Mathematical Geology, Vol. 36, No. 1, January 2004 (© 2004)

Detecting Randomness in Spatial Point Patterns: A "Stat-Geometrical" Alternative[1]

Paulo Sérgio Lucio[2] and Nilson Luiz Castelucio de Brito[3]

There are several methods to test the hypothesis of complete spatial randomness of point patterns. This work involves a new geometrical-based strategy to detect spatial arrangements, which takes into account both Euclidean and angular distances, defining a triangle-based network. An asymptotic test based on the Kolmogorov–Smirnov statistic is proposed to accommodate this situation. To assess the usefulness of this method (Stat-Geo), simulations based on Monte Carlo procedures, conducted using SPLUS™, give satisfactory results with a high degree of accuracy. As expected, the new technique proposed in this paper, performs better than traditional ones like distance-based or angle-based, since more information (combining distance and angle) is introduced in the decision-making system. This approach is a very simple way to offer high efficiency results for a low computational cost. Furthermore, this alternative method allows barycentric interpolation of the unsampled points into a two-dimensional simplex (triangular) framework.

KEY WORDS: spatial pattern analysis, spatial clusters, triangulation, barycentric interpolation, Kolmogorov–Smirnov test.

INTRODUCTION

Spatial point patterns can be observed in many natural phenomena. An important task of the analysis of spatial point patterns is to examine the dependence between points. Spatial data analysis is differentiated from typical data analysis by the inclusion of spatial information in models. Conventional statistical methods cannot detect whether there is spatial dependence configuration in the observational units. Point pattern data arise when locations themselves are the variable of interest. Spatial point patterns consist of a finite number of locations observed in the space, essentially planar region. Identification of spatial randomness, clustering, or regularity is often the first analysis performed when looking at point patterns (Venables and Ripley, 1999; S+SpatialStats, 2000). The standard against which spatial point

[1]Received 14 November 2002; accepted 13 October 2003.

[2]Centro de Geofísica de Évora (CGE), Universidade de Évora, Apartado 94, 7002-554 Évora, Portugal; e-mail: pslucio@uevora.pt

[3]Departamento de Estatística (EST), Universidade Federal de Minas Gerais, Minas Gerais, Brazil.

patterns are compared is a complete spatial random process. The complete spatial randomness (CSR) model uses some assumptions and several methods to ascertain whether there is a tendency for events to exhibit a random, systematic or clustered pattern (Fig. 1). Under CSR, events are independent and the number of events in any specified area of fixed size is Poisson distributed (Gatrell and others, 1996).

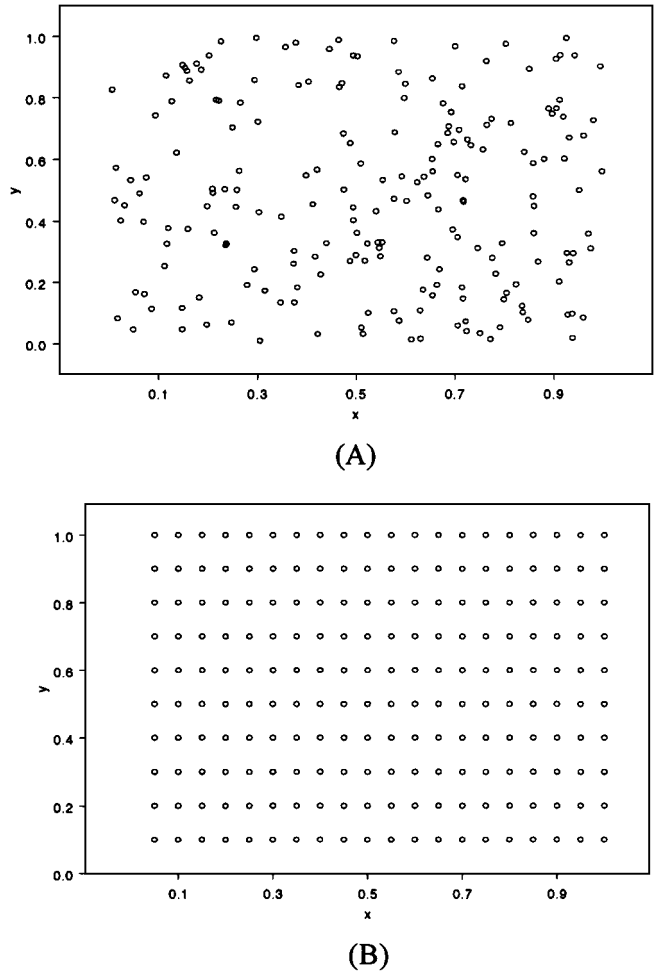


Figure 1. Three spatial point patterns generated in $[0, 1]^2$. Simulations of 200 points in (A), random field design by a Neyman–Scott stochastic process, in (B), regular arrangement and in (C), clustered configuration by a Strauss stochastic process.

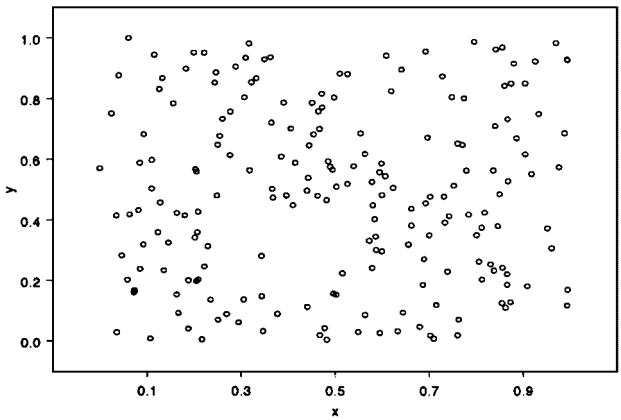


Figure 1. (*Continued.*)

Rejection of CSR is a prerequisite for any serious attempt to model any observed pattern. Statistical tests are used to explore a set of data and assist in the formulation of alternatives to CSR (Ripley, 1976; Besag and Newell, 1991; Bonetti and Pagano, 2001; Rogerson, 1999). In this paper, various methods for the analysis of a set of event locations are considered; many of these tests for CSR are well known.

One aim of modeling spatial-point pattern is to construct specific hypothesis test to help explain observed patterns. A new tessellation-based strategy for the detection of spatial clusters of arbitrary geometric form in a map of geo-referenced populations and cases is proposed. The present work is based on two fundamental measures—distance and angle—used to facilitate barycentric interpolation in preestablished influence zones. Thus, a geometrical tool is presented to analyze spatial-point patterns. This analysis can be accomplished subdividing areas of a region into triangles to assess if a spatial clustering structure exists. The basic ideas of the test are given below.

The design of this study is first to construct elementary cells (simplex structure) by triangulation of the convex subregion $S \subseteq R$. Next, to evaluate the CSR hypothesis, test statistic is computed on the basis of the areas of each elementary triangle, taking into account angles and distances. The performance of the simplex-based test is illustrated using simulated data.

This study deals with sparsely sampled point patterns. Some points are sampled from an ensemble and the spatial pattern is studied through its local properties, in the neighborhood of the sampled locations. To distinguish the sampled locations or the observed occurrences from the arbitrary locations composing the ensemble or the population, one calls the former events and the latter points.

Two worked "live" examples where geosciences implications are clear are included at the end of this paper, one in geomorphology and the other in seismology.

TESTS FOR CSR

CSR is defined by the following criteria: the intensity (the expected number of points per unit area) of the point pattern does not vary over the bounded sampling region, A. Hence, the hypothesis of complete spatial randomness (CSR) for a spatial-point pattern asserts that H_0: *The number of events m in any planar region A with area* $|A|$ *follows a homogeneous Poisson distribution with mean* $\lambda|A|$, *where the expected constant intensity is* λ. The distribution function is given by

$$f_{\lambda|A|}(m) = \frac{e^{-\lambda|A|}(\lambda|A|)^m}{m!}. \tag{1}$$

The m events in A constitute an independent random sample on A and the first-order properties (the expected intensity) of a spatial point process describe how the mean number of points per unit area varies through space. Well-known methods for detecting CSR include Quadrat test, Nearest-Neighbor test, and Angle tests. Notice that the angle-based tests are useful in regions with terrains that make distance measurements difficult.

SIMPLE QUADRAT TESTS

The dataset comprise counts of events $(m_1, m_2, \ldots, m_k)$ in k quadrats. A simple test statistics uses the Poisson distribution characteristics of equal mean and variance and computes the sample variance-to-mean ratio, or the index of dispersion (Bailey and Gatrel, 1995; Diggle, 1983; Okabe and Yamada, 2001):

$$s_m^2/\bar{m},$$

given by

$$I = \frac{(k-1)s_m^2}{\bar{m}} = \frac{\sum\limits_{n=1,k}(m_n - \bar{m})^2}{\bar{m}}. \tag{2}$$

Under CSR, $I \sim \chi^2_{k-1}$. Hence, I can be compared with percentage points of χ^2_{k-1}. Significant large values indicate clustering or aggregation, whereas small values indicate regularity. However, the main limitation of the quadrats-based test is the

losts of spatial information: it takes no account the position of quadrats or events within quadrats and there is no information about the neighborhood.

It is known that the neighborhood information is crucial in geology (locations of rock-type), seismology (epicenters of earthquakes in space–time clustering analysis), volcanology (locations of craters in a volcanic field), climatology (occurrences of spatial climate extremes), ecology (locations of seedlings in a forest), or sociology (locations of criminality indices). Hence, concerning to these fields and several others, the initial exploratory analysis of point patterns frequently requires a reliable test of the CSR hypothesis.

NEAREST-NEIGHBOR TESTS

In the nearest-neighborhood test, the empirical distribution function of the point-to-point (or the origin-to-point) nearest-neighbor distances is determined. Under CSR, the theoretical distribution function is given by Equation (1), where the intensity (the number of points per unit area) is λ, d is the distance between events (Besag and Cleaves, 1973; Hines and Hines, 1979), and πd^2 is the area of the circle centered on the target event. Then the probability that no events fall within a circle of radius d around any randomly chosen point is $e^{-\lambda\pi d^2}$. The distribution function, $F_0(m)$, of nearest neighbor point–event distances for CSR is given by

$$F_0(m) = P\{X \leq m\} = 1 - e^{-\lambda\pi d^2}. \tag{3}$$

The area of the circle, πd^2, follows an exponential distribution with parameter λ. Therefore $2\pi\lambda\, d^2 \sim \chi^2_2$ and $2\pi\lambda \sum d_i^2 \sim \chi^2_{2k}$. The distribution theory for this test is based on the assumption of independence of k nearest-neighbor measurements randomly sampled from a region R. This assumption of independence may be violated in the case of small numbers of events and if the proportion of them is large (Diggle, 1983; Diggle, Besag, and Gleaves, 1976). Since the distribution theory assumes independence, situations in which distances are not independent will promote inappropriate rejection of the null hypothesis of CSR. In particular, nearest-neighbor distances for events near the boundary will be biased, so it becomes important to apply edge corrections. Nearest-neighbor tests of CSR include those of Clarke–Evans, Hopkins, Blyth–Ripley; as well as those based on Ripley's K function.

Distance sampling refers to a class of methods where the basic sampling unit is the sample point and the configuration's information is obtained from distances to the nearest events. In practice, the proportion of points on a grid that is within a preestablished distance of the nearest point is computed. For a CSR process without edge effects, the theoretical distribution of this proportion is given by $F_0(d)$, where the intensity is the number of points per unit area (Bailey and Gatrell,

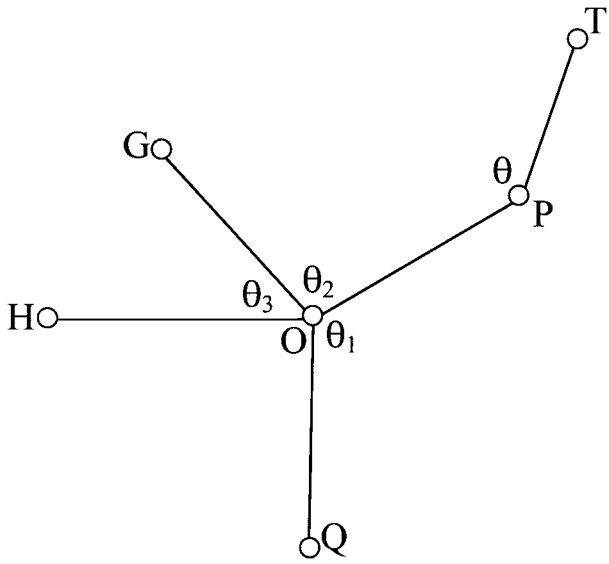


Figure 2. Illustrative scheme for the distance-based and the angle-based CSR tests. In the distance-based tests, the spatial neighbors are the nearest neighbors on the basis of the distances calculated using the point locations. The angle-based tests are adaptations of the nearest-neighbor test and are based on the smallest angle between two vectors that link the target event (O) to their two nearest neighbors. A slight modification of the angle-based tests incorporates the information concerning the third- and the fourth-nearest-neighbor events.

1995). Powerful tests, based on distance from the nearest neighbors, are generated by the T^2 method proposed by Besag and Gleaves (1973) and their modifications suggested by Hines and Hines (1979).

One can describe the idea of the T^2 method, following the scheme shown in Figure 2, where O is the target event, P and Q are the first- and the second-nearest neighbors of O, and the event T is denoted by the T^2 neighbor, the nearest neighbor of P under the constraint that the angle $O\hat{P}T$ is not inferior to $\pi/2$. This method uses the squared distance between OP and PT to calculate the statistics for the CSR test (Assunção and Reis, 2000; Diggle, Besag, and Gleaves, 1976).

ANGLE TESTS

When $m \geq 3$ events are sampled inside a convex subregion $S \subseteq R$, they span at least m measures of angles θ for each point. With n sampled points in the two-dimensional space distributed over a region R, the sample intensity is defined

by $m/n(m < n)$. If the intensity is less than 10%, the angles can be considered independent (Diggle, 1983; Diggle, Besag, and Gleaves, 1976); under CSR, $\theta_i \sim U[0, \pi]$, and the hypothesis can be tested using the Kolmogorov–Smirnov (K–S) procedure. In clustered patterns, the angle will tend to be smaller than expected under CSR, whereas in regular patterns, the angle will tend to be larger than expected under CSR.

The angle test (Assunção, 1994) is an adaptation of the nearest-neighbor test and is based on the smallest angle $P\hat{O}Q$, denoted by θ, between two vectors that link the target event to their two nearest neighbors. Under CSR configurations, $\theta \sim U[0, \pi]$. Assunção and Reis (2000) propose a modification of the angle test that incorporates the information concerning the third- and the fourth-nearest-neighbor events. The event O is randomly selected, P is its nearest neighbor, Q, G, and H are the second-, third-, and fourth-nearest-neighbor events, respectively, and T is the T^2 neighbor of P. The additional events span two tests: the first one uses the angles θ, θ_1 and the second one uses the angles $\theta, \theta_2, \theta_3$. Under CSR conditions, they are uniformly distributed on $[0, \pi]$ and independent (Fig. 2).

An adaptation of the angle test and its modification are very efficient for determining the structure of a spatial point pattern. However, it is possible to think of a trivial adaptation of these tests using measures based on Euclidian and angular distances, taking into account the construction of a tessellated network and the barycentric coordinates. Hence, in the present work a geometrical alternative is adopted using the areas of a region subdivided in triangles (elementary cells) to detect the spatial pattern of an ensemble of points. Here, the spatial-point pattern pertains to the structural distribution of an ensemble of punctual phenomena or its triangulated network. The Voronoi region of an object is the region of space closer to the given object than to any other object of the sample. The set of Voronoi triangulations for a set of spatial objects, called a Voronoi diagram (also known as a Dirichlet tessellation or Thiessen polygons), provides a partition of a point pattern according to its spatial structure. Features of this kind can also be used for analysis of the underlying point process.

In practice, the triangulated network of a subregion of the two-dimensional convex envelope must be determined. Let $m \geq 3$ *events* of a random sample over a subregion $S \subseteq R$ and assume that the two-dimensional convex hull of this subregion has area $|A|$ and that the partition produces $(M = 2m - \nu - 2)$ triangles, where ν is the number of extreme points of the two-dimensional partition of the unity A, with areas$|A_1|, |A_2|, \ldots, |A_{2m-\nu-2}|$, respectively. Under CSR, these areas are independent and uniformly distributed over the convex A. A sample of m punctual events from a population of n points is required to construct the empirical distribution of triangles $\hat{F}_M(x)$. The K-S test is used in a Monte Carlo fashion to test the hypothesis of CSR; that is, the values of the empirical distribution of the triangles are computed and the maximum deviation between $F_0(x)$—the distribution

under H_0—and $\hat{F}_M(x)$ is determined:

$$D_M = \sup_x |\hat{F}_M(x) - F_0(x)|. \tag{4}$$

Equation (4) can be modified to accommodate a sample size correction factor $\sqrt{M}$ as follows (Appendix):

$$D_M = \sqrt{M}\sup_x |\hat{F}_M(x) - F_0(x)|. \tag{5}$$

In the preceding equations, $\hat{F}_M(x)$ is the empirical distribution function and $F_0(x) = \frac{x}{A}$, $x \in [0, A]$, where x is the area of each one of the triangles of the subregion under study. If the points have a regular pattern, these areas will trend to have equal values, whereas if the points have a cluster pattern, these areas will tend to have many small values and a few number of large values or vice-versa.

Finally, a significance level α is chosen that determines the solution k_α in the following equation:

$$P\{D_M \geq k_\alpha | \hat{F}_M(x) = F_0(x)\} \leq \alpha \Rightarrow P_{H_0}\{D_M \geq k_\alpha\} \leq \alpha. \tag{6}$$

An attractive feature of this test is that the distribution of the K-S test statistic itself does not depend on the underlying cumulative distribution function being tested. Another advantage is that it is an exact test. Despite these advantages, the K-S test has two important restrictions:

(1) It applies only to continuous distributions and,
(2) It tends to be more sensitive near the centre of the distribution than at the tails.

As a consequence of these limitations, many analysts prefer to use the Anderson–Darling (A–D) goodness-of-fit test. However, the A–D test is available only for a few specific distributions. Furthermore, one considers the precision of the K-S is to be satisfactory.

For the numerical implementation, the Delaunay's triangulation scheme was used as a tessellation tool whose vertices are events, with the property that no point falls in the interior of the circle that passes through all three vertices of any triangle in the triangulated network. The numerical algorithm was adapted to produce the counterclockwise orientation of the angles around their own barycenter.

TESSELLATION AND ORIENTED AREAS

A closed ensemble K of the n-dimensional space $\Re^n$ is convex if for any $x \in K$, $y \in K$ and $0 \leq \lambda \leq 1$, the linear combination $\{\lambda x + (1 - \lambda)y \in K\}$. A point

$w \in K$ is an extreme point of K if it may not be written as a convex combination of k different elements of w. A two-dimensional convex envelope of a finite set $C = \{p_1, p_2, \ldots, p_m\}$ of m events of $\Re^n$ is defined as the set of all the convex combinations of elements of C:

$$\text{conv}\,(C) = \left\{ \sum_{i=1}^{m} \lambda_i p_i | \lambda_i \geq 0 \quad \text{and} \quad \sum_{i=1}^{n} \lambda_i = 1 \right\}. \tag{7}$$

The polygon $C = \{p_1, p_2, \ldots, p_m\}$ is convex if, and only if, each internal angle is convex, i.e., if each triangle p_{i-1}, p_i, p_{i+1} has the same polygon orientation. A triangle defines a coordinate system in the plane (Farin, 1993). Let us consider p_{i-1}, p_i, p_{i+1} noncollinear events onto a triangle $\Delta \subset \Re^2$. Each point $p \in \Delta \subset \Re^2$ can be written as a unique linear combination satisfying

$$\left\{ \sum_{i=1}^{3} \lambda_i p_i : \lambda_i \geq 0 \quad \text{and} \quad \sum_{i=1}^{3} \lambda_i = 1 \right\}. \tag{8}$$

The parameters $(\lambda_1, \lambda_2, \lambda_3)$ are the barycentric coordinates of p (the relative centroids) in relation to (p_{i-1}, p_i, p_{i+1}). For $(p, p_{i-1}, p_i, p_{i+1})$ with $p = (x, y)$ and $p_j = (x_{ji}, y_{ji})$, $j = i-1, i, i+1$, the parameters $(\lambda_{i-1}, \lambda_i, \lambda_{i+1})$ satisfying some initial conditions are solutions of the system represented below:

$$\begin{cases} \lambda_{i-1} x_{i-1} + \lambda_i x_i + \lambda_{i+1} x_{i+1} = x \\ \lambda_{ii-1} y_{i-1} + \lambda_i y_i + \lambda_{i+1} y_{i+1} = y \\ \lambda_{i-1} + \lambda_i + \lambda_{i+1} = 1 \end{cases}. \tag{9}$$

The determinant (Δ) of the solution matrix of the system above is the scalar $2S$,

$$\Delta = \begin{vmatrix} x_{i-1} & x_i & x_{i+1} \\ y_{i-1} & y_i & y_{i+1} \\ 1 & 1 & 1 \end{vmatrix} = 2S, \tag{10}$$

where S is the area of the triangle (p_{i-1}, p_i, p_{i+1}). The values of each one of the elements $(\lambda_{i-1}, \lambda_i, \lambda_{i+1})$ can be determined by Cramer's rule:

$$\lambda_{i-1} = \frac{S_{pp_i p_{i+1}}}{S_{p'_{-1} p_i p_{i+1}}} = \frac{S_{i-1}}{S}, \qquad \lambda_i = \frac{S_{p_{i-1} p p_{i+1}}}{S_{p_{i-1} p_i p_{i+1}}} = \frac{S_i}{S},$$
$$\lambda_{i+1} = \frac{S_{p'_{-1} p_i p}}{S_{p_{i-1} p_i p_{i+1}}} = \frac{S_{i+1}}{S}. \tag{11}$$

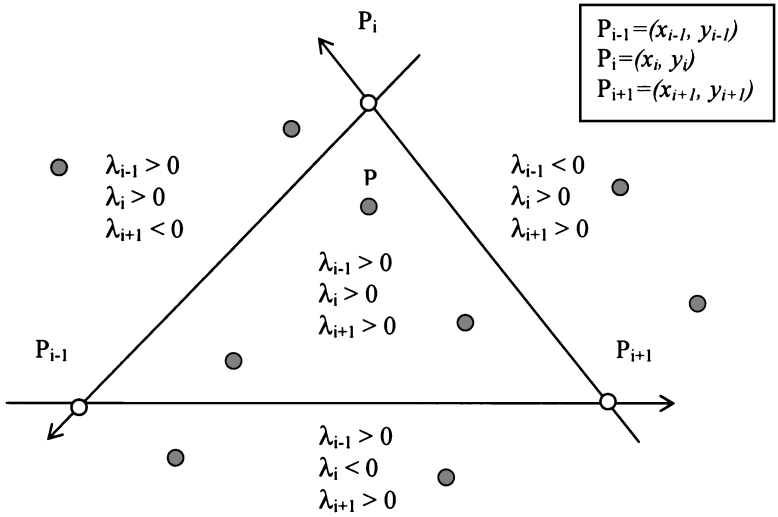


Figure 3. Oriented area for an ensemble of three events (P_{i-1}, P_i, P_{i+1}) and the relative barycentric coordinates (λ_{i-1}, λ_i, λ_{i+1}) to determine whether unsampled georeferenced points are inside the triangle defined by the three sampled events and the weights for the barycentric interpolation.

The system (9) determines oriented areas. The weights λ_j, $j = i-1, i, i+1$ will be positive if and only if S_j, $j = i-1, i, i+1$, and S has the same orientation (signal). On the basis of barycentric coordinates, it can be determined whether points are inside a triangle, because if a point is interior its barycentric coordinates are all positive (Fig. 3). The barycentric linear interpolation can be used to determine, for continuous phenomena, the unknown values at unsampled points in the spatial point pattern.

BARYCENTRIC INTERPOLATION

Neighbor relationships can also be weighted. Weights based on barycentric coordinates are the subject of this section. Given an ensemble $C = \{x_1, x_2, \ldots, x_m\}$ of events and real values $f(x_i)$, $i = 1, \ldots, m$, a piecewise linear function F, defined inside an adequate domain D, such as $F(x) = \sum f(x_i)$, $i = 1, \ldots, m$, can be obtained. The natural choice for this domain D is a conv(C). However, given a point $x \in \text{conv}(C)$, the calculation of $F(x)$ is not obvious. The basic idea is to write $x \in \text{conv}(C)$ as a disjoint union of an ensemble of triangles (the simplex). On the basis of the construction of this triangle network, given $x \in \text{conv}(C)$, it can be determined whether x belongs to a particular triangle p_{i-1}, p_i, p_{i+1}, and then $F(x)$ can be computed using Equations (9), (10), and (11).

The fundamental step in this approach consists of solving the related problem of the triangulation of an ensemble $C = \{x_1, x_2, \ldots, x_m\}$ based on the construction of conv(C). Notice that, on a two-dimensional space, the triangulation does not exhibit the property of unicity (i.e. there are several ways to triangulate a convex network). However, it is possible to determine the optimal number of triangles on the conv(C).

DATA SIMULATIONS AND ANALYSES

The data simulation, construction, and visualization of the two-dimensional convex, including the ordering of the points, construction of the triangulated network, execution of the Stat-Geo test, and computing the power function are accomplished using SPLUS™. For the simulation set-up, n points are generated inside an area of a two-dimensional space and m events are randomly sampled with which to test the CSR hypothesis. The power of the test is obtained through Monte Carlo simulation of two types of point processes.

Processes with irregular patterns are generated according to a Strauss process (Strauss, 1975). This process involves a parameter called the inhibition radius. Two events are considered neighbors if their distance is smaller than this parameter. Another parameter is the reflected inhibition, $0 \leq c \leq 1$; randomness is indicated when $c = 0$ (Fig. 4). The Strauss aggregated patterns are producing using the following procedure: the parent points are distributed randomly in the study region according to a homogeneous Poisson process with density λ per unit area; each parent independently produces a random number of children, following a Poisson distribution with mean $\mu \ll \lambda$. For each value of the parameters, 200 realizations of the punctual process are generated. In each realization, 20 sample points are independently chosen and the area is standardized. The sample sizes are selected in such way to keep the sample intensity smaller than 10%, guaranteeing that observations from different sample points can be considered independent (Diggle, Besag, and Gleaves, 1976). In the graphics "coordnova" denotes the barycentric coordinates and the "areat" the areas of each triangle of the network.

Illustration "I"

In this case a binomial process generates a spatially random pattern of 200 points within the given boundary. In essence it is a homogeneous Poisson process conditional on the given number of points. A random sample of 20 events is selected (Fig. 5). On the basis of the area of each triangle (total number of triangles is 30 on a convex constructed with 8 boundary vertices) the modified Kolmogorov-Smirnov test can be applied. The statistics of the one-sample Kolmogorov-Smirnov test with hypothesized distribution given by a uniform [0, $A = 0.932$] distribution is $ks = 0.2$, p value $= 0.5941$. The null hypothesis is not rejected (i.e., the true cumulative

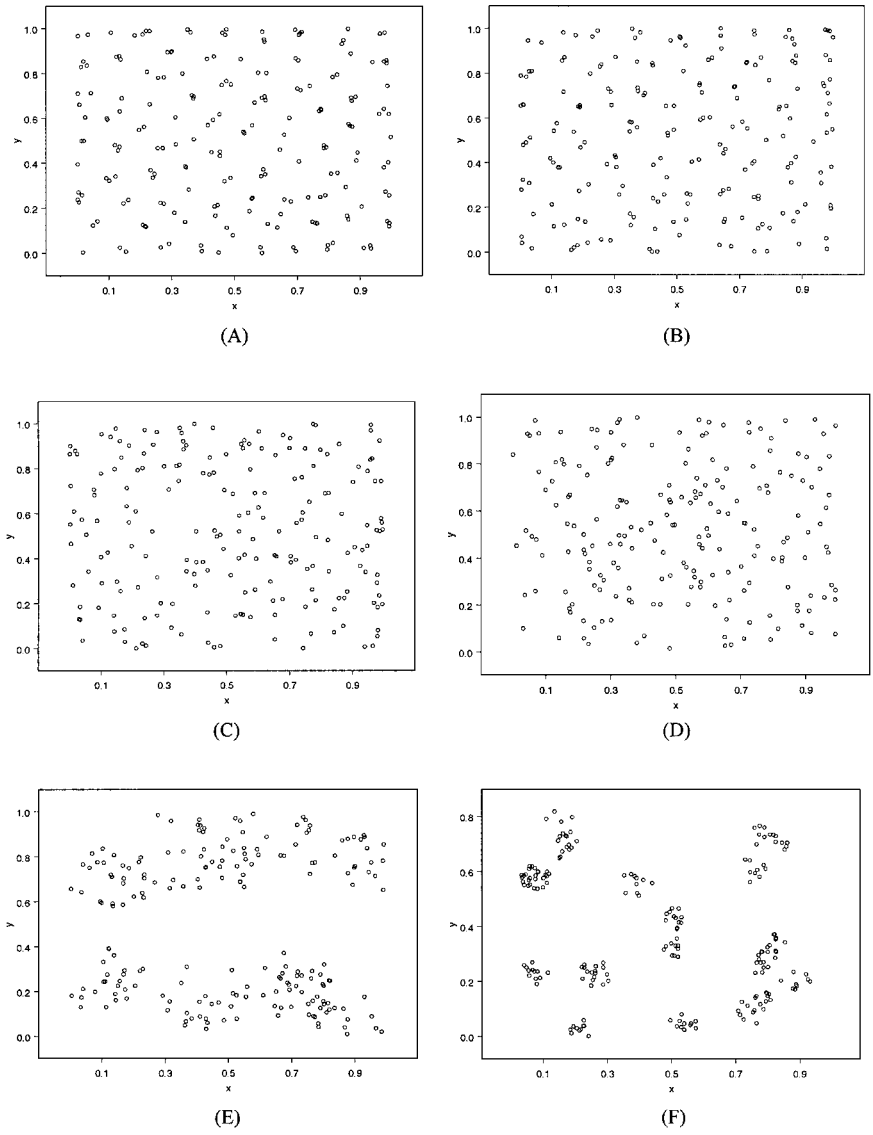


Figure 4. Generating spatial point patterns from a Strauss process of nonrandom point designs with the inhibition radius $r = 0.1$, from (A), (the inhibition parameter $c = 0$.) apparent regularity pattern to (*D*) (the inhibition parameter $c = 1$.) indicating clusters. Two Strauss aggregated cluster pattern on (E) and (F), where the spatial pattern configurations are immediately obvious.

distribution function of the triangulated network has $U \sim [0, A = 0.952]$, which means that there is evidence that the original process has a CSR pattern over the study region). For the alternative pure significance test, the critical value obtained is $2.165516 > 1.96$, which means that the hypothesis that the true value of the distribution parameter is 0.567, the area of the triangulated network, is rejected. Concerning the power function, the power for $\theta_0 = 0.567$ is $0.02037775 < 0.05$, which is acceptable (Fig. 5).

A robust first-order scheme based on barycentric coordinates is used to interpolate the observations at elementary cell vertices on a denser grid. For each unsampled location, the values are evaluated and updated by linear interpolation using the values at the vertices of the triangle. Notice that the precision of the linear interpolation can be estimated with the same properties as the kriging methodology; and without loss of generality the variogram model can be considered linear.

So, concerning the interpolation let $P = (0, 0)$. This point is clearly inside the triangle with vertices $P_{i-1} = (0.084, -0.0456)$, $P_i = (0.056, 0.276)$ and $P_{i+1} = (-0.074, 0.088)$, with area equal to 0.047. The barycentric coordinates satisfy the system (9), and the output of the program gives the solution (12): $\lambda_1 = 0.4$, $\lambda_2 = 0.07$, $\lambda_3 = 0.53$. Taking an associated continuous function into account, one can obtain an estimate for each point using barycentric interpolation. This value is given by the expression $T(x) = \sum_{j=i-1}^{i+1} \lambda_j T(x_j)$, where $T(x_j)$ is the attribute on P_j.

Illustration "II"

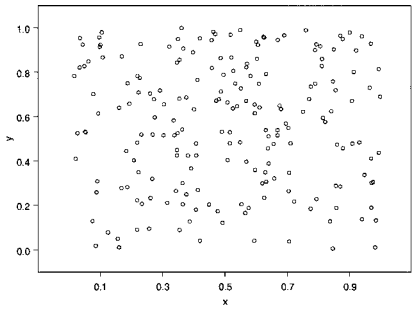
This is the case where a regular lattice process in $\Re^2$ generates a pattern of points within the given boundary. In essence this is a regular spatial process. A random sample of 20 events was selected (Fig. 6). The modified K-S test is applied to the empirical distribution of the triangle areas (total of 30) to test the hypothesis that the distribution is uniform $[0, A = 0.932]$. The test yields $ks = 0.8793$ and p value $= 0.0299$, suggesting no CSR pattern.

Illustration "III"

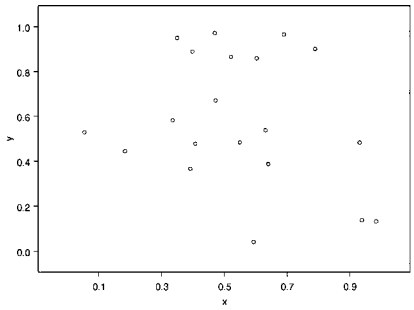
A random sample of 20 points is selected from an aggregated spatial process with a parent from a Poisson distribution with mean equal to 15 and a "big" radius ($r = 0.1$). The triangulation pattern resulting from the subdivision of the region under study (the simplex) is shown in Figure 7. Applying the modified K-S test to the empirical distribution of the triangle areas produces $ks = 0.8929$ and p value $=$ 0.0396. The hypothesis of CSR is rejected.

Illustration "IV"

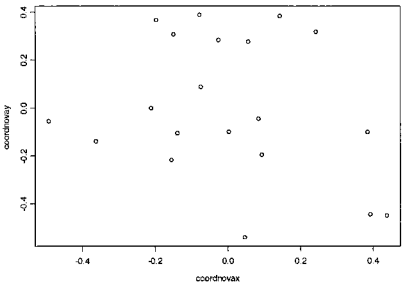
A random sample of 20 points is selected from a configuration generated by the Strauss process with "small" radius ($r = 0.01$), shown in Figure 8. The K-S test



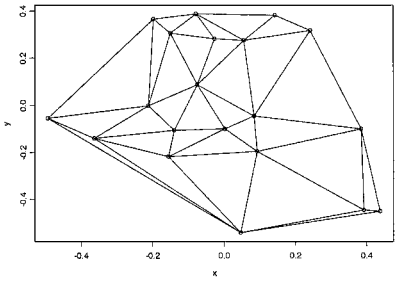
(A) $m = 200$, *total area = 0.931653*



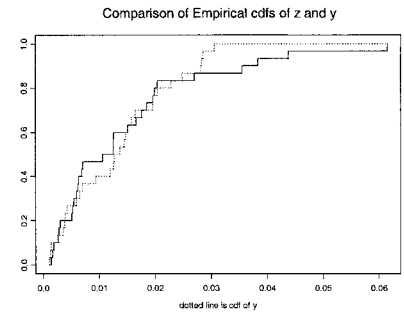
(B) $n = 20$, *polygon's area = 0.5670266*



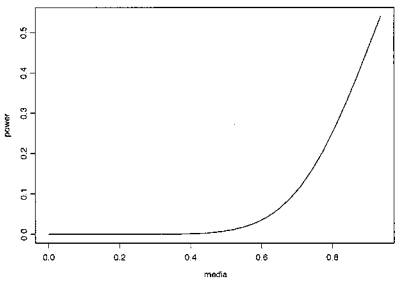
(C) *polygon's area = 0.5670266*



(D) *8 vertices* $\Rightarrow$ *number of triangles (ntriang) = 30*

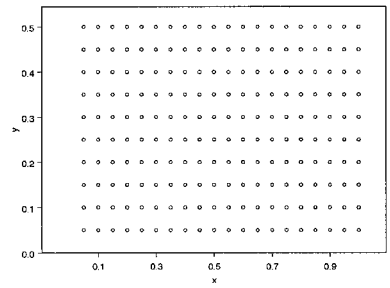


(E) *z = matrix(c(areat)), y ~ (ntriang,0,totalarea/(ntriang))*

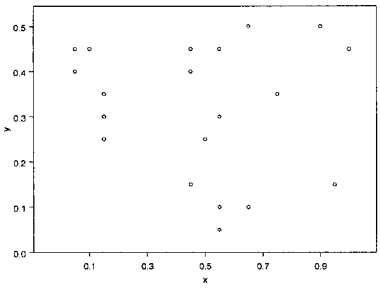


(F) *Power function*

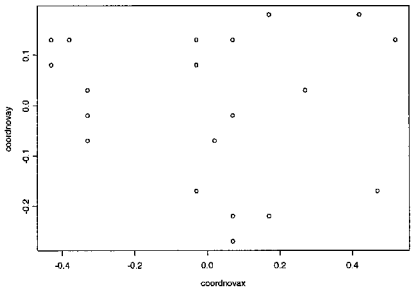
Figure 5. Simulating spatial-point patterns from a binomial process, to guarantee randomness. It is supposed to span a spatially random stochastic process (A), $n = 200$ points sampled (the intensity $\lambda = 214.67$ points per unity area) on (B), the $m = 20$ events ($\lambda = 35.27$ per unity area) under the new barycentric coordinates, on (C), the triangulated network, on (D), and on (E), the comparison between the empirical distribution of the triangle areas and the uniform distribution [0, 0.931653], on (F), the power function over the convex with area 0.57, where the supposed area for each triangle under uniform distribution is 0.0285.



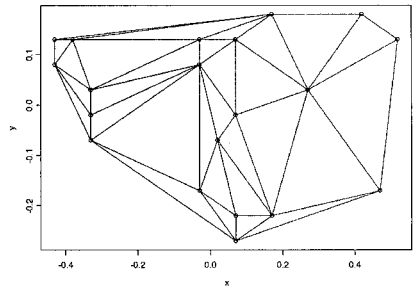
(A) *m=200, total area = 0.9322787*



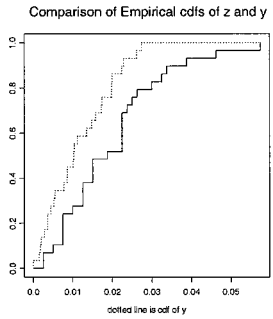
(B) *n=20, polygon's area=0.6006419*



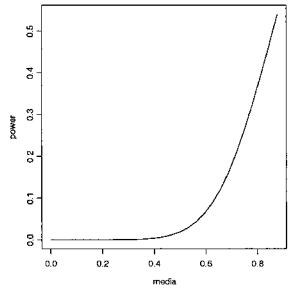
(C) *polygon's area=0.6006419*



(D) *8 vertices ⇒ number of triangles (ntriang) = 30*



(E) *z = matrix(c(areat)), y ~ (ntriang,0,totalarea/(ntriang))*



(F) *Power function*

Figure 6. Simulating spatial point patterns from a regular lattice process, to guarantee nonrandomness by means the regularity. It is supposed to span a spatially nonrandom stochastic process (A), $n = 200$ points sampled (the intensity $\lambda = 227.92$ points per unity area) on (B), the $m = 20$ events ($\lambda = 34.63$ per unity area) under the new barycentric coordinates, on (C), the triangulated network, on (D), and on (E), the comparison between the empirical distribution of the triangle areas and the uniform distribution [0, 0.932], on (F), the power function over the convex with area 0.60, where the supposed area for each triangle under uniform distribution is 0.02.

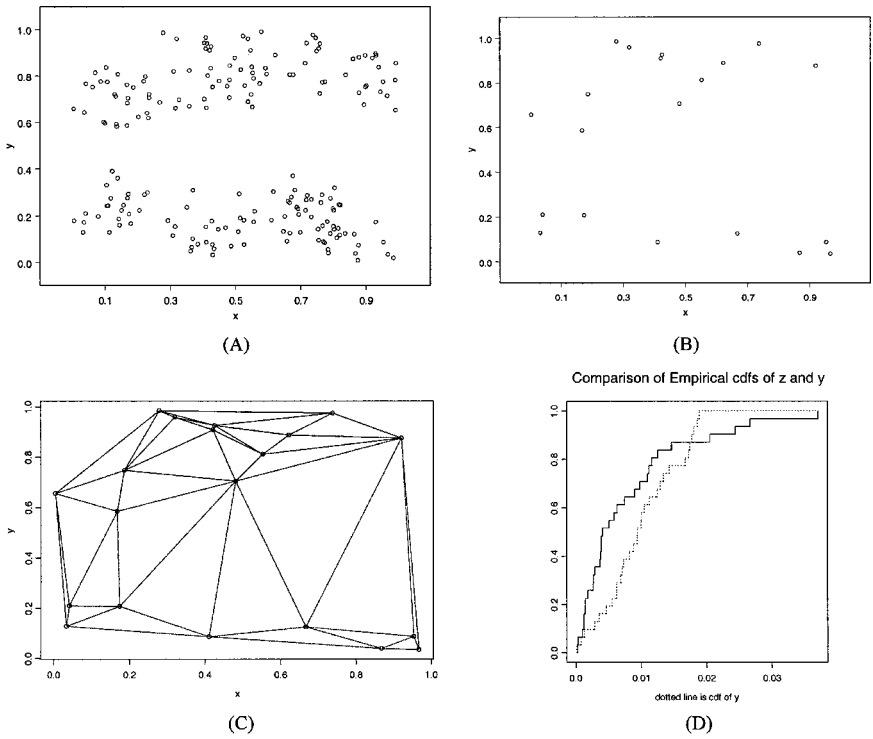


Figure 7. Simulating spatial-point patterns from a cluster process, to guarantee nonrandomness by means the spatially agglomerated pattern. It is supposed to span a spatially nonrandom stochastic process (A), $n = 200$ points sampled (the intensity $\lambda = 315.88$ points per unity area) on (B), the $m = 20$ events ($\lambda = 41.20$ per unity area), on (C), the triangulated network, on (D), the comparison between the empirical distribution of the triangle areas and the uniform distribution [0, 0.633].

of hypothesis that the triangle areas of the simplex have uniform distribution, yields $k\text{s} = 0.6333$ and p value $= 0.0065$. The hypothesis of CSR is rejected suggesting the test is capable of detecting the spatial nonrandomness even in complex patterns.

TWO APPLICATIONS

The preceding methodology leads to similar results when applied to "live" data situations.

Case-Study "I"—Geomorphology

In a situation involving limestone karst (hazard assessment associated with sinkhole occurring in several regions of California (Rowlingson and Diggle, 1993;

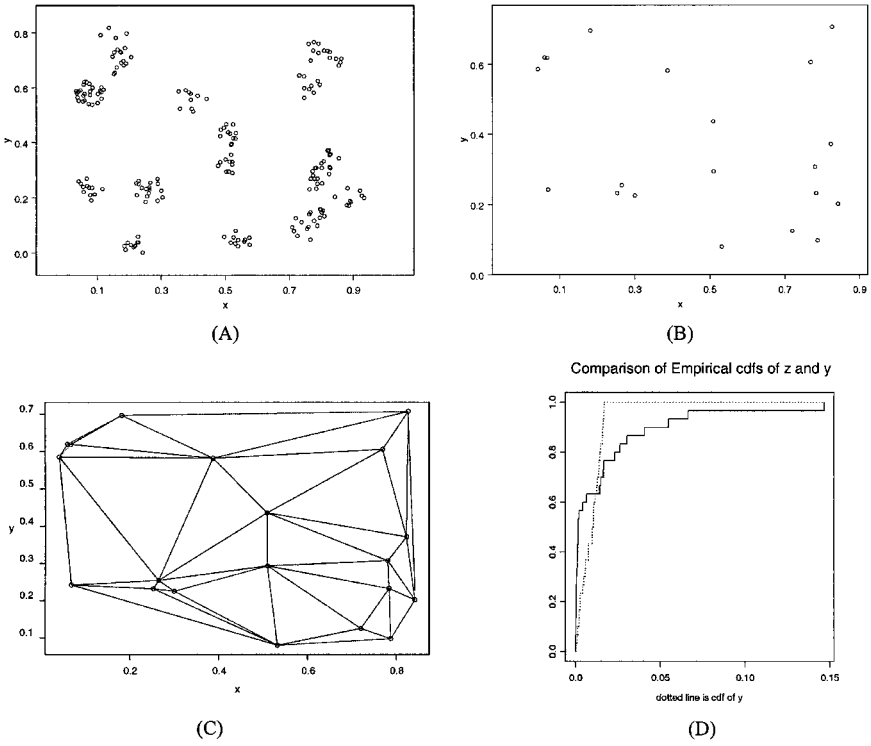


Figure 8. Simulating spatial-point patterns from a cluster process, to guarantee non-randomness by means the spatially agglomerated pattern. It is supposed to span a spatially nonrandom stochastic process (A), $n = 200$ points sampled (the intensity $\lambda = 368.04$ points per unity area) on (B), the $m = 20$ events ($\lambda = 41.89$ per unity area), on (C), the triangulated network, on (D), the comparison between the empirical distribution of the triangle areas and the uniform distribution [0, 0.596].

Vincent, 1987). An initial random sample of 200 events was selected (Fig. 9) and subsequently reduced to 20 events. The modified K-S test applied to the empirical distribution of the areas of the constructed triangles (total of 34) resulted in $ks = 0.8529$ and p value $= 0.01935$, when the underlying distribution was assumed to be uniform [0, $A = 545$]. The hypothesis of CSR was rejected.

Case-Study "II"—Seismology

A similar analysis was conducted on the coordinates of the earthquakes occurring in the San Francisco Bay area from 1962 to 1981.

Initially, a random sample of 200 events was selected (Fig. 10) and subsequently reduced to 20. Applying the modified K-S test to the empirical distribution

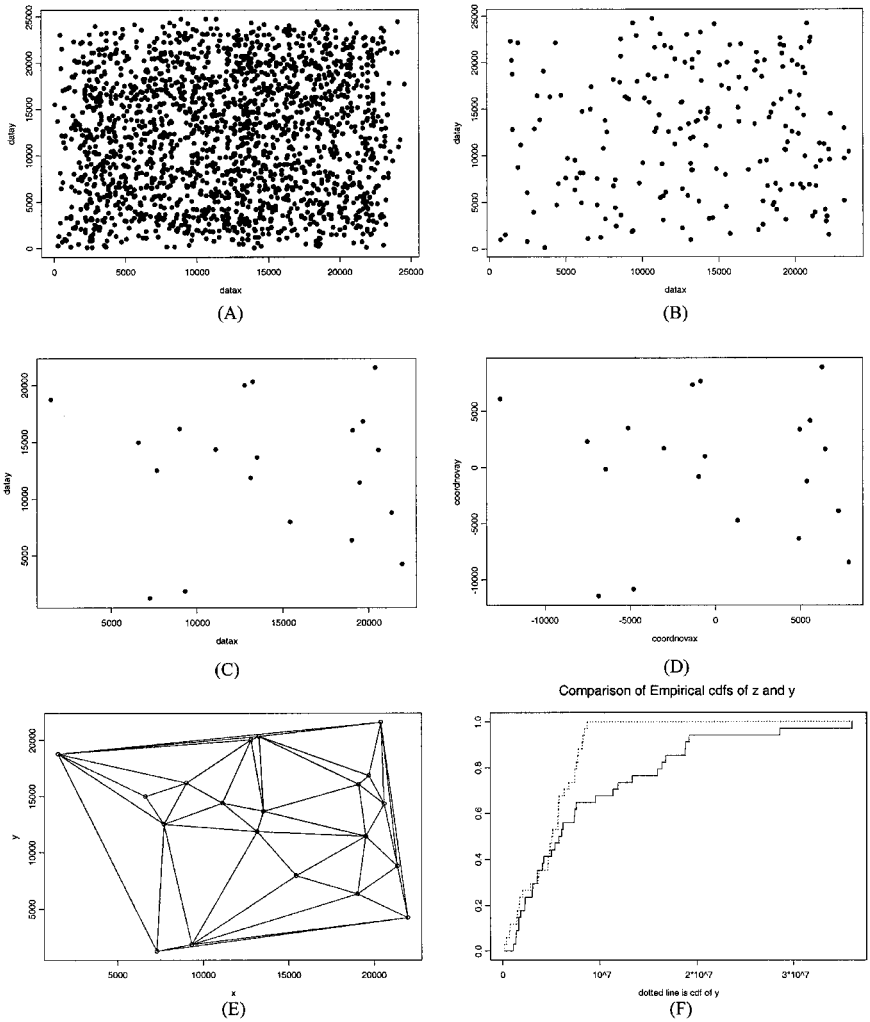


Figure 9. Pattern of sinkhole locations in several regions of limestone karst (rock consisting mainly of calcium carbonate) formations in a sinkhole database of 2360 points were sampled (A), into an established region and from 200 one has kept 20 events. The conclusion follows the simulation setup section. There is evidence that the original process has no CSR pattern.

of the constructed triangles (total of 34), again assuming the underlying distribution to be uniform $[0, A = 3{,}347]$, results in $k\text{s} = 0.529$ and p value $= 0.00282$. The null hypothesis is rejected, suggesting there is no evidence for CSR pattern, moreover the temporal quake processes is clustered in space.

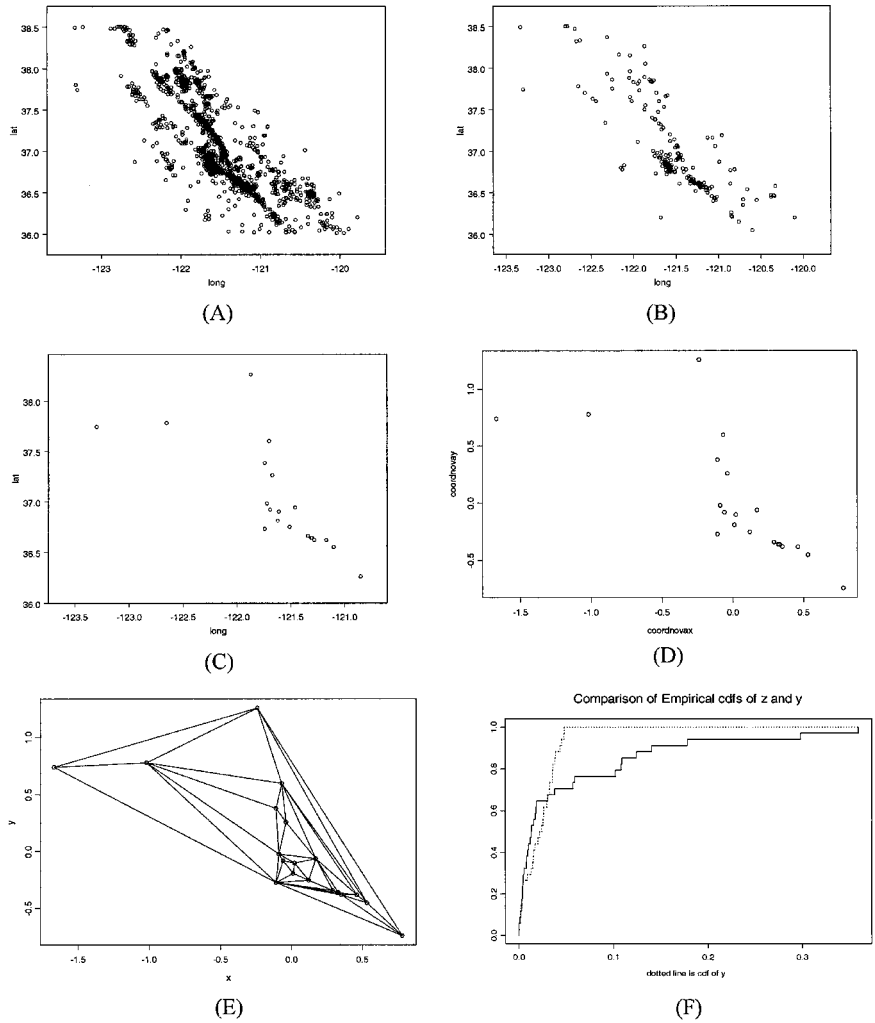


Figure 10. Pattern of the earthquakes occurring in the San Francisco Bay area from 1962 to 1981. The database of 2049 points were sampled A, into an established region over a target period and from 200 one has kept 20 representing the latest 3 years. The conclusion follows the simulation setup section. The null hypothesis is rejected, i.e., there is evidence that the original kept process (200 points) has no CSR pattern, besides the quake processes is clustering.

CONCLUSIONS

The asymptotic tests that have been proposed, present satisfactory results with a high accuracy, accepting the CSR hypothesis when the data has effectively this

configuration, and rejecting the CSR otherwise. This approach seems to be robust and offer a high efficiency results for a low computational cost in comparison with the classical and traditional CSR detection. Moreover, our method allows us to determine an interpolation criterion, similar to kriging methodology since the variogram has to be linear, based on the barycentric coordinates in influence zones.

ACKNOWLEDGMENT

The authors thank the referees. Their comments were very useful in improving this paper.

REFERENCES

Assunção, R., 1994, Testing spatial randomness by means of angle: Biometrics, v. 50, p. 531–537.

Assunção, R, and Reis, I. A., 2000, Testing spatial randomness: A comparison between T^2 methods and modifications of angle test: Braz. J. Prob. Stat., v. 14, p. 71–86.

Bailey, T. C., and Gatrel, A. C., 1995, Interactive spatial data analysis: Longman Scientific & Technical, London, 413 p.

Besag, J. E., and Cleaves, J. T., 1973, On the detection of spatial pattern in plant communities: Bull. Int. Stat. Inst., v. 45, p. 153–158.

Besag, J., and Newell, J., 1991, The detection of clusters in rare diseases: J. R. Stat. Soc. A, v. 154, p. 143–155.

Bonetti, M., and Pagano, M., 2001, On detecting clustering, *in* Proceedings of the Biometrics Section: American Statistical Association, p. 24–33.

Diggle, P. J., 1983, Statistical analysis of spatial point patterns: Academic Press, London, 148 p.

Diggle, P. J., Besag, J. E., and Gleaves, J. T., 1976, Statistical analysis of spatial point patterns by means of distance methods: Biometrics, v. 32, p. 659–667.

Farin, G. E., 1993, Curves and surfaces for computed aided geometrical design, 3rd ed.: Academic Press, London, 444 p.

Gatrell, A., Bailey, T., Diggle, P., and Rowlingson, B., 1996, Spatial point pattern analysis and its application in geographical epidemiology: Trans. Inst. B. Geogr., v. 21, p. 256–274.

Hines, W. G. S., and O'Hara Hines, R. J., 1979, The Eberhardt statistic and the detection of non-randomness of spatial point distributions: Biometrika, v. 66, p. 73–79.

Okabe, A., and Yamada, I., 2001, The K function method on a network and its computational implementation: Geogr. Anal., v. 33, p. 271–290.

Ripley, B. D., 1976, The second-orders analysis of stationary point processes: J. Appl. Prob., v. 13, p. 225–266.

Rogerson, P. A., 1999, The detection of clusters using a spatial version of the chi-square goodness-of-fit statistic: Geogr. Anal., v. 31, p. 130–147.

Rowlingson, B. S., and Diggle, P. J., 1993, Splancs: Spatial point pattern analysis code in S-Plus: Technical Report, Dep. of Mathematics and Statistics, Lancaster University, Lancaster, UK.

S+SpatialStats, 2000, Professional Release 3: Mathsoft, Seatlet, Washington, DC.

Strauss, D. J., 1975, A model for clustering: Biometrika, v. 62, p. 467–475.

Venables, W. N., Ripley, B. D., 1999, Modern applied statistics with S-Plus, 4th edn.: Springer, New York, 495 p.

Vincent, P. J., 1987, Spatial dispersion of polygonal Karst Sinks: Zeitschrift fur Geomorphologie, v. 31, p. 65–72.

APPENDIX

Often, decisions have to be made in situations where spatial variation plays an important role and it is crucial to determine the range of an intrinsic influence concerning cluster-shape patterns.

On the basis of the scope of the statistical test, the null hypothesis H_0 of CSR is founded on the supposition that the spatial stochastic process generated has a Poisson probability function $Y \sim \wp(\lambda|A|)$ that describes the number of points on the domain A with intensity λ, i.e. the standard models for CSR assume that the number of events $Y(A)$ follows a homogeneous Poisson process over the region A, with expectancy equal to $E(Y(A)) = \lambda|A|$. Considering a stationary process $Y(A_i) \sim \wp(\lambda|A_i|)$, the maximum-likelihood estimator (MLE) of the parameter $\lambda|A_i|$ is the sample mean $\bar{Y}(A_i)$. One can estimate the parameter $\lambda(A)$, using statistical tests.

Taking into account the asymptotic properties of the maximum-likelihood estimators, $\hat{\lambda}|A| \sim N(\lambda|A|, I^{-1}_{\lambda|A|})$ where $I_{\lambda|A|}$ is the Fisher's information, under H_0

$$Z = \frac{\hat{\lambda}|A| - \lambda|A|}{\sqrt{I_{\lambda|A|}}} \sim N(0, 1), \tag{A1}$$

with $I_{\lambda|A|} = \frac{\lambda|A|}{m}$ and m the number of events in A. The rejection of H_0 happens if

$$\frac{\left|\bar{Y}|A| - \lambda_0|A|\right|}{\lambda_0|A|}\sqrt{m} > k_\alpha, \tag{A2}$$

where $\hat{\lambda}|A|_{\mathrm{MLE}} = \bar{Y}|A|$ the intensity into the area A, λ_0 is the value of λ under H_0 and k_α is the critical value preestablished depending on the significance level α of the test, such as

$$\alpha = P\left(\frac{\bar{Y}|A| - \lambda|A|}{\sqrt{\lambda|A|}}\sqrt{m} > k_\alpha|\mathrm{H}_0\right). \tag{A3}$$

The power function of this test is given by:

$$\eta(\lambda|A|) = P\left(Z > c + \frac{\bar{Y}|A| - \lambda|A|}{\sqrt{\lambda|A|}}\sqrt{m}||\lambda|A|\right). \tag{A4}$$

In fact, one is interested in the null hypothesis H_0, from which a certain function $F_0(x)$ is the spatial distribution of a triangulated punctual (network) population. Such a hypothesis arises from a theoretical consideration, as a Poisson spatial process.