voronoi diagrams.[11] To avoid the boundary effect on M_n, m_n, μ_1 and μ_2,

[11] Each one of the six-thousand instances of Voronoi diagrams come from uniformly distributed points.

the bordering polygons are only considered to be neighbors of interior polygons. Last, S and λ_1 are calculated using all of the polygons. Table .3 shows that regular lattices are not capable of emulating the topological characteristics of real lattices in any of the indicators: $\mu_2 = 0$ and M_n, m_n, $\mu_1 = 4$ and 6 (for squares and hexagons, respectively) are values that are far from those of real lattices. The values obtained for λ_1 and S indicate that regular lattices of hexagons are more connected than real lattices, while regular lattices of squares are less connected than real lattices. With regard to Voronoi diagrams, M_n and m_n indicate that they are not capable of generating atypically connected polygons. The values of μ_1 are close to real lattices. Finally, Voronoi diagrams are more ordered than real lattices, with values of μ_2 close to 1.7, while real lattices report values of μ_2 that are close to 8.

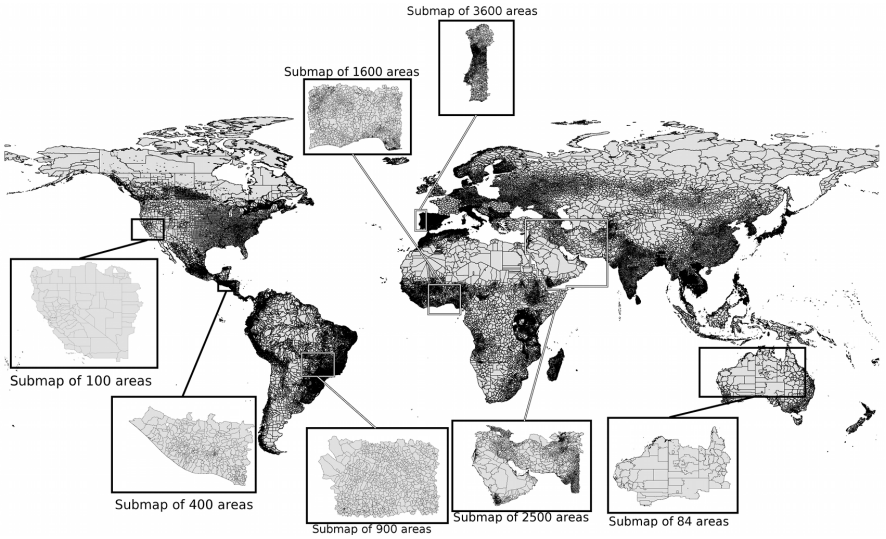


Fig. .4 Base map and example of a random irregular lattice obtained from it.

4 RI-Maps: An algorithm for generating realistic irregular lattices

This section is divided into two parts. The first part introduces an algorithm that generates irregular polygons based on a mean reverting process in polar coordinates, and the second part proposes a novel method to create polymorphic irregular aperiodic lattices with topological characteristics that are similar of those from real lattices.

Table .3 Average topological characteristics for real maps, regular lattices and Voronoi diagrams

		Number of polygons						
		81	100	400	900	1,600	2,500	3,600
Real lattices	M_n	12.28 ±7.52	13.22 ±9.90	23.22 ±29.89	29.55 ±36.27	42.77 ±50.59	48.53 ±55.07	60.64 ±64.58
	m_n	2.33 ±1.11	2.13 ±1.06	1.55 ±0.86	1.23 ±0.60	1.04 ±0.22	1.01 ±0.10	1.00 ±0.00
	μ_1	5.57 ±0.65	5.59 ±0.61	5.67 ±0.49	5.69 ±0.45	5.70 ±0.46	5.72 ±0.37	5.72 ±0.37
	μ_2	5.85 ±13.85	6.72 ±22.58	9.76 ±28.79	7.90 ±15.35	8.85 ±12.82	7.73 ±9.39	8.00 ±8.11
	S	5.98 ±0.51	4.91 ±0.43	1.30 ±0.11	0.58 ±0.046	0.33 ±0.02	0.21 ±0.01	0.15 ±0.01
	λ_1	5.96 ±0.53	6.09 ±0.65	6.89 ±1.52	7.30 ±1.82	8.03 ±2.42	8.33 ±2.62	8.92 ±3.02
Reg. Lattice *(Squares)*	M_n	4	4	4	4	4	4	4
	m_n	4	4	4	4	4	4	4
	μ_1	4	4	4	4	4	4	4
	μ_2	0	0	0	0	0	0	0
	S	4.44	3.64	0.95	0.43	0.24	0.16	0.11
	λ_1	3.80	3.84	3.96	3.98	3.99	3.99	3.99
Reg. Lattice *(Hexagons)*	M_n	6	6	6	6	6	6	6
	m_n	6	6	6	6	6	6	6
	μ_1	6	6	6	6	6	6	6
	μ_2	0	0	0	0	0	0	0
	S	6.30	5.19	1.39	0.64	0.36	0.23	0.16
	λ_1	5.55	5.62	5.88	5.94	5.96	5.97	5.98
Voronoi Diagrams	M_n	9.15 ±0.77	9.36 ±0.79	10.37 ±0.75	10.90 ±0.74	11.26 ±0.70	11.49 ±0.67	11.71 ±0.68
	m_n	3.36 ±0.48	3.26 ±0.44	3.00 ±0.03	3.00 ±0.00	3.00 ±0.00	3.00 ±0.00	3.00 ±0.03
	μ_1	5.75 ±0.07	5.77 ±0.05	5.88 ±0.02	5.92 ±0.01	5.94 ±0.00	5.95 ±0.00	5.96 ±0.00
	μ_2	1.68 ±0.31	1.70 ±0.27	1.75 ±0.13	1.76 ±0.09	1.76 ±0.07	1.77 ±0.05	1.77 ±0.04
	S	6.67 ±0.08	5.47 ±0.05	1.44 ±0.00	0.65 ±0.00	0.37 ±0.00	0.24 ±0.00	0.17 ±0.00
	λ_1	5.88 ±0.05	5.96 ±0.05	6.20 ±0.03	6.26 ±0.02	6.28 ±0.02	6.29 ±0.02	6.30 ±0.02

4.1 Mean reverting polygons (MR-Polygons)

The problem of characterizing the shape of irregular polygons is commonly addressed in two ways, that is, evaluating its similitude with a circle (Haggett, 1977) or describing its boundary roughness through its fractal dimension (Batty and Lon-

gley, 1994; Frankhauser, 1998).[12] In this paper, we apply both concepts in different stages during the creation of a polygon: the similitude with a circle to guide a mean reverting process in polar coordinates, and the fractal dimension to parameterize the mean reverting process.

4.1.1 Mean reverting process in polar coordinates

Different indexes are used to compare irregular polygons with a circle: Elongation ratio (Weeitty, 1969), form ratio (Horton, 1932), circularity ratio (Miller, 1953), compactness ratio (Gibbs, 1961; Cole, 1964; Richardson, 1961), ellipticity index (Stoddart, 1965) and the radial shape index (Clark, 1964). As Chen (2011) states, all of these indexes are based on comparisons between the irregular polygon and its area-equivalent circle. Under this relationship, an irregular polygon can be conceptualized as an irregular boundary with random variations following a circle, which lead us to use a mean reverting process in polar coordinates to create irregular polygons. A mean reverting process is a stochastic process that takes values that follow a long-term tendency in the presence of short-term variations. Formally, the process x at the moment t is the solution of the stochastic differential equation (.1), where μ is the long-term tendency, α is the mean reversion speed, σ is the gain in the diffusion term, $x(t_0)$ is the value of the process when $t = 0$ and $\{B_t\}_{t\geq 0}$ is an unidimensional Brownian (Mao, 1997). Equation (.2) shows the general solution; however, for practical purposes, hereafter we use the Euler discretization method, which is given by equation (.3), where ε_t is white noise.

$$dX_t = \alpha(\mu - X_t)d_t + \sigma dB_t \tag{.1}$$

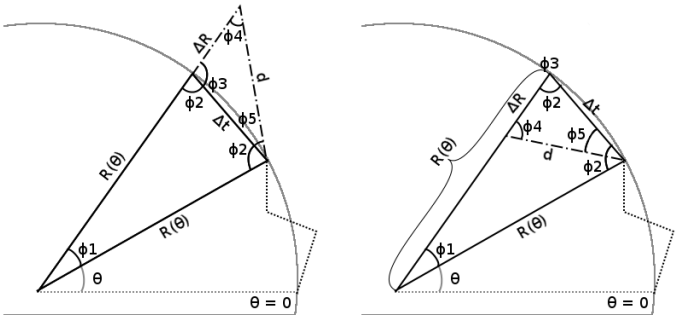
$$x(t) = e^{-\alpha(t-t_0)}\left(x(t_0) + \int_{t_0}^{t} e^{\alpha(s-s_0)}\alpha\mu ds + \int_{t_0}^{t} e^{\alpha(s-s_0)}\sigma dB(s)\right), \tag{.2}$$

$$X_t = X_{t-1} + \alpha(\mu - X_{t-1})\Delta_t + \sigma\sqrt{\Delta_t}\varepsilon_t \tag{.3}$$

Algorithm 1 presents the procedure for generating an irregular polygon P in polar coordinates using, as a data generator, a mean reverting process (X_t). This algorithm guarantees that the distance between two points in X_t, following the process X_t, is equal to the distance between the same two points in P when following the process P counterclockwise. The purpose of this equivalence is to preserve the fractal dimension of X_t in P. The angles Δ_R and ϕ_1 in algorithm 1 are the result of solving the geometric problem in Fig. .5. These two angles are used in equation (.4) to establish the location of the next point in P. The points of P are denoted as P_θ, with θ between 0 and 2π.

$$P_{\theta+\phi_1} = \begin{cases} P_\theta + \Delta_R & \text{if } X_{t+\Delta_t} \geq Xt \\ P_\theta - \Delta_R & \text{if } X_{t+\Delta_t} < Xt. \end{cases} \tag{.4}$$

[12] Chen (2011) established a relationship between these two approaches.



(a) $X_t \geq X_{t-\Delta_t}$ (b) $X_t < X_{t-\Delta_t}$

Fig. .5 Geometric problem to preserve the length and the fractal dimension of the mean reverting process when it is used to create an irregular polygon

Algorithm 1 MR-Polygon: Mean Reverting Polygon

function MEANREVERTINGPOLYGON($\alpha, \sigma, \mu, X_0, \Delta_t$)
  $X_{t-\Delta_t} = X_0$ ▷ Initial point of the mean reverting process
  $P = [(0, X_0)]$ ▷ Irregular polygon in polar coordinates
  while $\theta < 2\pi$ **do**
    $\varepsilon_t \leftarrow RandomNormal(0,1)$
    $X_t = X_{t-\Delta_t} + \alpha(\mu - X_{t-\Delta_t})\Delta_t + \sigma\sqrt{\Delta_t}\varepsilon_t$
    $d \leftarrow distance(X_t, X_{t-\Delta_t})$
    $R_\theta \leftarrow$ Last radius of the irregular polygon
    $\phi_1 = \arccos\left(\frac{2R_\theta - \Delta_t^2}{2R_\theta^2}\right)$
    if $Xt \geq X_{t-\Delta_t}$ **then**
      $\Delta_R = d\left(\cos\left(\arcsin\left(\frac{\Delta_t}{d}\cos\left(\frac{\phi_1}{2}\right)\right) - \sin\left(\frac{\phi_1}{2}\right)\frac{\Delta_t}{d}\right)\right)$
    else
      $\Delta_R = -d\left(\cos\left(\arcsin\left(\frac{\Delta_t}{d}\cos\left(\frac{\phi_1}{2}\right)\right) + \sin\left(\frac{\phi_1}{2}\right)\frac{\Delta_t}{d}\right)\right)$
    end if
    $R_{\theta+\phi_1} = R_\theta + \Delta_R$
    Add $(\theta + \phi_1, R_{\theta+\phi_1})$ to P
    Increase θ in ϕ_1
  end while
  Replace last point of P to $(0, X_0)$
  return P
end function

Because the process P depends on the parameters α, μ and σ, it is worthwhile to clarify their effect on the shape of polygon P: α is the speed at which the process reverts to the circle with radius μ, and σ is the scaling factor of the irregularity of the polygon. High values of α and low values of σ generate polygons that have shapes that are close to a circle with radius μ. Finally, Δ_t is utilized to preserve the

fractal dimension of both processes, X and P, and determines the angular step, ϕ_1 (see Fig. .5).

4.1.2 MR-Polygon parameterization

The process of establishing the values for α, μ, σ, Δ_t and X_0 is not an easy task, and their values must be set in such a way that the shape of P is similar to a real irregular polygon. However, how do we determine whether a polygon P satisfies this condition? In this case, the fractal dimension appears to be a tool that offers strong theoretical support to assess the shape of a given polygon.

According to Richardson (1961), the fractal dimension D of an irregular polygon (such as a coast) is a number between 1 and 2 (1 for smooth boundaries and 2 for rough boundaries) that measures the way in which the length of an irregular boundary L (equation .5) changes when the length of the measurement instrument (ε) changes. The fractal dimension in given by equation (.6), where $\hat{C}$ is a constant.

In general, an object is considered to be a fractal if it is endowed with irregular characteristics that are present at different scales of study (Mandelbrot, 1982). For practical purposes, D is obtained using equation (.6) and is given by 1 minus the slope of $\log(L(\varepsilon))$. This procedure is commonly known as the Richardson plot.

$$L(\varepsilon) = \hat{C}\varepsilon^{1-D} \tag{.5}$$

$$\log(L(\varepsilon)) = (1-D)\log(\varepsilon) - \log(\hat{C}) \tag{.6}$$

In almost all cases, the Richardson plot can be explained with two line segments that have different slopes; then, two fractal dimensions can be obtained: textural, for small scales, and structural, for large scales (Kindratenko and Treiger, 1996). As illustrated, Fig. .6 shows a segment of the United States east coast taken from Google maps in two resolutions. Note that as the resolution increases, some irregularities that were imperceptible at low resolution become visible. In this sense, it can be said that irregularities at low resolution define the general shape and are related to the structural dimension, while irregularities at high resolution capture the noise and are related to the textural dimension. Regional scientists tend to use highly sampled maps, which preserve the general shape but remove the small variations. This simplification does not change the topological configuration of the maps (Douglas, 1973). Fig. .7 presents the Richardson plot of the external boundary of the United States and its textural and structural fractal dimension.

In the field of stochastic processes, some approaches, which are based on different estimations of the length, have been made to characterize them through their fractal dimension. In our case, an experimental approach based on the fractal dimension of real polygons is proposed to select an appropriate combination of the parameters α and σ to generate realistic irregular polygons. Because our interest is on general shape rather than small variations, we account only for the struc-

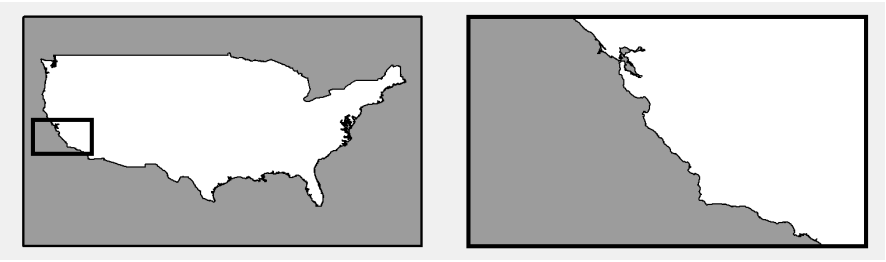


Fig. .6 Illustrative example of irregularities explained by the structural and textural dimension.

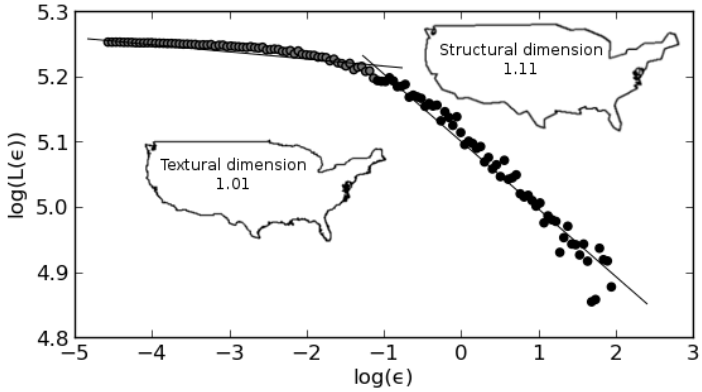


Fig. .7 Richarson plot to estimate the textural and structural dimension of the external boundary of the United States.

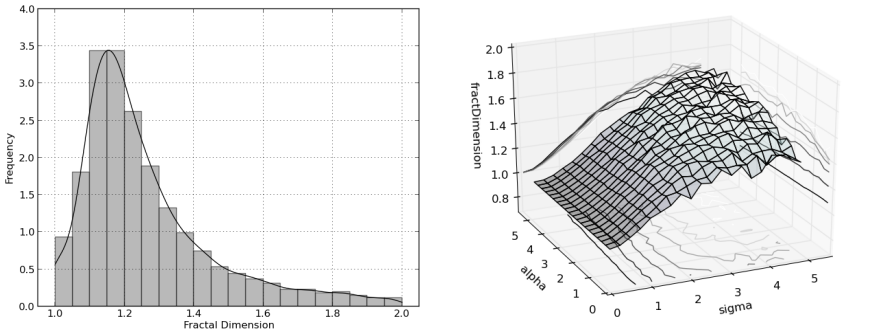
tural dimension[13] The parameterization process is divided into two parts: in the first part, the frequency histogram of the fractal dimensions of the real polygons is constructed. In the second part, we propose a range of possible values for α and σ, given μ, X_0, Δ_t, which generates fractal dimensions that are close to those obtained in the first part. Because the level of the long-term tendency μ does not affect the length of X and because algorithm 1 guarantees that the length is preserved, μ can be defined as a constant without affecting the fractal dimension. Hereafter, it is assumed that $\mu = X_0 = 10$. The value of Δ_t is set to be 0.001 to properly infer both of the fractal dimensions.

The empirical distribution of the fractal dimension of the irregular polygons is calculated over a random sample of 10,000 polygons from the world map used in section 3. The result of this empirical distribution is presented in Fig. .8.a. To find the fractal dimension of the *MR-Polygons*, we generate a surface of the average dimensions as a function of the values of α and σ, which range from 0.01 to 5 with

[13] To calculate the structural dimension, we use the EXACT procedure, which is devised by Allen et al. (1995), with a small value for Δ_t. Next, both of the dimensions were determined by using a k-means clustering algorithm over the cloud of points on the Richardson plot.

steps of 0.1 (Fig .8.b.). The resulting surface indicates that the fractal dimension is mainly affected by σ, especially when looking at small dimensions. Additionally, it is found that fractal dimensions close to 1.23 are obtained when σ takes on values between 1.2 and 1.5, regardless of the value of α.

Fig. .9 presents some examples of polygons using different values of α and σ. The polygons in the second row, which correspond to $\sigma = 1.5$, produce irregular polygons that have a realistic structural fractal dimension. Additionally, in the same figure, both the original (gray line) and sampled (black line) polygons reinforce the fact that sampling a polygon does not affect the structural dimension. From now on, we will use sampled polygons to improve the computational efficiency.



(a) Fractal dimensions of real polygons.

(b) Fractal dimension of simulated polygons as a function of α and σ.

Fig. .8 Stages to find the values of α and σ.

4.2 Recursive Irregular maps (RI-Maps)

Up to this point, we were able to generate irregular polygons with fractal dimensions that are similar to those from real maps. The next step is to use these polygons to create irregulars lattices of any size whose topological characteristics are close to the average values obtained for these characteristics in real lattices around the world. For this step, we formulate a recursive algorithm on which an irregular frontier is divided into a predefined number of polygons using *MR-Polygons*. Our conceptualization of the algorithm was made under three principles: (1) *Scalability:* Preserving the computational complexity of the algorithm when the number of polygons increases; (2) *Fractality:* Preserving the fractal characteristics of the map at any scale; and (3) *Correlativity:* Encouraging the presence of spatial agglomerations of polygons with similar sizes, which is commonly present in real maps in which there are clusters of small polygons that correspond to urban areas.

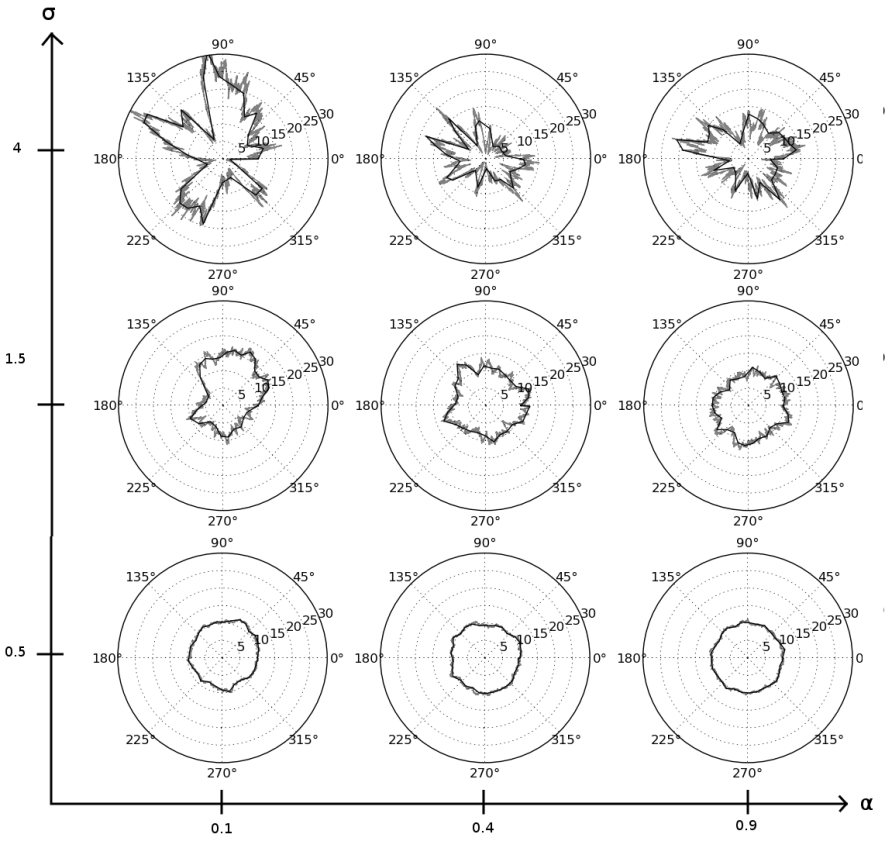


Fig. .9 Examples of stochastic polygons generated using (Algorithm 1) with different values of σ and α.

Algorithm 2 presents the *RI-Maps* algorithm to create polymorphic irregular aperiodic asymmetric lattices with realistic topological characteristics. This algorithm starts with an initial empty irregular polygon, *pol*, (the outer border of the *RI-Map*) and the number of polygons, *n*, to fit inside. In a recursive manner, a portion of the initial polygon *pol* starts being divided following a depth-first strategy until that portion is divided into small polygons. This process is repeated for a new uncovered portion of *pol* until the whole area of *pol* is covered. Because the recursive partitions are made by using *MR-Polygons*, we take the values of α and σ from a uniform distribution between 0.1 and 0.5 and 1.2 and 1.5, respectively, which correspond to the ranges of values that generate realistic polygons (see section 4.1). Regarding μ, X_0 and Δ_t, these variables take the same values proposed in section 4.1. Last, to guarantee the computational treatability of the geometrical operations, each polygon comes from a sampling process of 30 points.

To ensure a full understanding of the *RI-Maps* algorithm, we summarize in Fig. .10 its main steps, which indicates the link between the figure and the lines in the *RI-Maps* algorithm.

The *RI-Maps* algorithm has three unknown parameters:

- p_1: Because each polygon is created by the *MR-Polygons* using a polar coordinate system that is unrelated to the map being constructed with *RI-Maps*, it is necessary to apply a scaling factor, $\sqrt{\frac{p_1 \times area(pol)}{n \times \pi \times \mu^2}}$, that adjusts the size of the *MR-Polygon* before being included into the *RI-Map*.
- p_2: When a new polygon is used to divide its predecessor, its capacity to contain new polygons (measured by the number of polygons) is proportional to its share of the unused area of its predecessor. However, to enforce the appearance of spatial agglomerations of small polygons, the number of polygons that the new polygon can hold is increased with a probability of p_2.
- p_3: When p_2 indicates that a new polygon will hold more polygons, the number of extra polygons is calculated as the p_3 percent of the number of missing polygons that are expected to fit into the unused area of its predecessor polygon. The number of extra polygons is subtracted from the unused area to keep constant the final number of polygons (n).

Table .4 illustrates the effect of the parameters p_2 and p_3 on the topological characteristics of *RI-Maps*. In the first row, p_2 and p_3 equal 0, which generates highly ordered lattices without spatial agglomerations. The second and third rows are more disordered than the first row and have spatial agglomerations, with those in the second row less frequent and evident than those in the third row. As will be shown in the next section, lattices in the third row are more realistic in terms of their topological characteristics.

To find a combination of p_1, p_2, and p_3 that generates realistic *RI-Maps* in terms of their topological characteristics, we use a standard genetic algorithm, where the population γ at iteration i, denoted as γ^i, is formed by the genomes $\gamma^i_j = [p^i_{j_1}, p^i_{j_2}, p^i_{j_3}]$, where $p^i_{j_1}$, $p^i_{j_2}$, and $p^i_{j_3}$ are real numbers between 0 and 1, representing instances of p_1, p_2, p_3, which are denoted as phenomes. In this case, $i \in \mathbb{N}$ between 0 and 20 and $j \in \mathbb{N}$ between 0 and 100. To evaluate the quality of each genomes fitness function, $F(\gamma^i_j)$ is defined in equation (.7), where θ is a set of polygons, ϕ_k is the relative importance for a map of k polygons, and $f_k(\gamma^i_j)$ is a function given by equation (.8) that measures the average difference between the values of the topological indicators of real lattices and those values of *RI-Maps* formed by k polygons using the phenome γ^i_j. For the sake of simplicity, in equation (.8), $\Psi_k = [M_n, m_n, \mu_1, \mu_2, S, \lambda_1]$ denotes the vector of real indicators, and $\Psi_k(\gamma^i_j)$ denotes the vector for the mean values of *RI-Maps* with k polygons using γ^i_j. The superindex l is used in Ψ^l_k and $\Psi^l_k(\gamma^i_j)$ to refer to the l^{th} indicator in the real and simulated values, respectively. Finally, ns is the number of simulations to be generated with each genome.

Algorithm 2 RI-Map: Recursive Irregular Map

function RECURSIVEIRREGULARMAP(n, pol)
  $(\alpha_{min}, \alpha_{max}, \sigma_{min}, \sigma_{max}, \mu, X_0, \Delta_t) = (0.1, 0.5, 1.2, 1.5, 10, 10, 0.001)$
  $p_1 \in \mathbb{R}, p_2 \in \mathbb{R}, p_3 \in \mathbb{R}$
  if $n > 2$ **then**
    $missingPolygons \leftarrow n$
    $uncoveredPolygon \leftarrow pol$
    $coveredPolygon \leftarrow \phi$
    $polygons \leftarrow []$
    $scalingFactor \leftarrow \sqrt{\frac{p_1 \times area(pol)}{n \times \pi \times \mu^2}}$
    while $\frac{area(uncoveredPolygon)}{area(pol)} >= 0.03$ **do**
      $uncovered2select \leftarrow$ Bigger part of $uncoveredPolygon$
      if $missingPolygons \times \frac{area(uncovered2select)}{area(uncoveredPolygon)} <= 1.5$ **then**
        $polygons.put(uncovered2select)$
        $coveredPolygon \leftarrow coveredPolygon \cup uncovered2slect$
        $missingPolygons \leftarrow missingPolygons - 1$
      else
        $\alpha \leftarrow RandomUniform(\alpha_{min}, \alpha_{max})$
        $\sigma \leftarrow RandomUniform(\sigma_{min}, \sigma_{max})$
        $pol_i \leftarrow$MEANREVERTINGPOLYGON$(\alpha, \sigma, \mu, X_0, \Delta_t)$
        $pol_i \leftarrow$ Multiply each ratio of pol_i by $scalingFactor$
        $pol_i \leftarrow$ Center pol_i randomly into $uncovered2select$
        $pol_i \leftarrow (pol_i - coveredPolygon) \cap pol$
        $pol_i \leftarrow$ Bigger part of pol_i
        $n_i \leftarrow missingPolygons \times \frac{area(pol_i)}{area(uncoveredPolygon)}$
        if $Uniform(0,1) < p_2$ **then**
          $n_i = n_i + missingPolygons \times p_3$
        end if
        $n_i \leftarrow Round(n_1)$
        if $n_i \geq 1$ **then**
          $polygons_i \leftarrow$ RECURSIVEIRREGULARMAP(n_i, pol_i) ▷ Recursive step
          $polygons \leftarrow polygons \cup polygons_i$
          $coveredPolygon \leftarrow coveredPolygon \cup polygons_i$
          $missingPolygons \leftarrow missingPolygons - n_i$
        end if
      end if
      $uncoveredPolygon \leftarrow pol - coveredPolygon$
    end while
    Append interior holes of $coveredPolygon$ to $polygons$
    $coveredArea \leftarrow \bigcup polygons$
    while $length(polygons) < n$ **do**
      Append the smaller polygon to its larger neighbor
    end while
    while $length(polygons) > n$ **do**
      Divide the larger polygon
    end while
  else if n = 1 **then** ▷ Terminating case
    $polygons \leftarrow [pol]$
  else ▷ Terminating case
    $pol_1, pol_2 \leftarrow$ Divide pol in 2
    $polygons \leftarrow [pol_1, pol_2]$
  end if
  return polygons
end function

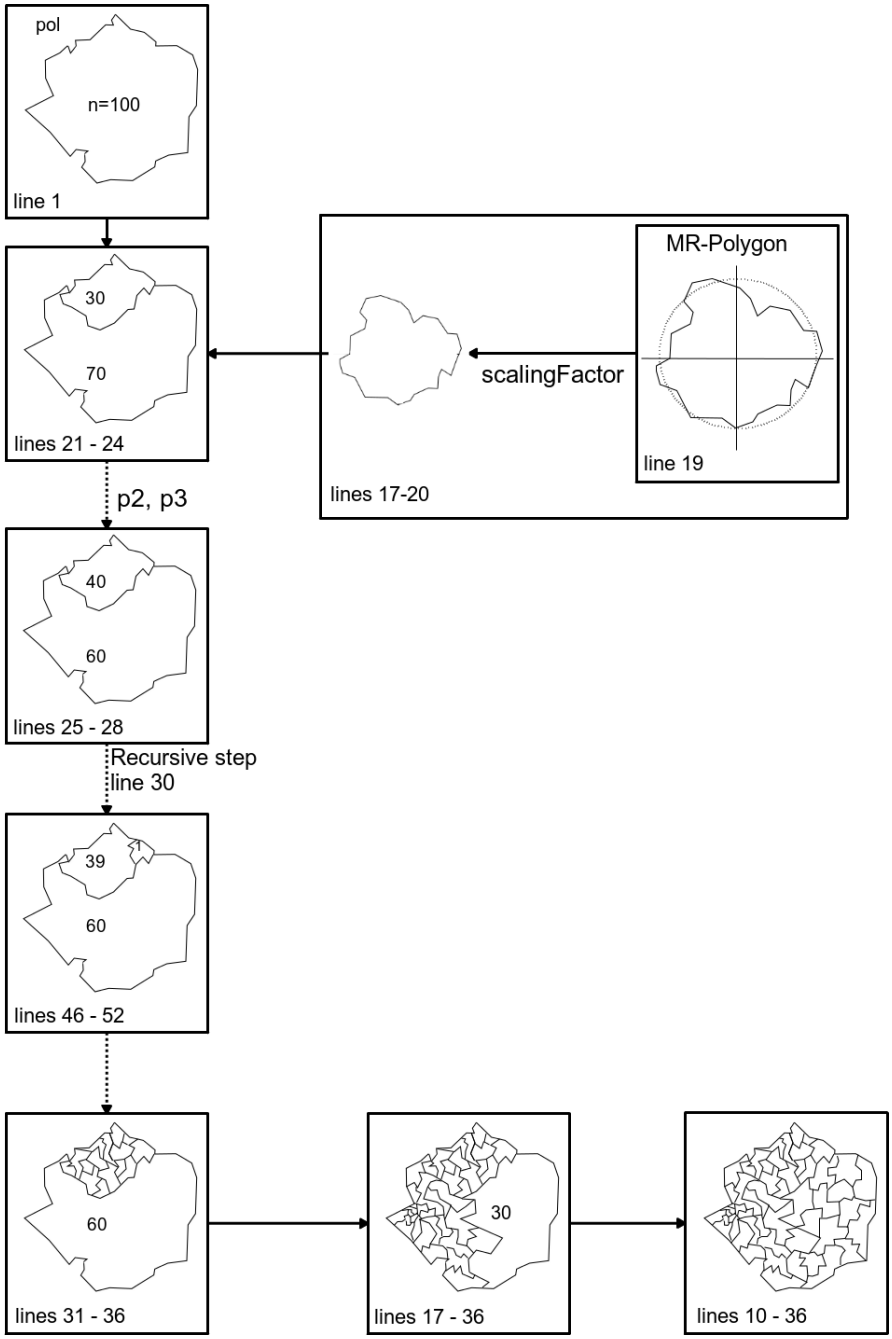
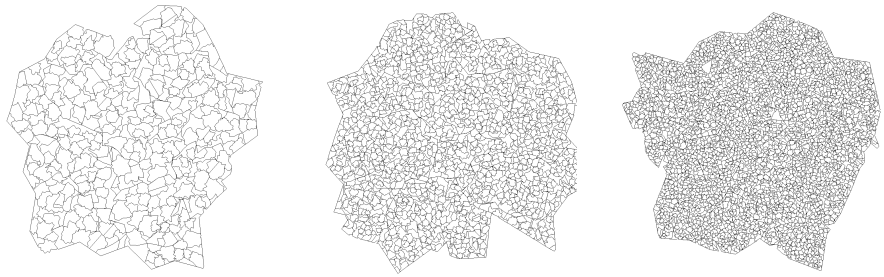


Fig. .10 Diagram of the main steps of the *RI-Maps*

Table .4 Examples of *RI-Maps* of 400, 1,600 and 3,600 polygons using different combinations of parameters



i) $p_1 = 0.1$, $p_2 = 0.0$, $p_3 = 0.0$

ii) $p_1 = 0.1$, $p_2 = 0.1$, $p_3 = 0.5$

iii) $p_1 = 0.010$, $p_2 = 0.050$, $p_3 = 0.315$

$$F(\gamma_j^i) = \frac{\left(\sum_{k\in\theta} \phi_k f_k(\gamma_j^i)\right)}{\sum_{k\in\theta} \phi_k} \tag{.7}$$

$$f_k(\gamma_j^i) = \frac{\sum_{l=1}^{6} \frac{(\sum_{s=i}^{ns} \Psi_k^l(\gamma_j^i)) - ns\Psi_k^l}{ns\Psi_k^l}}{6} \tag{.8}$$

The algorithm starts with an initial random population of 100 genomes to obtain the best four genomes. The subsequent populations are composed of two parts. The first 64 genomes are all of the possible combinations of the last best 4 genomes, and the other 36 genomes are random modifications of those 64 genomes. Because of the computational time required to evaluate equation (.7), only lattices of 400 and 1,600 were used, with an importance of $\phi_{400} = 1$ and $\phi_{1.600} = 2$, respectively. The algorithm reached the optimal value after 13 iterations with $p_1 = 0.010$, $p_2 = 0.050$, and $p_3 = 0.315$.

5 Results

Fig. .11 presents a graphical comparison of the topological characteristics of real *RI-Maps* and Voronoi diagrams. The values for the *RI-Maps* were obtained from 100 instances. The results show that *RI-Maps* have a maximum (M_n) and a minimum (m_n) number of neighbors that are very close to the values found in the real lattices. Regarding the average number of neighbors, both *RI-Maps* and Voronoi diagrams show similar values that are slightly higher than those observed in real lattices; however, because the number of neighbors is an integer value, it can be concluded for all three cases that the average number of neighbors is 6, which verifies the findings by Weaire and Rivier (2009) in irregular lattices. Regarding μ_2, *RI-Maps* are a better approach to simulate the level of disorder found in real lattices. To facilitate the visualization, the values of S are reported as $S * \sqrt{n}$. The results show that *RI-Maps* replicate the values of real lattices at any size, while Voronoi diagrams report higher values that tend to increase with the number of polygons. Last, *RI-Maps* have values of λ_1 that are closer to the values of real lattices, especially for large instances.

Table .6 presents the average and standard deviation of *RI-Maps* under the optimal parameters ($p_1 = 0.010$, $p_2 = 0.050$, $p_3 = 0.315$) found in the previous section. This table completes the topological information on lattices presented in Table .3.

6 Application of *RI-Maps*

In this section, we present an example of the use of *RI-Maps* based on the computational experiments designed by Duque et al. (2011a) to compare the efficiency of the improved AMOEBA algorithm. To present the results, Duque et al. (2011a)

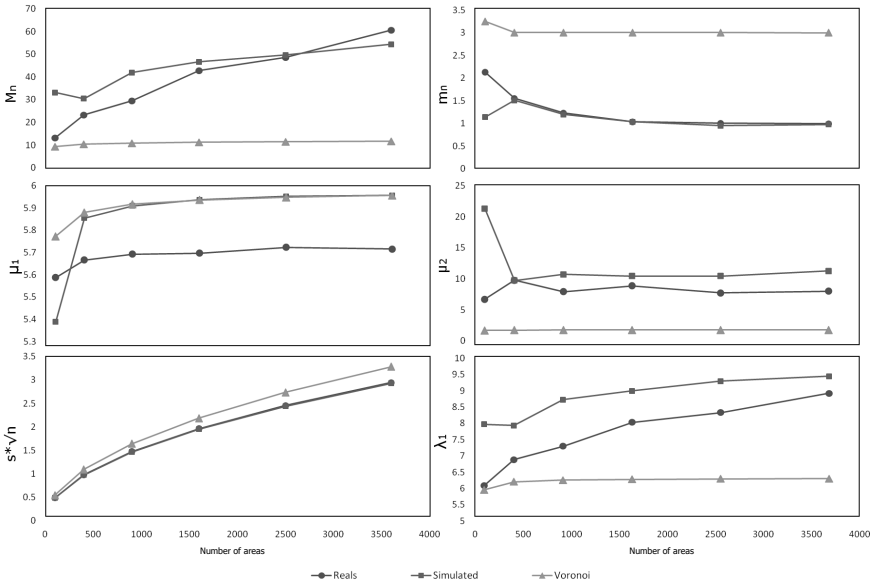


Fig. .11 Comparison of the topological characteristics of real lattices, *RI-Maps* and Voronoi diagrams.

Table .5 Average topological characteristics for *RI-Maps*

	Number of polygons						
	81	100	400	900	1,600	2,500	3,600
M_n	26.500 ±12.765	33.100 ±14.751	30.460 ±13.664	41.870 ±16.035	46.760 ±16.429	49.730 ±16.886	54.396 ±15.156
m_n	1.260 ±0.691	1.140 ±0.513	1.510 ±0.847	1.200 ±0.550	1.040 ±0.374	0.950 ±0.261	0.979 ±0.204
μ_1	5.347 ±0.333	5.388 ±0.304	5.855 ±0.093	5.909 ±0.052	5.937 ±0.036	5.952 ±0.032	5.957 ±0.027
μ_2	17.397 ±13.734	21.313 ±14.729	9.722 ±6.189	10.708 ±3.831	10.426 ±2.277	10.443 ±1.506	11.276 ±1.229
S	5.974 ±0.324	4.879 ±0.219	1.372 ±0.020	0.620 ±0.006	0.353 ±0.002	0.226 ±0.001	0.157 ±0.001
λ_1	7.431 ±0.804	7.969 ±0.808	7.931 ±0.979	8.724 ±0.962	8.993 ±0.960	9.299 ±0.957	9.449 ±0.798

proposed 3 computational experiments; one of them reports the running time of AMOEBA as the number of polygons of regular lattices increases. In this paper, we will run the same algorithm not only for regular lattices but also for real irregular and simulated irregular lattices (*RI-Maps*). First, we want to see whether running the computational experiment on regular lattices only is a good representation of the performance of the algorithm. (Can the conclusions that are obtained for regular lattices be extrapolated to irregular lattices?) Second, we want to see whether the results from using *RI-Maps* are also valid for real irregular maps.

For the experiment that we generated, for each type of lattice, there were 30 instances with 1,600 polygons. For each instance, we generated a spatial process that had four clusters, each using the methodology proposed by Duque et al. (2011a). Last, the instances for real maps were obtained from sampling the same world map that was used in previous sections. Fig. .12 presents the distribution of the running times obtained for each type of lattice, and Table .6 compares the distributions with the two-sided Kolmogorov-Smirnov test (George Marsaglia and Wang, 2003). The null hypothesis of the Kolmogorov-Smirnov test is that the two samples come from the same probability distribution. Regarding the first question, it is clear that using a regular lattice for testing the algorithm can lead to underestimating the execution times of the algorithm. On the other hand, the distribution of the running times obtained for real and *RI-Maps* is statistically equal, which shows the benefits of using *RI-Maps* because it can automatically generate instances without limits on their sizes.

Table .6 Kolmogorov-Smirnov test to compare the distributions of AMOEBA execution times using different lattices

	Regular Lattices	RI-Maps	Real Maps
Regular Lattices	0.00 (p=1)	0.51 (p=$0.0e^{-4}$)	0.61 (p=$2.8e^{-5}$)
RI-Maps	0.51 (p=$0.0e^{-4}$)	0.00 (p=1)	0.19 (p=0.607)
Real Maps	0.61 (p=$2.8e^{-5}$)	0.19 (p=0.607)	0.00 (p=1)

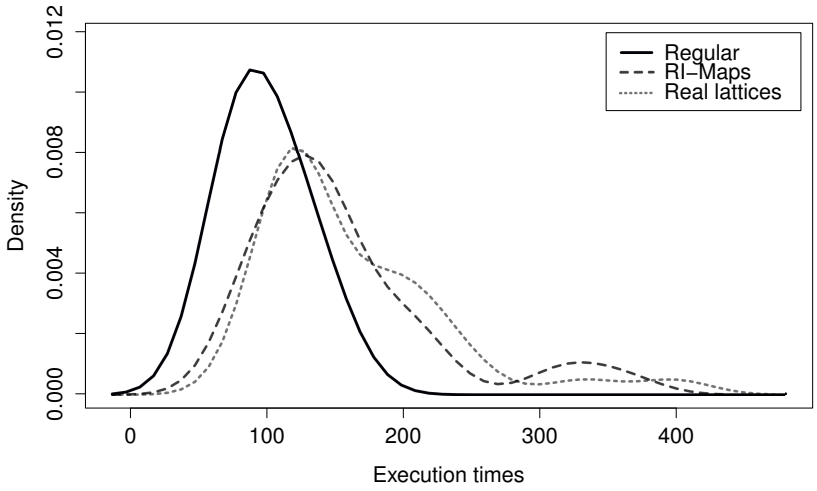


Fig. .12 Execution times of AMOEBA over regular lattices and *RI-Maps* of 1,600

7 Conclusions

This paper introduces an algorithm that combines fractal theory, the theory of stochastic processes and computational geometry for simulating realistic irregular lattices with a predefined number of polygons. The main goal of this contribution is to provide a tool that can be utilized for geocomputational experiments in the fields of exploratory spatial data analysis, spatial statistics and spatial econometrics. This tool will allow theoretical and empirical researchers to create irregular lattices of any size and with topological characteristics that are close to the average characteristics found in irregular lattices around the world.

As shown in the last section, the performance of some geocomputational algorithms can be affected by the topological characteristics of the lattices in which these algorithms are tested. This situation can lead to an unfair comparison of algorithm performances in the literature. With the algorithm proposed in this paper, the differences in the computational performances will not be affected by the topological characteristics of the lattices.

This paper also shows that the topological characteristics of regular lattices (with squared and hexagonal polygons) and Voronoi diagrams (commonly used to emulate irregular lattices) are far from the topological characteristics that are found in real lattices.

References

Aldstadt, J. and Getis, A. (2006). Using AMOEBA to create a apatial weights matrix and identify spatial clusters. *Geographical Analysis*, 38(4):327–343.

Allen, M., Brown, G., and Miles, N. (1995). Measurement of boundary fractal dimensions: review of current techniques. *Powder Technology*, 84(1):1–14.

Anselin, L. (1988). *Spatial econom trics: methods and models*. Kluwer Academic Publishers, Dordrecht, 1 edition.

Anselin, L., Bera, A., Florax, R., and Yoon, M. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26(1):77–104.

Anselin, L., Florax, R., and Rey, S. J. (2004). *Advances in spatial econometrics: methodology, tools and applications*. Springer.

Anselin, L. and Moreno, R. (2003). Properties of tests for spatial error components. *Regional Science and Urban Economics*, 33(5):595–618.

Anselin, L. and Smirnov, O. (1996). Efficient algorithms for constructing proper higher order spatial lag operators. *Journal of Regional Science*, 36(1):67–89.

Aste, T., Szeto, K., and Tam, W. (1996). Statistical properties and shell analysis in random cellular structures. *Physical Review. E, Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 54(5):5482–5492.

Bartlett, M. S. (1975). *Probability, statistics, and time: a collection of essays*. Chapman and Hall, New York, 1 edition.

Batty, M. and Longley, P. (1994). *Fractal cities: A geometry of form and function*. Harcout Brace & Company, London.

Blommestein, H. and Koper, N. (2006). Recursive algorithms for the elimination of redundant paths in spatial lag operators. *Journal of Regional Science*, 32(1):91–111.

Boots, B. (1982). Comments on the use of eigenfunctions to measure structural properties of geographic networks. *Environment and Planning A*.

Boots, B. (1984). Evaluating principal eigenvalues as measures of network structure. *Geographical Analysis*, 16(3):270–275.

Boots, B. (1985). Size effects in the spatial patterning of nonprincipal eigenvectors of planar networks. *Geographical Analysis*, 17(1):74–81.

Chen, Y. (2011). Derivation of the functional relations between fractal dimension of and shape indices of urban form. *Computers, Environment and Urban Systems*.

Church, R. L. (2008). BEAMR: An exact and approximate model for the p-median problem. *Computers & Operations Research*, 35(2):417–426.

Clark, W. (1964). The Concept of Shape in Geography. *American Geographical Society*, 54(4):561–572.

Cole, J. (1964). Study of major and minor civil divisions in political geography. In *20th International Geographical Congress*, Sheffield. University of Nottingham.

Coxeter, H. S. M. (1974). *Regular complex polytopes*. CUP Archive.

Douglas, D. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10(2):112–122.

Duque, J., Aldstadt, J., Velasquez, E., Franco, J., and Betancourt, A. (2011a). A computationally efficient method for delineating irregularly shaped spatial clusters. *Journal of Geographical Systems*, 13:355–372.

Duque, J. C., Anselin, L., and Rey, S. J. (2012). the max-p-regions problem*. *Journal of Regional Science*, 52(3):397–419.

Duque, J. C., Dev, B., Betancourt, A., and Franco, J. L. (2011b). ClusterPy: {Library} of spatially constrained clustering algorithms, {Version} 0.9.9.

Elhorst, J. P. (2003). Specification and estimation of spatial panel data models. *International Regional Science Review*, 26(3):244–268.

Frankhauser, P. (1998). The fractal approach: A new tool for the spatial analysis of urban agglomerations. *New Methodological Approaches in the Social Sciences*, 10(1):205–240.

Garrison, W. (1964). Factor-analytic study of the connectivity of a transportation network. *Papers in Regional Science*.

George Marsaglia, W. W. T. and Wang, J. (2003). Evaluating Kolmogorov s Distribution. *Journal of Statistical Software*, 8(18):1–4.

Ghyka, M. (2004). *The Geometry of art and life*. Kessinger Publishing.

Gibbs, J. (1961). A method for comparing the spatial shapes of urban units. In *Urban Research Methods*, pages 96–106. D. Van Nostrand Company, Inc, Princeton.

Gould, P. (1967). On the geographical interpretation of eigenvalues. *Transactions of the Institute of British Geographers*, 42(42):53–86.

Griffith, D. (1987). Toward a theory of spatial statistics: Another step forward. *Geographical Analysis*, 19(1):69–82.

Grunbaum, B. and Shephard, G. C. (2011). *Tilings and patterns*. Dover Publications.

Haggett, P. (1977). *Locational analysis in human geography*. John Wiley & Sons, London.

Haining, R. (2010). The nature of georeferenced data. In Fischer, M. M. and Getis, A., editors, *Handbook of Applied Spatial Analysis*, pages 197–217. Springer Berlin Heidelberg, Berlin, Heidelberg.

Hijmans, R., Guarino, L., Jarvis, A., O'Brien, R., Mathur, P., Bussink, C., Cruz, M., Barrantes, I., and Rojas, E. (2011). Diva-GIS.

Hooper, P. and Hewings, G. (1981). Some properties of space-time processes. *Geographical Analysis*, 13(3):203–223.

Horton, R. (1932). Drainage basin characteristics. *Transactions of the American Geophysical Union*, 13(1):350–361.

Johnson, D. L. (2001). *Symmetries*. Springer Verlag, London, 1 edition.

Kindratenko, V. and Treiger, B. (1996). Chemometrical approach to the determination of the fractal dimension (s) of real objects. *Chemometrics and Intelligent Laboratory Systems*, 34:103–108.

Le Caer, G. and Delannay, R. (1993). The administrative divisions of mainland France as 2D random cellular structures. *Journal de Physique I*, 3(8):1777–1800.

Le Caer, G. and Delannay, R. (1995). Topological models of 2D fractal cellular structures. *Journal de Physique I*, 5(11):1417–1429.

Mandelbrot, B. B. (1982). *The fractal geometry of nature*. W.H. Freeman.

Mao, X. (1997). *Stochastic differential equations and applications*. Horwood publishing limited, Chichester, 1 edition.

Miller, V. (1953). *A quantitative geomorphic study of drainage basin characteristics in the clinch mountain area Virginia and Tenesse.* Department of Geology, Columbia University.

Mur Lacambra, J. (1992). Contrastes de autocorrelación espacial: Un estudio de Monte Carlo. *Estadística Española*, 34(130):285–308.

Murray, A. T. and O'Kelly, M. E. (2002). Assessing representation error in point-based coverage modeling. *Journal of Geographical Systems*, 4(2):171–191.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70(349):120–126.

Pace, R. and LeSage, J. P. (2004). Chebyshev approximation of log-determinants of spatial weight matrices. *Computational Statistics & Data Analysis*, 45(2):179–196.

Penrose, R. (1974). The Rôle of Aesthetics in Pure and Applied Mathematical Research. *Journal of The Institute of Mathematics and its Aplications*, 10:266–271.

Peshkin, M., Strandburg, K., and Rivier, N. (1991). Entropic predictions for cellular networks. *Physical Review Letters*, 67(13):1803–1806.

Radin, C. (1993). Symmetry of tilings of the plane. *Bullettin of the American Mathematical Society*, 29(2):213–217.

ReVelle, C. and Swain, R. (1970). Central facilities location. *Geographical Analysis*, 2(1):30–42.

Richardson, L. (1961). The problem of contiguity: An appendix of statistics of deadly quarrels. *General Systems Yearbook*, 6(13):139–187.

Ross, I. C. and Harary, F. (1952). On the determination of redundancies in socioeconometric chains. *Psychometrika*, 17(2):195–208.

Smirnov, O. and Anselin, L. (2001). Fast maximum likelihood estimation of very large spatial autoregressive models: A characteristic polynomial approach. *Computational Statistics & Data Analysis*, 35(3):301–319.

Stoddart, D. (1965). The shape of atolls. *Marine Geology*, 3(5):269–283.

Tilley, R. (2006). *Crystals and crystal structures*. Wiley, England, 1 edition.

Tinkler, K. (1972). The physical interpretation of eigenfunctions of dichotomous matrices. *Transactions of the Institute of British Geographers*, 55(55):17–46.

Weaire, D. and Rivier, N. (2009). Soap, cells and statistics random patterns in two dimensions. *Contemporary Physics*, 50(1):199–239.

Weeitty, A. (1969). *On the form of drainage basins*. Dept. of Geography, Pennsylvania State University, Pennsylvania, 1 edition.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 41(3):434–449.

An algorithmic approach for simulating realistic irregular lattices*

Juan C. Duque[1]
Alejandro Betancourt[2]
Freddy Marin[3]

July 15 2013

Resumen/abstract
There is a wide variety of computational experiments, or statistical simulations, in which regional scientists require regular and irregular lattices with a predefined number of polygons. While most commercial and free GIS software offer the possibility of generating regular lattices of any size, the generation of instances of irregular lattices is not a straightforward task. The most common strategy in this case is to find a real map that matches as closely as possible the required number of polygons. This practice is usually conducted without considering whether the topological characteristics of the selected map are close to those for an "average" map sampled in different parts of the world.
In this paper, we propose an algorithm, *RI-Maps*, that combines fractal theory, stochastic calculus and computational geometry for simulating realistic irregular lattices with a predefined number of polygons. The irregular lattices generated with *RI-Maps* have guaranteed consistency in their topological characteristics, which reduces the potential distortions in the computational or statistical results due to an inappropriate selection of the lattices.

Palabras claves/Keywords
RI-Maps, MR-Polygons, Regional Science, Lattices, Computation, Experiment

JEL Clasification
C02, C63

[1] Research in Spatial Economics (RISE-group), Department of Economics,
EAFIT University,
e-mail: jduquec1@eafit.edu.co
Phone: (574) 261-9354, Fax: (574) 261-9294

[2] Research in Spatial Economics (RISE-group), Department of Economics,
EAFIT University

[3] Department of Fundamental Sciences,
EAFIT University

* The authors wish to thank Colciencias (Departamento Administrativo de Ciencia y Tecnología e Inovación) for their financial support under the program "Jovenes Investigadores." The authors also thank the Cyberinfrastructure Service for High Performance Computing, "Apolo", at EAFIT University, for allowing us to run our computational experiments in their supercomputer. The usual disclaimer applies.

1 Introduction

The complexity of computational experimentation in regional science has drastically increased in recent decades. Regional scientists are constantly developing more efficient methods, taking advantage of modern computational resources and geocomputational tools, to solve larger problem instances, generate faster solutions or approach asymptotics. The literature has many examples: The first formulation of the p-median problem provides a numerical example that required 1.51 minutes to optimally locate four facilities in a 10-node network (ReVelle and Swain, 1970); three decades later, Church (2008) located five facilities in a 500-node network in 1.68 minutes. As noted by Anselin et al. (2004), spatial econometrics has also benefitted from computational advances; the computation of the determinant required for maximum likelihood estimation of a spatial autoregressive model proposed by Ord (1975) was feasible to apply for data sets not larger than 1,000 observations. Later, Pace and LeSage (2004) proposed a procedure that was capable of computing this determinant for over a million observations. According to Blommestein and Koper (2006), one of the first algorithms for constructing higher-order spatial lag operators, which was devised by Ross and Harary (1952), required 8,000 seconds (approximate computation time) to calculate the sixth-order contiguity matrix in a 100x100 regular lattice. Anselin and Smirnov (1996) proposes new algorithms that are capable of computing a sixth-order contiguity matrix for the 3,111 U.S. contiguous counties in less than a second.

An important aspect when conducting computational experiments in regional science is the selection of the way that the spatial phenomena are represented or conceptualized. This aspect is of special relevance when using a discrete representation of continuous space, such as polygons (Haining, 2010). This representation can be accomplished in two ways: regular or irregular lattices; the use of one or the other could cause important differences in the computational times, solution qualities or statistical properties. We suggest four examples, as follows: (1) The method proposed by Duque et al. (2011a) for running the AMOEBA algorithm (Aldstadt and Getis, 2006) requires an average time of 109 seconds to delimit four spatial clusters on a regular lattice with 1,849 polygons. This time rises to 229 seconds on an irregular lattice with the same number of polygons. (2) For the location set covering problem, Murray and O'Kelly (2002) concluded that the spatial configuration, number of needed facilities, computational requirements and coverage error all varied significantly as the spatial representation was modified. (3) Elhorst (2003) warns that the parameters of the random effects spatial error and spatial lag model might not be an appropriate specification when the observations are taken from irregular lattices.[1] (4) Anselin and Moreno (2003) finds that the use of regular or irregular lattice affects the performance of test statistics against alternatives of the spatial error components form.

However, returning to the tendency toward the design of computational experiments with large instances, there is an important difference between generating large instances of regular and irregular lattices. On the one hand, regular lattices are easy to generate, and there is no restriction on the maximum number of polygons. On the other hand, instances of irregular lattices are usually made by sampling real maps. Table .1 shows some examples of this practice.

The generation of large instances of irregular lattices has several complications that are of special interest in this paper. First, the size is limited to the sizes of the available real lattices. Second, the possibility of generating a large number of different instances of a given size is also limited (e.g., generate 1,000 instances of irregular lattices with 3,000 polygons). Third, as shown in Fig. 1, the topological characteristics of irregular lattices built from real maps change drastically, depending on the region from where they are sampled, which could bias the results of the computational experiments.[2]

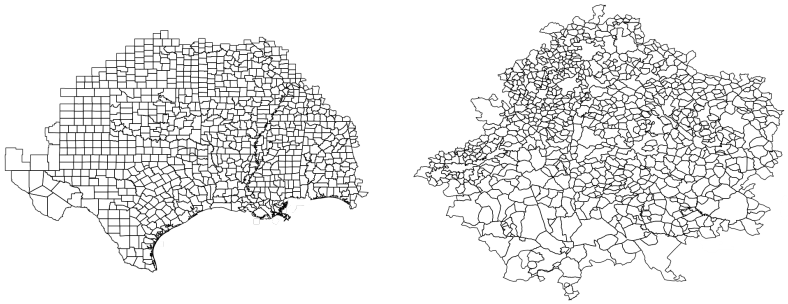
This paper seeks to contribute to the field of computational experiment design in regional science by proposing a scalable recursive algorithm (*RI-Maps*), which combines concepts from

[1] See also Anselin (1988), pg 51

[2] Later in this paper, we show that the topological characteristics of Voronoi diagrams are far from those for an "average" map sampled in different parts of the world

Table .1 Annotated chronological listing of studies that use irregular lattices generated by sampling real maps.

Study	Purpose	Source of irregular lattices
Mur Lacambra (1992)	Compares different methods to detect spatial autocorrelation.	Spain provinces in 1985 (sizes 14 and 48 polygons)
Anselin et al. (1996)	Performance of a diagnostic test for spatial dependence.	COROP and economic geographic regions in The Netherlands (sizes 40 and 81, respectively)
Smirnov and Anselin (2001)	Performance of a new method for evaluating the Jacobian term.	921 counties (Kreise) for Germany; 3,107 U.S. continental counties; 3,140 U.S. counties, and 29,762 U.S. postal zip codes.
Anselin and Moreno (2003)	Extend the knowledge about the properties of spatial correlation tests, especially in empirical applications.	Spatial grouping of Western U.S. counties for dimensions 46, 80, 124, 264, 413 and 1,013
Duque et al. (2012)	Performance of an algorithm for spatial clustering (the max-p-regions model)	Sacramento census tracks (403), Colombian municipalities (1,068) and U.S. census tracks (3,085).



(a) United States (b) Spain

Fig. .1 Examples of two instances of 900 irregular polygons.

stochastic calculus (mean reversing processes), fractal theory and computational geometry to generate instances of irregular lattices that have an unlimited size and topological characteristics that are a good representation of the irregular lattices sampled from around the world. The use of instances obtained from *RI-Maps* will guarantee that differences in the results in the computational experiment will not come from differences in the topological characteristics of the used lattices and will also make it easier to generate the unlimited number of large instances.[3]

The remainder of this paper is organized as follows: Section 2 introduces the basic definitions of the polygons and lattices and proposes a consensus taxonomy of the lattices. Section 3 presents a set of indicators that are used to characterize the topological characteristics of a lattice and shows the topological differences between regular and irregular lattices. Section 4 presents the algorithm

[3] This algorithm as well as all of the instances of *RI-Maps* used in this article will be publicly available for the academic community after its publication (Duque et al., 2011b).

for generating irregular lattices. Section 5 evaluates the capacity of the algorithm to generate realistic irregular lattices. Finally, Section 6 presents the conclusions.[4]

2 Conceptualizing polygons and lattices

A polygon is a plane figure enclosed by a set of finite straight line segments. Polygons can be categorized according to their boundaries, convexity and symmetry properties, as follows:

i) Boundary: A Polygons is *Simple* when it is formed by a single plain figure with no holes, and it is *Complex* when it contains holes or multiple parts.[5]
ii) Convexity: In a *convex* polygon, every pair of points can be connected by a straight line without crossing its boundary. A *Concave* polygon is simple and non-convex.
iii) Symmetry: A *regular* polygon has all of its angles of equal magnitude and all of its sides of equal length. A non-regular polygon is also called *irregular* (Johnson, 2001; Coxeter, 1974).

A lattice is a set of polygons of any type, with no gaps and no overlaps, that covers a subspace or the entire space. Next, a more formal definition: A lattice is the division of a subspace $S \subseteq R^n$ into k subsets $i \subseteq S$ such that $\cup_i = S$ and $\cap_i = \phi$, where ϕ is the empty set of R^n (Grunbaum and Shephard, 2011)[6]. There exist different taxonomies of lattices depending on the field of study. In an attempt to unify these taxonomies, a consensus lattice taxonomy is presented in Fig. .2. This taxonomy classifies lattices according to the shapes of their polygons, the spatial relationships between them, and the use, or not, of symmetric relationships to construct the lattice:[7]

i) According to the variety of the shapes of the polygons that form the lattice: **homomorphisms** are lattices that are formed by polygons that have the same shape, and **polymorphisms** are lattices that are formed by polygons that have different shapes.
ii) According to the regularity of the polygons that form the lattice and the way in which they intersect, each vertex:[8] **Regular**, lattices formed by regular polygons in which all of the vertexes join the same arrangement of polygons (Tilley, 2006); **Semi-regular**, when the polygons are regular but there are different configurations of vertexes; and **Irregular** otherwise (Ghyka, 2004).
iii) According to the existence of symmetric relationships within the lattice:[9] **Symmetric**, when the lattice implies the presence of at least one symmetric relationship; and **Asymmetric** otherwise.
iv) According to the symmetric relationship of translation: A lattice is **periodic** if and only if it implies the use of translation without rotation or reflection; it is **Aperiodic** otherwise (Tilley, 2006).

Table .2 shows an example of each category of this consensus taxonomy.

---

[4] A dataset of 700 RI-Maps is available at "http://www.rise-group.org/section/Research/Publication/AnAlgorithmicAproach/"

[5] Complex polygons do not refer to polygons that exist in the Hilbert plane (Coxeter, 1974).

[6] This paper focuses exclusively on bidimensional lattices (i.e., $n = 2$)

[7] An alternative category is proposed for lattices formed by fractal polygons that are informally defined by Mandelbrot (1982) as rough fragmented geometric shapes that could be infinitely divided into scalable parts.

[8] considering the vertexes to be all of the points of the lattice that intersect three or more polygons)

[9] There are three types of symmetrical relationships: *Translation*, when the lattice is formed by translating a subset of polygons; *reflection*, when there are axes of reflection in the lattice; and *rotation*, when it is possible to obtain the same lattice after a rotation process of less than 2π (Radin, 1993).

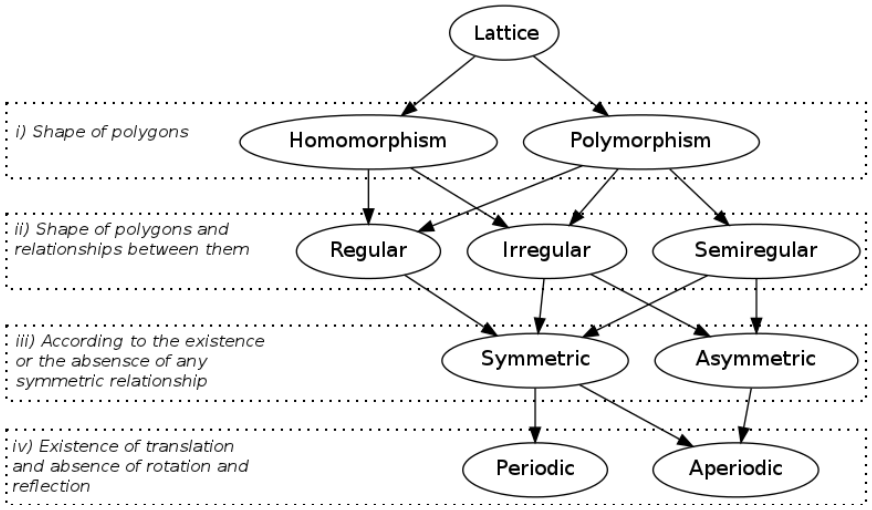
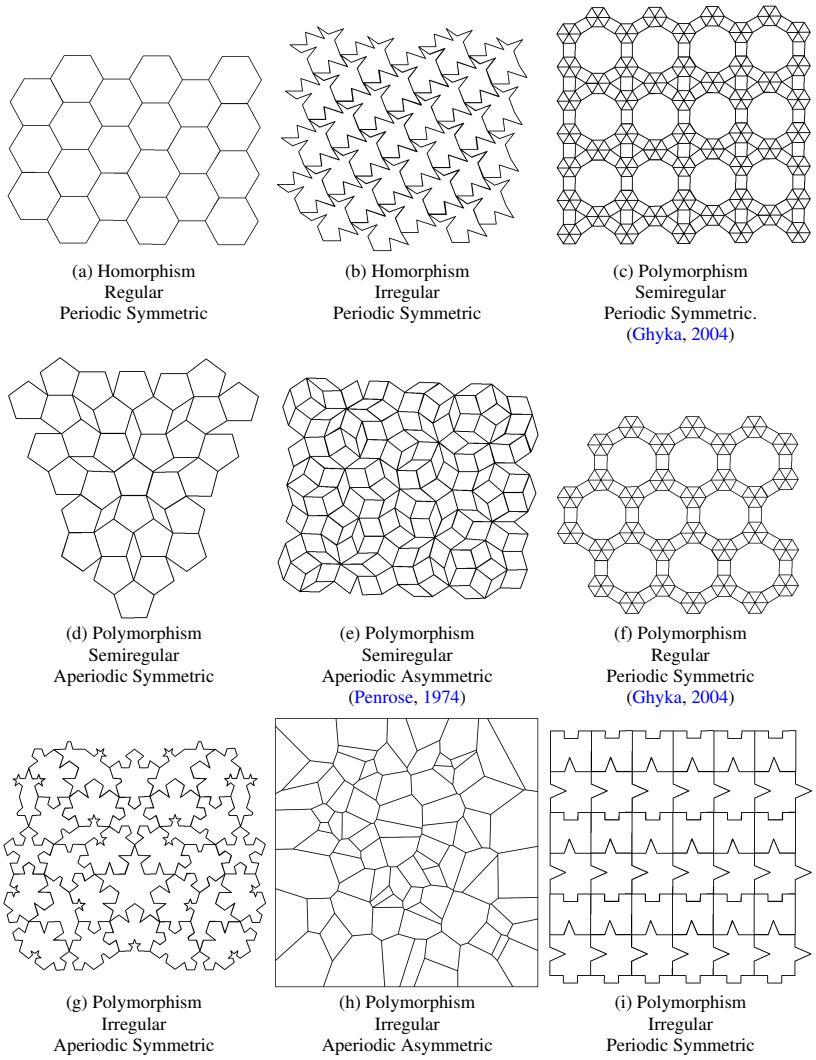


Fig. .2 Consensus taxonomy of lattices

The topological characteristics of lattices are usually summarized through the properties of the sparse matrix that represent the neighboring relationships between the polygons in the map, the so-called W matrix (Gould, 1967; Boots, 1982; Le Caer and Delannay, 1995; Aste et al., 1996; Peshkin et al., 1991).[10] This paper uses six indicators; the first three indicators are self-explained: the maximum (M_n), minimum (m_n) and average (μ_1) number of neighbors per polygon. The fourth indicator, Sparseness *(S)*, is the percentage of non-zero elements of the W matrix. The fifth indicator is the first eigenvalue of the W matrix (λ_1), which is an algebraic tool that is frequently used in graph theory (Garrison, 1964; Tinkler, 1972) and regional science (Gould, 1967; Boots, 1982, 1984, 1985) to summarize different aspects of the W matrix. The last indicator is the variance of the number of neighbors per polygon (μ_2), which measures the spatial disorder of a lattice.

Within the field of regional science, lattices are frequently used with two purposes: First, real lattices can be used to study real phenomena, e.g., to analyze spatial patterns, confirm spatial relationships between variables, and detect spatio-temporal regimes within a spatial panel, among others. Second, lattices can be used to evaluate the behavior of statistical tests (Anselin and Moreno, 2003; Mur Lacambra, 1992); algorithms (Duque et al., 2011a); and topological characteristics of lattices (Aste et al., 1996; Le Caer and Delannay, 1995, 1993). In these cases, it is necessary to use sets of lattices that satisfy some requirements imposed by the regional scientist, e.g., the number of polygons, regularity or irregularity of the polygons and the number of instances. To accomplish this goal, it is a common approach to use a geographical base for real or simulated data *polymorphism irregular aperiodic asymmetric* (e.g., real lattices and Voronoi diagrams) or *homomorphism regular periodic symmetric* (e.g., regular lattices). The next sections focus on the second use of lattices.

[10] See Anselin (1988) for more information about this matrix.



(a) Homorphism
Regular
Periodic Symmetric

(b) Homorphism
Irregular
Periodic Symmetric

(c) Polymorphism
Semiregular
Periodic Symmetric.
(Ghyka, 2004)

(d) Polymorphism
Semiregular
Aperiodic Symmetric

(e) Polymorphism
Semiregular
Aperiodic Asymmetric
(Penrose, 1974)

(f) Polymorphism
Regular
Periodic Symmetric
(Ghyka, 2004)

(g) Polymorphism
Irregular
Aperiodic Symmetric

(h) Polymorphism
Irregular
Aperiodic Asymmetric

(i) Polymorphism
Irregular
Periodic Symmetric

Table .2 Example Lattices

3 Topological characteristics of regular and irregular lattices

As stated above, regional scientists have the option of using regular or irregular lattices in their computational experiments. However, this section will show that there are important topological differences between these types of lattices.

Real lattices have topological characteristics that vary substantially from location to location. As an example, Fig. .3 presents the topological characteristics of lattices of different sizes (100, 400 and 900 polygons) sampled in Spain and the United States. Each box-plot summarizes 1,000 instances. Important differences emerge between these two places: Spanish polygons tend to have

more neighbors, are more disordered, and their first eigenvalues are higher in mean and variance. These differences in the topological characteristics have direct repercussions on the performance of algorithms whose complexity depends on the neighboring structure (Aldstadt and Getis, 2006; Duque et al., 2011a).

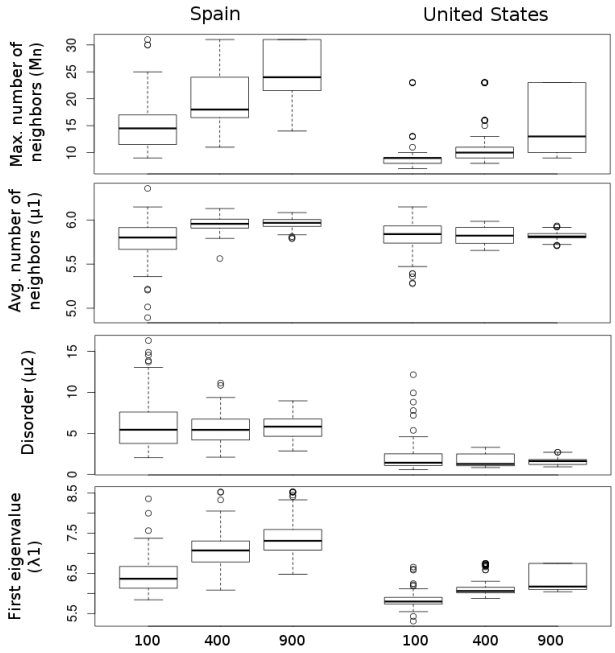


Fig. .3 Topological differences of lattices from Spain and the United States

Regular lattices and Voronoi diagrams are also commonly used for computational experiments because they are easy to generate, there is no restriction on the size of the instances (the number of polygons in the map), and their over-simplified structure allows for some mathematical simplifications or reductions (Whittle, 1954; Bartlett, 1975; Griffith, 1987). However, the topological characteristics of these lattices are substantially different from real, irregular lattices. These differences can lead to biased results in theoretical and empirical experiments, e.g., spatial stationarity in STARMA models (Hooper and Hewings, 1981), improper conclusions about the properties of the power and sample sizes in hypothesis testing (Anselin and Moreno, 2003; Mur Lacambra, 1992), and the over-qualification of the computational efficiency of the algorithms (Duque et al., 2011a; Aldstadt and Getis, 2006), among others. Table .3 shows the topological differences between real maps, two types of regular lattices and Voronoi diagrams.

To illustrate the magnitude of these differences, we calculated the topological indicators (M_n, m_n, μ_1, μ_2, S and λ_1) for six thousand lattices of different sizes (1,000 instances each of 100, 400, 900, 1,600, 2,500, and 3,600 polygons) that were sampled around the world at the smallest administrative division available in Hijmans et al. (2011). As an example, Fig. .4 shows seven of those instances. These real instances are then compared to regular lattices that have square and hexagonal polygons and Voronoi diagrams.[11] To avoid the boundary effect on M_n, m_n, μ_1 and μ_2,

[11] Each one of the six-thousand instances of Voronoi diagrams come from uniformly distributed points.

the bordering polygons are only considered to be neighbors of interior polygons. Last, S and λ_1 are calculated using all of the polygons. Table .3 shows that regular lattices are not capable of emulating the topological characteristics of real lattices in any of the indicators: $\mu_2 = 0$ and M_n, m_n, $\mu_1 = 4$ and 6 (for squares and hexagons, respectively) are values that are far from those of real lattices. The values obtained for λ_1 and S indicate that regular lattices of hexagons are more connected than real lattices, while regular lattices of squares are less connected than real lattices. With regard to Voronoi diagrams, M_n and m_n indicate that they are not capable of generating atypically connected polygons. The values of μ_1 are close to real lattices. Finally, Voronoi diagrams are more ordered than real lattices, with values of μ_2 close to 1.7, while real lattices report values of μ_2 that are close to 8.

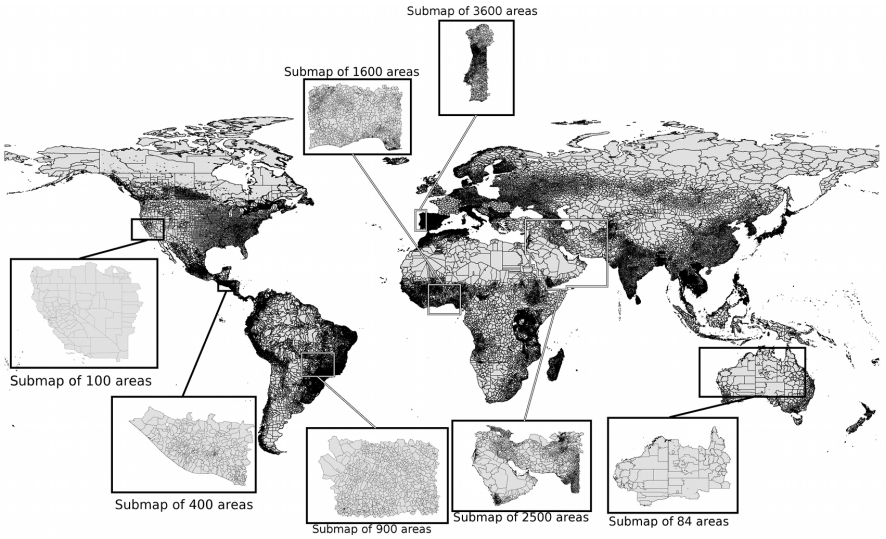


Fig. .4 Base map and example of a random irregular lattice obtained from it.

4 RI-Maps: An algorithm for generating realistic irregular lattices

This section is divided into two parts. The first part introduces an algorithm that generates irregular polygons based on a mean reverting process in polar coordinates, and the second part proposes a novel method to create polymorphic irregular aperiodic lattices with topological characteristics that are similar of those from real lattices.

Table .3 Average topological characteristics for real maps, regular lattices and Voronoi diagrams

		Number of polygons						
		81	100	400	900	1,600	2,500	3,600
Real lattices	M_n	12.28 ±7.52	13.22 ±9.90	23.22 ±29.89	29.55 ±36.27	42.77 ±50.59	48.53 ±55.07	60.64 ±64.58
	m_n	2.33 ±1.11	2.13 ±1.06	1.55 ±0.86	1.23 ±0.60	1.04 ±0.22	1.01 ±0.10	1.00 ±0.00
	μ_1	5.57 ±0.65	5.59 ±0.61	5.67 ±0.49	5.69 ±0.45	5.70 ±0.46	5.72 ±0.37	5.72 ±0.37
	μ_2	5.85 ±13.85	6.72 ±22.58	9.76 ±28.79	7.90 ±15.35	8.85 ±12.82	7.73 ±9.39	8.00 ±8.11
	S	5.98 ±0.51	4.91 ±0.43	1.30 ±0.11	0.58 ±0.046	0.33 ±0.02	0.21 ±0.01	0.15 ±0.01
	λ_1	5.96 ±0.53	6.09 ±0.65	6.89 ±1.52	7.30 ±1.82	8.03 ±2.42	8.33 ±2.62	8.92 ±3.02
Reg. Lattice *(Squares)*	M_n	4	4	4	4	4	4	4
	m_n	4	4	4	4	4	4	4
	μ_1	4	4	4	4	4	4	4
	μ_2	0	0	0	0	0	0	0
	S	4.44	3.64	0.95	0.43	0.24	0.16	0.11
	λ_1	3.80	3.84	3.96	3.98	3.99	3.99	3.99
Reg. Lattice *(Hexagons)*	M_n	6	6	6	6	6	6	6
	m_n	6	6	6	6	6	6	6
	μ_1	6	6	6	6	6	6	6
	μ_2	0	0	0	0	0	0	0
	S	6.30	5.19	1.39	0.64	0.36	0.23	0.16
	λ_1	5.55	5.62	5.88	5.94	5.96	5.97	5.98
Voronoi Diagrams	M_n	9.15 ±0.77	9.36 ±0.79	10.37 ±0.75	10.90 ±0.74	11.26 ±0.70	11.49 ±0.67	11.71 ±0.68
	m_n	3.36 ±0.48	3.26 ±0.44	3.00 ±0.03	3.00 ±0.00	3.00 ±0.00	3.00 ±0.00	3.00 ±0.03
	μ_1	5.75 ±0.07	5.77 ±0.05	5.88 ±0.02	5.92 ±0.01	5.94 ±0.00	5.95 ±0.00	5.96 ±0.00
	μ_2	1.68 ±0.31	1.70 ±0.27	1.75 ±0.13	1.76 ±0.09	1.76 ±0.07	1.77 ±0.05	1.77 ±0.04
	S	6.67 ±0.08	5.47 ±0.05	1.44 ±0.00	0.65 ±0.00	0.37 ±0.00	0.24 ±0.00	0.17 ±0.00
	λ_1	5.88 ±0.05	5.96 ±0.05	6.20 ±0.03	6.26 ±0.02	6.28 ±0.02	6.29 ±0.02	6.30 ±0.02

4.1 Mean reverting polygons (MR-Polygons)

The problem of characterizing the shape of irregular polygons is commonly addressed in two ways, that is, evaluating its similitude with a circle (Haggett, 1977) or describing its boundary roughness through its fractal dimension (Batty and Lon-

gley, 1994; Frankhauser, 1998).[12] In this paper, we apply both concepts in different stages during the creation of a polygon: the similitude with a circle to guide a mean reverting process in polar coordinates, and the fractal dimension to parameterize the mean reverting process.

4.1.1 Mean reverting process in polar coordinates

Different indexes are used to compare irregular polygons with a circle: Elongation ratio (Weeitty, 1969), form ratio (Horton, 1932), circularity ratio (Miller, 1953), compactness ratio (Gibbs, 1961; Cole, 1964; Richardson, 1961), ellipticity index (Stoddart, 1965) and the radial shape index (Clark, 1964). As Chen (2011) states, all of these indexes are based on comparisons between the irregular polygon and its area-equivalent circle. Under this relationship, an irregular polygon can be conceptualized as an irregular boundary with random variations following a circle, which lead us to use a mean reverting process in polar coordinates to create irregular polygons. A mean reverting process is a stochastic process that takes values that follow a long-term tendency in the presence of short-term variations. Formally, the process x at the moment t is the solution of the stochastic differential equation (.1), where μ is the long-term tendency, α is the mean reversion speed, σ is the gain in the diffusion term, $x(t_0)$ is the value of the process when $t = 0$ and $\{B_t\}_{t\geq 0}$ is an unidimensional Brownian (Mao, 1997). Equation (.2) shows the general solution; however, for practical purposes, hereafter we use the Euler discretization method, which is given by equation (.3), where ε_t is white noise.

$$dX_t = \alpha(\mu - X_t)d_t + \sigma dB_t \tag{.1}$$

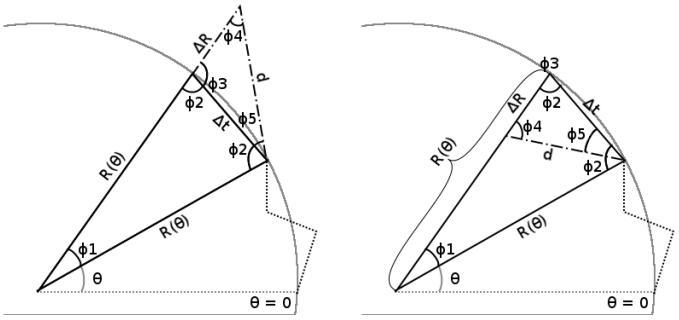
$$x(t) = e^{-\alpha(t-t_0)}\left(x(t_0) + \int_{t_0}^{t} e^{\alpha(s-s_0)}\alpha\mu ds + \int_{t_0}^{t} e^{\alpha(s-s_0)}\sigma dB(s)\right), \tag{.2}$$

$$X_t = X_{t-1} + \alpha(\mu - X_{t-1})\Delta_t + \sigma\sqrt{\Delta_t}\varepsilon_t \tag{.3}$$

Algorithm 1 presents the procedure for generating an irregular polygon P in polar coordinates using, as a data generator, a mean reverting process (X_t). This algorithm guarantees that the distance between two points in X_t, following the process X_t, is equal to the distance between the same two points in P when following the process P counterclockwise. The purpose of this equivalence is to preserve the fractal dimension of X_t in P. The angles Δ_R and ϕ_1 in algorithm 1 are the result of solving the geometric problem in Fig. .5. These two angles are used in equation (.4) to establish the location of the next point in P. The points of P are denoted as P_θ, with θ between 0 and 2π.

$$P_{\theta+\phi_1} = \begin{cases} P_\theta + \Delta_R & \text{if } X_{t+\Delta_t} \geq Xt \\ P_\theta - \Delta_R & \text{if } X_{t+\Delta_t} < Xt. \end{cases} \tag{.4}$$

[12] Chen (2011) established a relationship between these two approaches.



(a) $X_t \geq X_{t-\Delta_t}$ (b) $X_t < X_{t-\Delta_t}$

Fig. .5 Geometric problem to preserve the length and the fractal dimension of the mean reverting process when it is used to create an irregular polygon

Algorithm 1 MR-Polygon: Mean Reverting Polygon

function MEANREVERTINGPOLYGON($\alpha, \sigma, \mu, X_0, \Delta_t$)
  $X_{t-\Delta_t} = X_0$ ▷ Initial point of the mean reverting process
  $P = [(0, X_0)]$ ▷ Irregular polygon in polar coordinates
  while $\theta < 2\pi$ **do**
    $\varepsilon_t \leftarrow RandomNormal(0,1)$
    $X_t = X_{t-\Delta_t} + \alpha(\mu - X_{t-\Delta_t})\Delta_t + \sigma\sqrt{\Delta_t}\varepsilon_t$
    $d \leftarrow distance(X_t, X_{t-\Delta_t})$
    $R_\theta \leftarrow$ Last radius of the irregular polygon
    $\phi_1 = \arccos\left(\frac{2R_\theta - \Delta_t^2}{2R_\theta^2}\right)$
    if $Xt \geq X_{t-\Delta_t}$ **then**
      $\Delta_R = d\left(\cos\left(\arcsin\left(\frac{\Delta_t}{d}\cos\left(\frac{\phi_1}{2}\right)\right) - \sin\left(\frac{\phi_1}{2}\right)\frac{\Delta_t}{d}\right)\right)$
    else
      $\Delta_R = -d\left(\cos\left(\arcsin\left(\frac{\Delta_t}{d}\cos\left(\frac{\phi_1}{2}\right)\right) + \sin\left(\frac{\phi_1}{2}\right)\frac{\Delta_t}{d}\right)\right)$
    end if
    $R_{\theta+\phi_1} = R_\theta + \Delta_R$
    Add $(\theta + \phi_1, R_{\theta+\phi_1})$ to P
    Increase θ in ϕ_1
  end while
  Replace last point of P to $(0, X_0)$
  return P
end function

Because the process P depends on the parameters α, μ and σ, it is worthwhile to clarify their effect on the shape of polygon P: α is the speed at which the process reverts to the circle with radius μ, and σ is the scaling factor of the irregularity of the polygon. High values of α and low values of σ generate polygons that have shapes that are close to a circle with radius μ. Finally, Δ_t is utilized to preserve the

fractal dimension of both processes, X and P, and determines the angular step, ϕ_1 (see Fig. .5).

4.1.2 MR-Polygon parameterization

The process of establishing the values for α, μ, σ, Δ_t and X_0 is not an easy task, and their values must be set in such a way that the shape of P is similar to a real irregular polygon. However, how do we determine whether a polygon P satisfies this condition? In this case, the fractal dimension appears to be a tool that offers strong theoretical support to assess the shape of a given polygon.

According to Richardson (1961), the fractal dimension D of an irregular polygon (such as a coast) is a number between 1 and 2 (1 for smooth boundaries and 2 for rough boundaries) that measures the way in which the length of an irregular boundary L (equation .5) changes when the length of the measurement instrument (ε) changes. The fractal dimension in given by equation (.6), where $\hat{C}$ is a constant.

In general, an object is considered to be a fractal if it is endowed with irregular characteristics that are present at different scales of study (Mandelbrot, 1982). For practical purposes, D is obtained using equation (.6) and is given by 1 minus the slope of $\log(L(\varepsilon))$. This procedure is commonly known as the Richardson plot.

$$L(\varepsilon) = \hat{C}\varepsilon^{1-D} \tag{.5}$$

$$\log(L(\varepsilon)) = (1-D)\log(\varepsilon) - \log(\hat{C}) \tag{.6}$$

In almost all cases, the Richardson plot can be explained with two line segments that have different slopes; then, two fractal dimensions can be obtained: textural, for small scales, and structural, for large scales (Kindratenko and Treiger, 1996). As illustrated, Fig. .6 shows a segment of the United States east coast taken from Google maps in two resolutions. Note that as the resolution increases, some irregularities that were imperceptible at low resolution become visible. In this sense, it can be said that irregularities at low resolution define the general shape and are related to the structural dimension, while irregularities at high resolution capture the noise and are related to the textural dimension. Regional scientists tend to use highly sampled maps, which preserve the general shape but remove the small variations. This simplification does not change the topological configuration of the maps (Douglas, 1973). Fig. .7 presents the Richardson plot of the external boundary of the United States and its textural and structural fractal dimension.

In the field of stochastic processes, some approaches, which are based on different estimations of the length, have been made to characterize them through their fractal dimension. In our case, an experimental approach based on the fractal dimension of real polygons is proposed to select an appropriate combination of the parameters α and σ to generate realistic irregular polygons. Because our interest is on general shape rather than small variations, we account only for the struc-

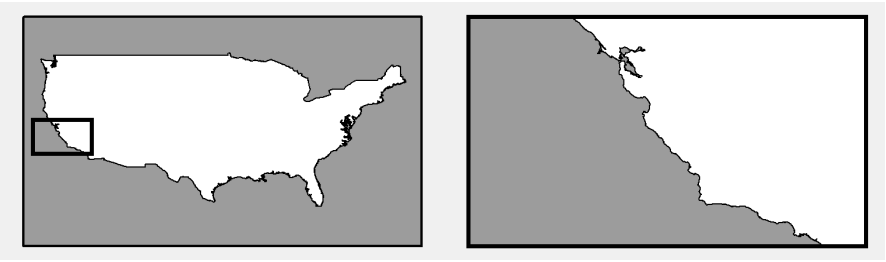


Fig. .6 Illustrative example of irregularities explained by the structural and textural dimension.

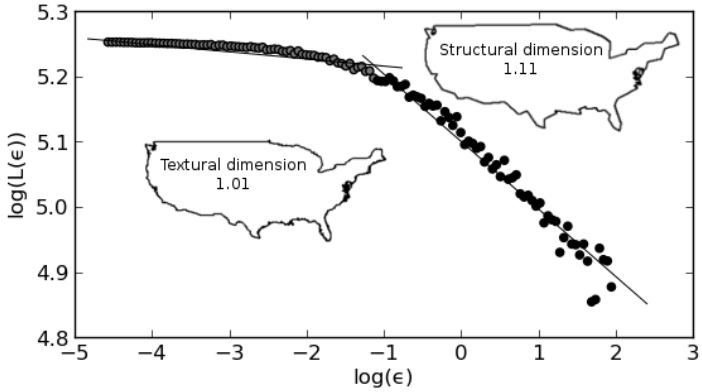


Fig. .7 Richarson plot to estimate the textural and structural dimension of the external boundary of the United States.

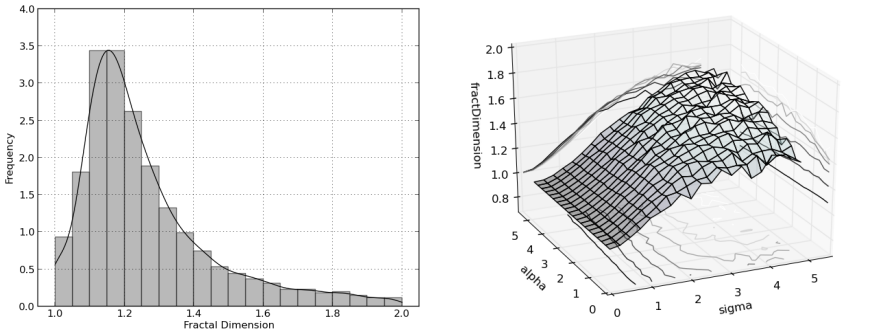
tural dimension[13] The parameterization process is divided into two parts: in the first part, the frequency histogram of the fractal dimensions of the real polygons is constructed. In the second part, we propose a range of possible values for α and σ, given μ, X_0, Δ_t, which generates fractal dimensions that are close to those obtained in the first part. Because the level of the long-term tendency μ does not affect the length of X and because algorithm 1 guarantees that the length is preserved, μ can be defined as a constant without affecting the fractal dimension. Hereafter, it is assumed that $\mu = X_0 = 10$. The value of Δ_t is set to be 0.001 to properly infer both of the fractal dimensions.

The empirical distribution of the fractal dimension of the irregular polygons is calculated over a random sample of 10,000 polygons from the world map used in section 3. The result of this empirical distribution is presented in Fig. .8.a. To find the fractal dimension of the *MR-Polygons*, we generate a surface of the average dimensions as a function of the values of α and σ, which range from 0.01 to 5 with

[13] To calculate the structural dimension, we use the EXACT procedure, which is devised by Allen et al. (1995), with a small value for Δ_t. Next, both of the dimensions were determined by using a k-means clustering algorithm over the cloud of points on the Richardson plot.

steps of 0.1 (Fig .8.b.). The resulting surface indicates that the fractal dimension is mainly affected by σ, especially when looking at small dimensions. Additionally, it is found that fractal dimensions close to 1.23 are obtained when σ takes on values between 1.2 and 1.5, regardless of the value of α.

Fig. .9 presents some examples of polygons using different values of α and σ. The polygons in the second row, which correspond to $\sigma = 1.5$, produce irregular polygons that have a realistic structural fractal dimension. Additionally, in the same figure, both the original (gray line) and sampled (black line) polygons reinforce the fact that sampling a polygon does not affect the structural dimension. From now on, we will use sampled polygons to improve the computational efficiency.



(a) Fractal dimensions of real polygons.

(b) Fractal dimension of simulated polygons as a function of α and σ.

Fig. .8 Stages to find the values of α and σ.

4.2 Recursive Irregular maps (RI-Maps)

Up to this point, we were able to generate irregular polygons with fractal dimensions that are similar to those from real maps. The next step is to use these polygons to create irregulars lattices of any size whose topological characteristics are close to the average values obtained for these characteristics in real lattices around the world. For this step, we formulate a recursive algorithm on which an irregular frontier is divided into a predefined number of polygons using *MR-Polygons*. Our conceptualization of the algorithm was made under three principles: (1) *Scalability:* Preserving the computational complexity of the algorithm when the number of polygons increases; (2) *Fractality:* Preserving the fractal characteristics of the map at any scale; and (3) *Correlativity:* Encouraging the presence of spatial agglomerations of polygons with similar sizes, which is commonly present in real maps in which there are clusters of small polygons that correspond to urban areas.

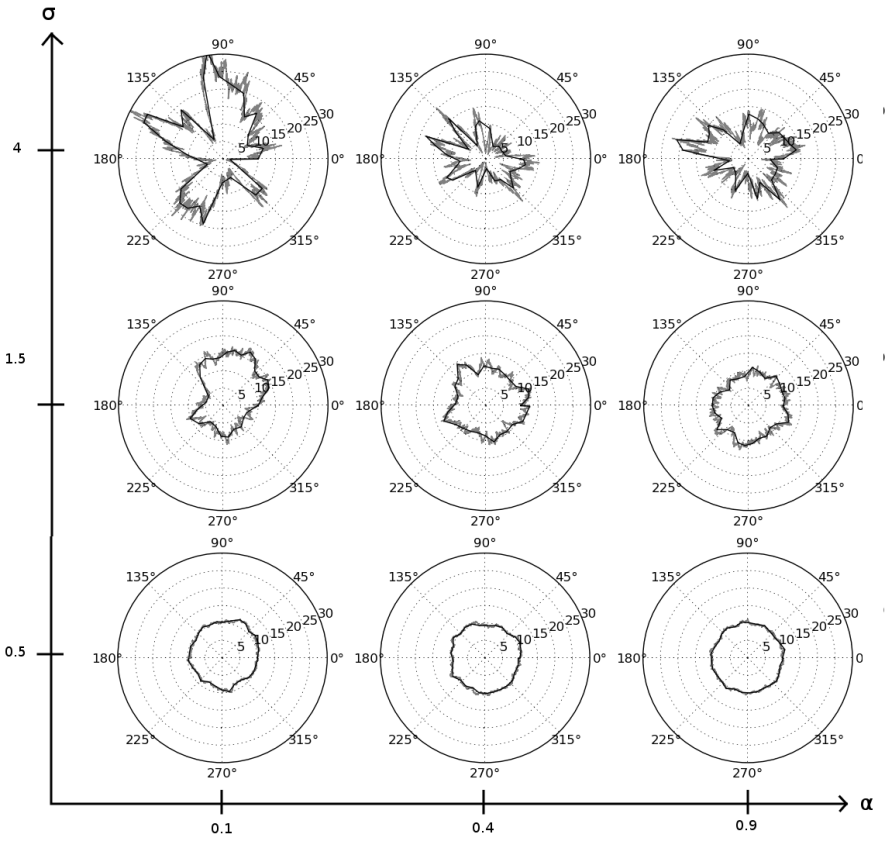


Fig. .9 Examples of stochastic polygons generated using (Algorithm 1) with different values of σ and α.

Algorithm 2 presents the *RI-Maps* algorithm to create polymorphic irregular aperiodic asymmetric lattices with realistic topological characteristics. This algorithm starts with an initial empty irregular polygon, *pol*, (the outer border of the *RI-Map*) and the number of polygons, *n*, to fit inside. In a recursive manner, a portion of the initial polygon *pol* starts being divided following a depth-first strategy until that portion is divided into small polygons. This process is repeated for a new uncovered portion of *pol* until the whole area of *pol* is covered. Because the recursive partitions are made by using *MR-Polygons*, we take the values of α and σ from a uniform distribution between 0.1 and 0.5 and 1.2 and 1.5, respectively, which correspond to the ranges of values that generate realistic polygons (see section 4.1). Regarding μ, X_0 and Δ_t, these variables take the same values proposed in section 4.1. Last, to guarantee the computational treatability of the geometrical operations, each polygon comes from a sampling process of 30 points.

To ensure a full understanding of the *RI-Maps* algorithm, we summarize in Fig. .10 its main steps, which indicates the link between the figure and the lines in the *RI-Maps* algorithm.

The *RI-Maps* algorithm has three unknown parameters:

- p_1: Because each polygon is created by the *MR-Polygons* using a polar coordinate system that is unrelated to the map being constructed with *RI-Maps*, it is necessary to apply a scaling factor, $\sqrt{\frac{p_1 \times area(pol)}{n \times \pi \times \mu^2}}$, that adjusts the size of the *MR-Polygon* before being included into the *RI-Map*.
- p_2: When a new polygon is used to divide its predecessor, its capacity to contain new polygons (measured by the number of polygons) is proportional to its share of the unused area of its predecessor. However, to enforce the appearance of spatial agglomerations of small polygons, the number of polygons that the new polygon can hold is increased with a probability of p_2.
- p_3: When p_2 indicates that a new polygon will hold more polygons, the number of extra polygons is calculated as the p_3 percent of the number of missing polygons that are expected to fit into the unused area of its predecessor polygon. The number of extra polygons is subtracted from the unused area to keep constant the final number of polygons (n).

Table .4 illustrates the effect of the parameters p_2 and p_3 on the topological characteristics of *RI-Maps*. In the first row, p_2 and p_3 equal 0, which generates highly ordered lattices without spatial agglomerations. The second and third rows are more disordered than the first row and have spatial agglomerations, with those in the second row less frequent and evident than those in the third row. As will be shown in the next section, lattices in the third row are more realistic in terms of their topological characteristics.

To find a combination of p_1, p_2, and p_3 that generates realistic *RI-Maps* in terms of their topological characteristics, we use a standard genetic algorithm, where the population γ at iteration i, denoted as γ^i, is formed by the genomes $\gamma^i_j = [p^i_{j_1}, p^i_{j_2}, p^i_{j_3}]$, where $p^i_{j_1}$, $p^i_{j_2}$, and $p^i_{j_3}$ are real numbers between 0 and 1, representing instances of p_1, p_2, p_3, which are denoted as phenomes. In this case, $i \in \mathbb{N}$ between 0 and 20 and $j \in \mathbb{N}$ between 0 and 100. To evaluate the quality of each genomes fitness function, $F(\gamma^i_j)$ is defined in equation (.7), where θ is a set of polygons, ϕ_k is the relative importance for a map of k polygons, and $f_k(\gamma^i_j)$ is a function given by equation (.8) that measures the average difference between the values of the topological indicators of real lattices and those values of *RI-Maps* formed by k polygons using the phenome γ^i_j. For the sake of simplicity, in equation (.8), $\Psi_k = [M_n, m_n, \mu_1, \mu_2, S, \lambda_1]$ denotes the vector of real indicators, and $\Psi_k(\gamma^i_j)$ denotes the vector for the mean values of *RI-Maps* with k polygons using γ^i_j. The superindex l is used in Ψ^l_k and $\Psi^l_k(\gamma^i_j)$ to refer to the l^{th} indicator in the real and simulated values, respectively. Finally, ns is the number of simulations to be generated with each genome.

Algorithm 2 RI-Map: Recursive Irregular Map

function RECURSIVEIRREGULARMAP(n, pol)
  $(\alpha_{min}, \alpha_{max}, \sigma_{min}, \sigma_{max}, \mu, X_0, \Delta_t) = (0.1, 0.5, 1.2, 1.5, 10, 10, 0.001)$
  $p_1 \in \mathbb{R}, p_2 \in \mathbb{R}, p_3 \in \mathbb{R}$
  if $n > 2$ **then**
    $missingPolygons \leftarrow n$
    $uncoveredPolygon \leftarrow pol$
    $coveredPolygon \leftarrow \phi$
    $polygons \leftarrow []$
    $scalingFactor \leftarrow \sqrt{\frac{p_1 \times area(pol)}{n \times \pi \times \mu^2}}$
    while $\frac{area(uncoveredPolygon)}{area(pol)} >= 0.03$ **do**
      $uncovered2select \leftarrow$ Bigger part of $uncoveredPolygon$
      if $missingPolygons \times \frac{area(uncovered2select)}{area(uncoveredPolygon)} <= 1.5$ **then**
        $polygons.put(uncovered2select)$
        $coveredPolygon \leftarrow coveredPolygon \cup uncovered2slect$
        $missingPolygons \leftarrow missingPolygons - 1$
      else
        $\alpha \leftarrow RandomUniform(\alpha_{min}, \alpha_{max})$
        $\sigma \leftarrow RandomUniform(\sigma_{min}, \sigma_{max})$
        $pol_i \leftarrow$MEANREVERTINGPOLYGON$(\alpha, \sigma, \mu, X_0, \Delta_t)$
        $pol_i \leftarrow$ Multiply each ratio of pol_i by $scalingFactor$
        $pol_i \leftarrow$ Center pol_i randomly into $uncovered2select$
        $pol_i \leftarrow (pol_i - coveredPolygon) \cap pol$
        $pol_i \leftarrow$ Bigger part of pol_i
        $n_i \leftarrow missingPolygons \times \frac{area(pol_i)}{area(uncoveredPolygon)}$
        if $Uniform(0,1) < p_2$ **then**
          $n_i = n_i + missingPolygons \times p_3$
        end if
        $n_i \leftarrow Round(n_1)$
        if $n_i \geq 1$ **then**
          $polygons_i \leftarrow$ RECURSIVEIRREGULARMAP(n_i, pol_i) ▷ Recursive step
          $polygons \leftarrow polygons \cup polygons_i$
          $coveredPolygon \leftarrow coveredPolygon \cup polygons_i$
          $missingPolygons \leftarrow missingPolygons - n_i$
        end if
      end if
      $uncoveredPolygon \leftarrow pol - coveredPolygon$
    end while
    Append interior holes of $coveredPolygon$ to $polygons$
    $coveredArea \leftarrow \bigcup polygons$
    while $length(polygons) < n$ **do**
      Append the smaller polygon to its larger neighbor
    end while
    while $length(polygons) > n$ **do**
      Divide the larger polygon
    end while
  else if n = 1 **then** ▷ Terminating case
    $polygons \leftarrow [pol]$
  else ▷ Terminating case
    $pol_1, pol_2 \leftarrow$ Divide pol in 2
    $polygons \leftarrow [pol_1, pol_2]$
  end if
  return polygons
end function

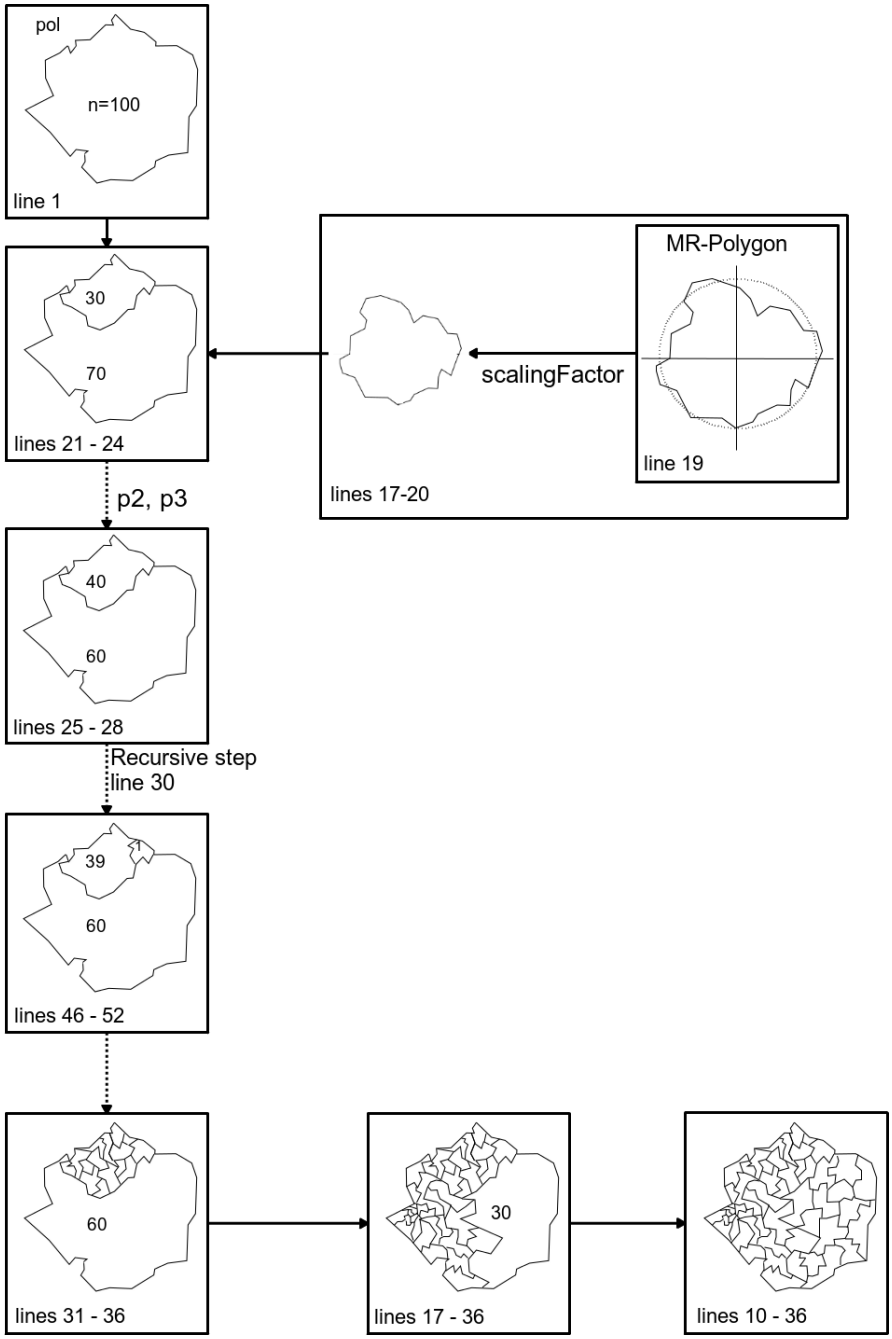
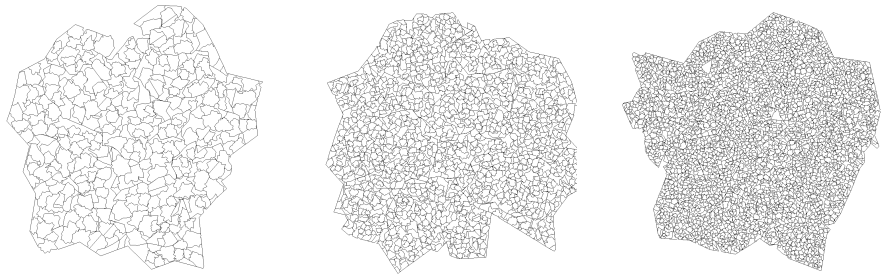


Fig. .10 Diagram of the main steps of the *RI-Maps*

Table .4 Examples of *RI-Maps* of 400, 1,600 and 3,600 polygons using different combinations of parameters



i) $p_1 = 0.1$, $p_2 = 0.0$, $p_3 = 0.0$

ii) $p_1 = 0.1$, $p_2 = 0.1$, $p_3 = 0.5$

iii) $p_1 = 0.010$, $p_2 = 0.050$, $p_3 = 0.315$

$$F(\gamma_j^i) = \frac{\left(\sum_{k\in\theta} \phi_k f_k(\gamma_j^i)\right)}{\sum_{k\in\theta} \phi_k} \tag{.7}$$

$$f_k(\gamma_j^i) = \frac{\sum_{l=1}^{6} \frac{(\sum_{s=i}^{ns} \Psi_k^l(\gamma_j^i)) - ns\Psi_k^l}{ns\Psi_k^l}}{6} \tag{.8}$$

The algorithm starts with an initial random population of 100 genomes to obtain the best four genomes. The subsequent populations are composed of two parts. The first 64 genomes are all of the possible combinations of the last best 4 genomes, and the other 36 genomes are random modifications of those 64 genomes. Because of the computational time required to evaluate equation (.7), only lattices of 400 and 1,600 were used, with an importance of $\phi_{400} = 1$ and $\phi_{1.600} = 2$, respectively. The algorithm reached the optimal value after 13 iterations with $p_1 = 0.010$, $p_2 = 0.050$, and $p_3 = 0.315$.

5 Results

Fig. .11 presents a graphical comparison of the topological characteristics of real *RI-Maps* and Voronoi diagrams. The values for the *RI-Maps* were obtained from 100 instances. The results show that *RI-Maps* have a maximum (M_n) and a minimum (m_n) number of neighbors that are very close to the values found in the real lattices. Regarding the average number of neighbors, both *RI-Maps* and Voronoi diagrams show similar values that are slightly higher than those observed in real lattices; however, because the number of neighbors is an integer value, it can be concluded for all three cases that the average number of neighbors is 6, which verifies the findings by Weaire and Rivier (2009) in irregular lattices. Regarding μ_2, *RI-Maps* are a better approach to simulate the level of disorder found in real lattices. To facilitate the visualization, the values of S are reported as $S * \sqrt{n}$. The results show that *RI-Maps* replicate the values of real lattices at any size, while Voronoi diagrams report higher values that tend to increase with the number of polygons. Last, *RI-Maps* have values of λ_1 that are closer to the values of real lattices, especially for large instances.

Table .6 presents the average and standard deviation of *RI-Maps* under the optimal parameters ($p_1 = 0.010$, $p_2 = 0.050$, $p_3 = 0.315$) found in the previous section. This table completes the topological information on lattices presented in Table .3.

6 Application of *RI-Maps*

In this section, we present an example of the use of *RI-Maps* based on the computational experiments designed by Duque et al. (2011a) to compare the efficiency of the improved AMOEBA algorithm. To present the results, Duque et al. (2011a)

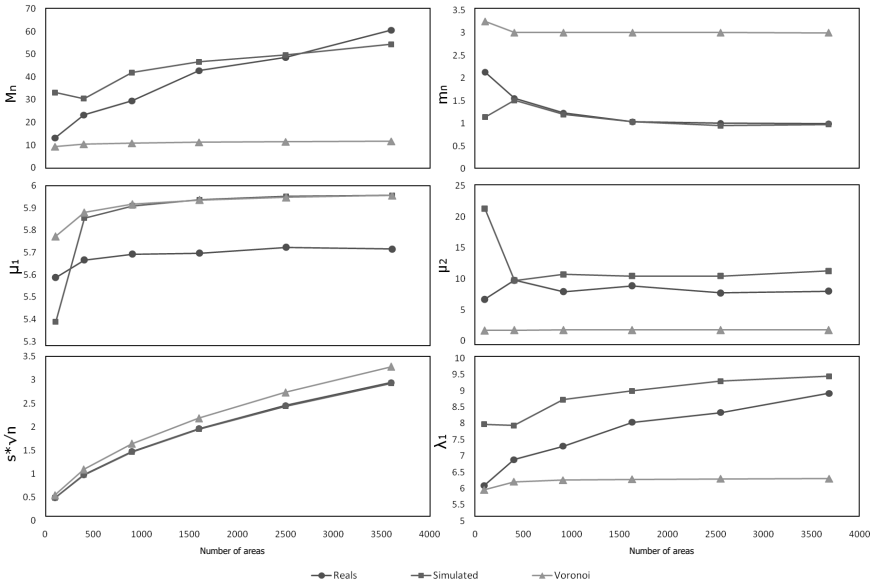


Fig. .11 Comparison of the topological characteristics of real lattices, *RI-Maps* and Voronoi diagrams.

Table .5 Average topological characteristics for *RI-Maps*

	Number of polygons						
	81	100	400	900	1,600	2,500	3,600
M_n	26.500 ±12.765	33.100 ±14.751	30.460 ±13.664	41.870 ±16.035	46.760 ±16.429	49.730 ±16.886	54.396 ±15.156
m_n	1.260 ±0.691	1.140 ±0.513	1.510 ±0.847	1.200 ±0.550	1.040 ±0.374	0.950 ±0.261	0.979 ±0.204
μ_1	5.347 ±0.333	5.388 ±0.304	5.855 ±0.093	5.909 ±0.052	5.937 ±0.036	5.952 ±0.032	5.957 ±0.027
μ_2	17.397 ±13.734	21.313 ±14.729	9.722 ±6.189	10.708 ±3.831	10.426 ±2.277	10.443 ±1.506	11.276 ±1.229
S	5.974 ±0.324	4.879 ±0.219	1.372 ±0.020	0.620 ±0.006	0.353 ±0.002	0.226 ±0.001	0.157 ±0.001
λ_1	7.431 ±0.804	7.969 ±0.808	7.931 ±0.979	8.724 ±0.962	8.993 ±0.960	9.299 ±0.957	9.449 ±0.798

proposed 3 computational experiments; one of them reports the running time of AMOEBA as the number of polygons of regular lattices increases. In this paper, we will run the same algorithm not only for regular lattices but also for real irregular and simulated irregular lattices (*RI-Maps*). First, we want to see whether running the computational experiment on regular lattices only is a good representation of the performance of the algorithm. (Can the conclusions that are obtained for regular lattices be extrapolated to irregular lattices?) Second, we want to see whether the results from using *RI-Maps* are also valid for real irregular maps.

For the experiment that we generated, for each type of lattice, there were 30 instances with 1,600 polygons. For each instance, we generated a spatial process that had four clusters, each using the methodology proposed by Duque et al. (2011a). Last, the instances for real maps were obtained from sampling the same world map that was used in previous sections. Fig. .12 presents the distribution of the running times obtained for each type of lattice, and Table .6 compares the distributions with the two-sided Kolmogorov-Smirnov test (George Marsaglia and Wang, 2003). The null hypothesis of the Kolmogorov-Smirnov test is that the two samples come from the same probability distribution. Regarding the first question, it is clear that using a regular lattice for testing the algorithm can lead to underestimating the execution times of the algorithm. On the other hand, the distribution of the running times obtained for real and *RI-Maps* is statistically equal, which shows the benefits of using *RI-Maps* because it can automatically generate instances without limits on their sizes.

Table .6 Kolmogorov-Smirnov test to compare the distributions of AMOEBA execution times using different lattices

	Regular Lattices	RI-Maps	Real Maps
Regular Lattices	0.00 (p=1)	0.51 (p=$0.0e^{-4}$)	0.61 (p=$2.8e^{-5}$)
RI-Maps	0.51 (p=$0.0e^{-4}$)	0.00 (p=1)	0.19 (p=0.607)
Real Maps	0.61 (p=$2.8e^{-5}$)	0.19 (p=0.607)	0.00 (p=1)

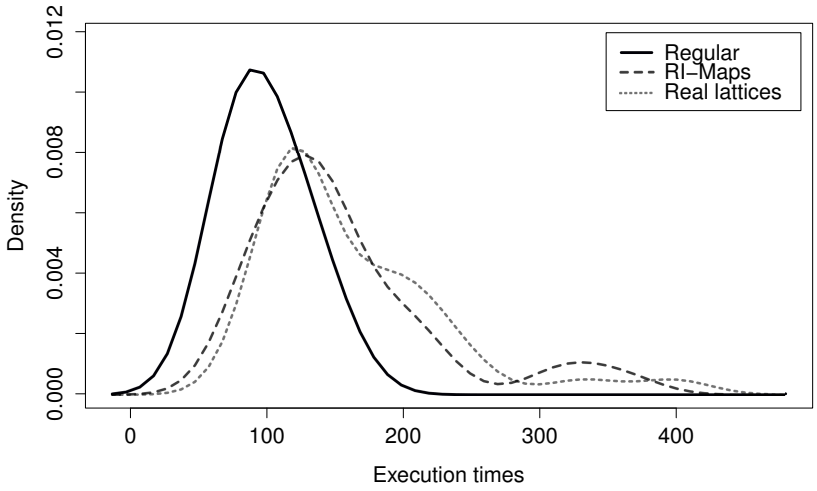


Fig. .12 Execution times of AMOEBA over regular lattices and *RI-Maps* of 1,600

7 Conclusions

This paper introduces an algorithm that combines fractal theory, the theory of stochastic processes and computational geometry for simulating realistic irregular lattices with a predefined number of polygons. The main goal of this contribution is to provide a tool that can be utilized for geocomputational experiments in the fields of exploratory spatial data analysis, spatial statistics and spatial econometrics. This tool will allow theoretical and empirical researchers to create irregular lattices of any size and with topological characteristics that are close to the average characteristics found in irregular lattices around the world.

As shown in the last section, the performance of some geocomputational algorithms can be affected by the topological characteristics of the lattices in which these algorithms are tested. This situation can lead to an unfair comparison of algorithm performances in the literature. With the algorithm proposed in this paper, the differences in the computational performances will not be affected by the topological characteristics of the lattices.

This paper also shows that the topological characteristics of regular lattices (with squared and hexagonal polygons) and Voronoi diagrams (commonly used to emulate irregular lattices) are far from the topological characteristics that are found in real lattices.

References

Aldstadt, J. and Getis, A. (2006). Using AMOEBA to create a apatial weights matrix and identify spatial clusters. *Geographical Analysis*, 38(4):327–343.

Allen, M., Brown, G., and Miles, N. (1995). Measurement of boundary fractal dimensions: review of current techniques. *Powder Technology*, 84(1):1–14.

Anselin, L. (1988). *Spatial econom trics: methods and models*. Kluwer Academic Publishers, Dordrecht, 1 edition.

Anselin, L., Bera, A., Florax, R., and Yoon, M. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26(1):77–104.

Anselin, L., Florax, R., and Rey, S. J. (2004). *Advances in spatial econometrics: methodology, tools and applications*. Springer.

Anselin, L. and Moreno, R. (2003). Properties of tests for spatial error components. *Regional Science and Urban Economics*, 33(5):595–618.

Anselin, L. and Smirnov, O. (1996). Efficient algorithms for constructing proper higher order spatial lag operators. *Journal of Regional Science*, 36(1):67–89.

Aste, T., Szeto, K., and Tam, W. (1996). Statistical properties and shell analysis in random cellular structures. *Physical Review. E, Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 54(5):5482–5492.

Bartlett, M. S. (1975). *Probability, statistics, and time: a collection of essays*. Chapman and Hall, New York, 1 edition.

Batty, M. and Longley, P. (1994). *Fractal cities: A geometry of form and function*. Harcout Brace & Company, London.

Blommestein, H. and Koper, N. (2006). Recursive algorithms for the elimination of redundant paths in spatial lag operators. *Journal of Regional Science*, 32(1):91–111.

Boots, B. (1982). Comments on the use of eigenfunctions to measure structural properties of geographic networks. *Environment and Planning A*.

Boots, B. (1984). Evaluating principal eigenvalues as measures of network structure. *Geographical Analysis*, 16(3):270–275.

Boots, B. (1985). Size effects in the spatial patterning of nonprincipal eigenvectors of planar networks. *Geographical Analysis*, 17(1):74–81.

Chen, Y. (2011). Derivation of the functional relations between fractal dimension of and shape indices of urban form. *Computers, Environment and Urban Systems*.

Church, R. L. (2008). BEAMR: An exact and approximate model for the p-median problem. *Computers & Operations Research*, 35(2):417–426.

Clark, W. (1964). The Concept of Shape in Geography. *American Geographical Society*, 54(4):561–572.

Cole, J. (1964). Study of major and minor civil divisions in political geography. In *20th International Geographical Congress*, Sheffield. University of Nottingham.

Coxeter, H. S. M. (1974). *Regular complex polytopes*. CUP Archive.

Douglas, D. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10(2):112–122.

Duque, J., Aldstadt, J., Velasquez, E., Franco, J., and Betancourt, A. (2011a). A computationally efficient method for delineating irregularly shaped spatial clusters. *Journal of Geographical Systems*, 13:355–372.

Duque, J. C., Anselin, L., and Rey, S. J. (2012). the max-p-regions problem*. *Journal of Regional Science*, 52(3):397–419.

Duque, J. C., Dev, B., Betancourt, A., and Franco, J. L. (2011b). ClusterPy: {Library} of spatially constrained clustering algorithms, {Version} 0.9.9.

Elhorst, J. P. (2003). Specification and estimation of spatial panel data models. *International Regional Science Review*, 26(3):244–268.

Frankhauser, P. (1998). The fractal approach: A new tool for the spatial analysis of urban agglomerations. *New Methodological Approaches in the Social Sciences*, 10(1):205–240.

Garrison, W. (1964). Factor-analytic study of the connectivity of a transportation network. *Papers in Regional Science*.

George Marsaglia, W. W. T. and Wang, J. (2003). Evaluating Kolmogorov s Distribution. *Journal of Statistical Software*, 8(18):1–4.

Ghyka, M. (2004). *The Geometry of art and life*. Kessinger Publishing.

Gibbs, J. (1961). A method for comparing the spatial shapes of urban units. In *Urban Research Methods*, pages 96–106. D. Van Nostrand Company, Inc, Princeton.

Gould, P. (1967). On the geographical interpretation of eigenvalues. *Transactions of the Institute of British Geographers*, 42(42):53–86.

Griffith, D. (1987). Toward a theory of spatial statistics: Another step forward. *Geographical Analysis*, 19(1):69–82.

Grunbaum, B. and Shephard, G. C. (2011). *Tilings and patterns*. Dover Publications.

Haggett, P. (1977). *Locational analysis in human geography*. John Wiley & Sons, London.

Haining, R. (2010). The nature of georeferenced data. In Fischer, M. M. and Getis, A., editors, *Handbook of Applied Spatial Analysis*, pages 197–217. Springer Berlin Heidelberg, Berlin, Heidelberg.

Hijmans, R., Guarino, L., Jarvis, A., O'Brien, R., Mathur, P., Bussink, C., Cruz, M., Barrantes, I., and Rojas, E. (2011). Diva-GIS.

Hooper, P. and Hewings, G. (1981). Some properties of space-time processes. *Geographical Analysis*, 13(3):203–223.

Horton, R. (1932). Drainage basin characteristics. *Transactions of the American Geophysical Union*, 13(1):350–361.

Johnson, D. L. (2001). *Symmetries*. Springer Verlag, London, 1 edition.

Kindratenko, V. and Treiger, B. (1996). Chemometrical approach to the determination of the fractal dimension (s) of real objects. *Chemometrics and Intelligent Laboratory Systems*, 34:103–108.

Le Caer, G. and Delannay, R. (1993). The administrative divisions of mainland France as 2D random cellular structures. *Journal de Physique I*, 3(8):1777–1800.

Le Caer, G. and Delannay, R. (1995). Topological models of 2D fractal cellular structures. *Journal de Physique I*, 5(11):1417–1429.

Mandelbrot, B. B. (1982). *The fractal geometry of nature*. W.H. Freeman.

Mao, X. (1997). *Stochastic differential equations and applications*. Horwood publishing limited, Chichester, 1 edition.

Miller, V. (1953). *A quantitative geomorphic study of drainage basin characteristics in the clinch mountain area Virginia and Tenesse.* Department of Geology, Columbia University.

Mur Lacambra, J. (1992). Contrastes de autocorrelación espacial: Un estudio de Monte Carlo. *Estadística Española*, 34(130):285–308.

Murray, A. T. and O'Kelly, M. E. (2002). Assessing representation error in point-based coverage modeling. *Journal of Geographical Systems*, 4(2):171–191.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70(349):120–126.

Pace, R. and LeSage, J. P. (2004). Chebyshev approximation of log-determinants of spatial weight matrices. *Computational Statistics & Data Analysis*, 45(2):179–196.

Penrose, R. (1974). The Rôle of Aesthetics in Pure and Applied Mathematical Research. *Journal of The Institute of Mathematics and its Aplications*, 10:266–271.

Peshkin, M., Strandburg, K., and Rivier, N. (1991). Entropic predictions for cellular networks. *Physical Review Letters*, 67(13):1803–1806.

Radin, C. (1993). Symmetry of tilings of the plane. *Bullettin of the American Mathematical Society*, 29(2):213–217.

ReVelle, C. and Swain, R. (1970). Central facilities location. *Geographical Analysis*, 2(1):30–42.

Richardson, L. (1961). The problem of contiguity: An appendix of statistics of deadly quarrels. *General Systems Yearbook*, 6(13):139–187.

Ross, I. C. and Harary, F. (1952). On the determination of redundancies in socioeconometric chains. *Psychometrika*, 17(2):195–208.

Smirnov, O. and Anselin, L. (2001). Fast maximum likelihood estimation of very large spatial autoregressive models: A characteristic polynomial approach. *Computational Statistics & Data Analysis*, 35(3):301–319.

Stoddart, D. (1965). The shape of atolls. *Marine Geology*, 3(5):269–283.

Tilley, R. (2006). *Crystals and crystal structures*. Wiley, England, 1 edition.

Tinkler, K. (1972). The physical interpretation of eigenfunctions of dichotomous matrices. *Transactions of the Institute of British Geographers*, 55(55):17–46.

Weaire, D. and Rivier, N. (2009). Soap, cells and statistics random patterns in two dimensions. *Contemporary Physics*, 50(1):199–239.

Weeitty, A. (1969). *On the form of drainage basins*. Dept. of Geography, Pennsylvania State University, Pennsylvania, 1 edition.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 41(3):434–449.