

Research Article

Characterizing the Relative Spatial Structure of Point Patterns

Eric Marcon,¹ Florence Puech,² and Stéphane Traissac¹

¹AgroParisTech, UMR EcoFoG, BP 709, 97310 Kourou, French Guiana

²LET (Université de Lyon, CNRS, ENTPE), Institut des Sciences de l'Homme, 14 avenue Berthelot, 69363 Lyon Cedex 07, France

Correspondence should be addressed to Eric Marcon, eric.marcon@ecofog.gf

Received 26 May 2012; Revised 23 July 2012; Accepted 21 August 2012

Academic Editor: Cajo J. F. ter Braak

Copyright © 2012 Eric Marcon et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We generalize Ripley's K function to get a new function, M , to characterize the spatial structure of a point pattern relatively to another one. We show that this new approach is pertinent in ecology when space is not homogenous and the size of objects matters. We present how to use the function and test the data against the null hypothesis of independence between points. In a tropical tree data set we detect intraspecific aggregation and interspecific competition.

1. Introduction

Investigating the spatial structure of point patterns has been a long-time challenge for ecologists. Pielou [1] claimed that the information an ecologist wants to have immediately when observing a point set representing a vegetal community is the density of each species and the existence of interactions between plants. Density can be estimated by various methods [2] but interactions have motivated a living literature for more than half a century. In ecology, Ripley's K function [3], or its square-root transformation L [4], is the most used tool to characterize them [5], assuming the pattern is a realization of a homogenous point process; that is to say the probability to find a point is the same everywhere.

However, identifying interactions under the assumption of nonhomogeneity of space is still an open question. Twenty years ago, Cuzick and Edwards [6] followed by Diggle and Chetwynd [7] paved the way by introducing some specific tests. The D function proposed by Diggle and Chetwynd is defined as the difference between the K function for the studied points (called cases) and the K function for others (called controls). It is not completely satisfactory yet because both K functions are computed separately so all the data contained in the relative position of cases and controls is lost. A more recent important advance was proposed by Baddeley et al. [8] who generalized K to inhomogeneous point processes. They developed a complete theoretical framework but practical applications are still difficult because assumptions are

necessary about the relative scales of heterogeneity and interactions, leading to possibly opposite analyses of the results [9]. A recent review of these methods can be found in [5].

Other tools were also developed by economists [10] with a different approach, comparing the distribution of points of interest relatively to that of other points. Brühlhart and Traeger [11] call them *relative* measures, as opposed to *topographic* measures such as K which take space (measured by areas) as their benchmark.

In this study, we introduce a relative measure of spatial structure, namely, the M function [12] to extend the ecologists' toolbox and allow a more pertinent approach when the null hypothesis is that points of interest are distributed like others. It bypasses the issue of heterogeneity and allows weighting points. Moreover, its computation is easy.

The paper is organized as follows: first, we derive the M function as a generalization of Ripley's K function; then, we apply it to theoretical examples and real data sets in a tropical forest and in epidemiology; we finally discuss the way it can be usefully applied after clarifying the assumptions it relies on.

2. Methods

2.1. Ripley's K Function

2.1.1. Definition. The theoretical framework is a point process whose realization is observed in a window of area

A. Nonparametric methods such as Ripley's K are used to reject the null hypothesis of independence of points. To use Ripley's K , we will assume that the point process is *stationary* (i.e., its intensity does not vary by translation). All point processes used in this paper are *second-order stationary* (i.e., interactions between points do not vary under translation) and *isotropic* (i.e., they do not depend on direction). The null hypothesis to reject is therefore complete spatial randomness (CSR); that is, the point pattern is a realization of a homogenous Poisson process. More details on K can be found in [5] or [13]. We focus on its estimator here.

Points are denoted x . We call a point x 's neighbors all the points less than r apart from it (all points in a disk of radius r centered on the point x).

Ripley's K function estimator was built by counting neighbors (indexed by n) around reference points (indexed by f), which can belong to the same type (*univariate* K function) or not (*intertype* or *bivariate* K function). N points are found in the window, and we denote $\mathbf{1}(\|x_f - x_n\| \leq r)$ the indicator function equal to 1 if the distance between x_f and x_n is less than or equal to r , 0 else.

An unbiased estimator of univariate K [14] with no edge-effect correction is

$$\hat{K}(r) = \frac{A}{N(N-1)} \sum_{f=1}^N \sum_{n=1, n \neq f}^N \mathbf{1}(\|x_f - x_n\| \leq r). \quad (1)$$

The bivariate version of K (denoted $K_{f,n}$) is very similar. We denote N_f the number of reference points and N_n the number of neighbors. We have

$$\hat{K}_{f,n}(r) = \frac{A}{N_f N_n} \sum_{f=1}^{N_f} \sum_{n=1}^{N_n} \mathbf{1}(\|x_f - x_n\| \leq r). \quad (2)$$

2.1.2. Edge-Effect Correction. Points located close to the window borders are problematic because a part of the circle inside which points are supposed to be counted is outside the window. Various answers have been proposed to correct for this [15, 16]. We prefer Besag's [4] correction. Let us denote A_{fr} the part of the area of the circle of radius r centered on the point x_f located inside the window. We count the number of neighbors inside the circle, and we correct it by the ratio between the circle's area and its inside part. We suppose that the outside part of the circle would have contained the same neighbor density than the inside part. Finally, an unbiased estimator of K with edge-effect correction is

$$\hat{K}(r) = \frac{A}{N(N-1)} \sum_{f=1}^N \sum_{n=1, n \neq f}^N \mathbf{1}(\|x_f - x_n\| \leq r) \frac{\pi r^2}{A_{fr}}. \quad (3)$$

2.1.3. Normalization. Besag [4] proposed to normalize K to obtain a benchmark of r rather than πr^2 . The well-known L function is defined as $L(r) = \sqrt{\hat{K}(r)/\pi}$. It can be interpreted as a distance [17]: $L(r) = r + l$ means that as many neighbors are found around reference points up to distance r as would be expected at distance $r + l$ under CSR. We believe that $\hat{K}(r)/\pi r^2$ is a better normalization. Its reference value is 1, and it can be interpreted as the density of neighbors around reference points divided by the density of neighbors anywhere.

2.2. The M Function

2.2.1. Definition of M . Equation (3) can be rearranged:

$$\frac{\hat{K}(r)}{\pi r^2} = \frac{\left[(1/N) \sum_{f=1}^N \left(\sum_{n=1, n \neq f}^N \mathbf{1}(\|x_f - x_n\| \leq r) / A_{fr} \right) \right]}{(N-1)/A}. \quad (4)$$

Around each point x_f , $[\sum_{n=1, n \neq f}^N \mathbf{1}(\|x_f - x_n\| \leq r)] / A_{fr}$ is the number of neighbors divided by the area where it is counted. Its average value is compared to what it is expected to be all over the window, $(N-1)/A$.

Topographic measures like K use space as their benchmark; that is, the number of points is divided by an area. The benchmark may also be another point pattern; for example, the number of trees of a species under study may be divided by the total number of neighbor trees, defining *relative* measures.

We transpose K into a relative framework. The ratio is now built comparing a number of neighbors of interest to the total number of neighbors. Weights can be associated to points without changing the construction of the measure. Reference points are indexed by f (x_f is a reference point), neighbor points by n ; all points whatever their type (i.e., the benchmark) by a ; their numbers are N_f , N_n , and N_a . w_i is the weight of point x_i , $W_i = \sum_{i=1}^{N_i} w_i$ is the total weight of this type of points.

The average weighted ratio of neighbor points around reference points is $(1/N_f) \sum_{f=1}^{N_f} (\sum_{n=1, x_n \neq x_f}^{N_n} \mathbf{1}(\|x_f - x_n\| \leq r) w_n / \sum_{a=1, x_a \neq x_f}^{N_a} \mathbf{1}(\|x_f - x_a\| \leq r) w_a)$.

In the whole window, the same ratio is $(1/N_f) \sum_{f=1}^{N_f} ((W_n - w_f) / (W_a - w_f))$ if neighbor points and reference points belong to the same type, $(1/N_f) \sum_{f=1}^{N_f} (W_n / (W_a - w_f))$ else. If reference and neighbor points belong to the same type, $N_f = N_n$ and $W_f = W_n$.

We define the univariate M function as

$$M(r) = \frac{\sum_{f=1}^{N_f} \left(\sum_{n=1, x_n \neq x_f}^{N_n} \mathbf{1}(\|x_f - x_n\| \leq r) w_n / \sum_{a=1, x_a \neq x_f}^{N_a} \mathbf{1}(\|x_f - x_a\| \leq r) w_a \right)}{\sum_{f=1}^{N_f} \left((W_f - w_f) / (W_a - w_f) \right)}. \quad (5)$$

The bivariate $M_{f,n}$ function is

$$M_{f,n}(r) = \frac{\sum_{f=1}^{N_f} \left(\sum_{n=1}^{N_n} \mathbf{1}(\|x_f - x_n\| \leq r) w_n / \sum_{a=1, x_a \neq x_f}^{N_a} \mathbf{1}(\|x_f - x_a\| \leq r) w_a \right)}{\sum_{f=1}^{N_f} (W_n / (W_a - w_f))}. \quad (6)$$

Equations (5) and (6) are simplified when points are not weighted:

$$M(r) = \frac{N_a - 1}{N_f(N_f - 1)} \sum_{f=1}^{N_f} \frac{\sum_{n=1, x_n \neq x_f}^{N_n} \mathbf{1}(\|x_f - x_n\| \leq r)}{\sum_{a=1, x_a \neq x_f}^{N_a} \mathbf{1}(\|x_f - x_a\| \leq r)},$$

$$M_{f,n}(r) = \frac{N_a - 1}{N_f N_n} \sum_{f=1}^{N_f} \frac{\sum_{n=1}^{N_n} \mathbf{1}(\|x_f - x_n\| \leq r)}{\sum_{a=1, x_a \neq x_f}^{N_a} \mathbf{1}(\|x_f - x_a\| \leq r)}. \quad (7)$$

2.2.2. Case-Control Design. A particular attention must be paid to case-control designs. In practical terms, all points of interest (called *cases*) are carefully referenced, and the

benchmark point set (called *controls*) is just sampled. This approach has been widely used for spatial clustering of diseases [7, 9, 18–20, among others]: sick people are the *cases* and the rest of the population the *controls*. Case-control design is of course not limitative to geographical epidemiology and can easily be applied to ecology questions.

The M function defined previously can be slightly modified to take into account this feature. Since the controls are chosen to be a representative sample of the population at every scale, neighbors of any kind are replaced by controls, indexed by a . Reference and neighbor points are N_c cases; their total weight is W_c . After simplifications, M_{cases} can be written as follows:

$$M_{\text{cases}}(r) = \frac{\sum_{f=1}^{N_c} \left(\sum_{n=1, x_n \neq x_f}^{N_c} \mathbf{1}(\|x_f - x_n\| \leq r) w_n / \sum_{a=1}^{N_a} \mathbf{1}(\|x_f - x_a\| \leq r) w_a \right)}{W_c(N_c - 1)/W_a}. \quad (8)$$

2.2.3. Significance. The first-order property (intensity) of the process must be controlled to allow the detection of the second-order property (nonindependence of points, that is to say interactions between the objects they represent). Thus, a point distribution generated according to the null hypothesis must respect, on the one hand, the local values of the density of the process the point distribution is a realization of and, on the other hand, its points must be distributed independently from each other.

The practical difficulty comes from the lack of knowledge of the point process that gave the point distribution, which is its unique available realization. Its first-order property is consequently widely unknown. We can only assume that the actual set of point locations is a good approximation of it, following Duranton and Overman [10]. Consequently, we generate random data sets for the univariate and case-control M functions by redistributing the actual point set (type and weight couples) on the actual location set (coordinates). The confidence interval of the null hypothesis is then computed by the Monte Carlo technique [21].

The intertype function must support two null hypotheses [22]. The random labeling hypothesis is simulated by permuting the point types, keeping point locations and weights unchanged. The population independence hypothesis is more complex to test. The reference points are kept unchanged, so that the spatial structure of the reference point type is maintained, and all other points are redistributed across the available locations. This allows

testing the independence of populations considering the structure of the reference point type. Then, the reference and neighbor types are interchanged and the test is repeated. If both $M_{f,n}$ and $M_{n,f}$ functions leave their null hypothesis confidence envelope in a range of distances, then population independence is rejected. This test requires that some points do not belong to either the reference or the neighbor point type or there will be nothing to redistribute. More generally, testing the relative spatial structure only makes sense if the tested point types are a small part of the point pattern; see discussion Section 4.1.

The tests based on Monte Carlo simulations are actually not correct because they are repeated at each step of the function (see [23] for an extensive discussion). A global test, without the need of simulations, is available only for K against CSR [14]. We can follow Loosmore and Ford's goodness-of-fit (GoF) test to obtain a correct P -value to reject the null hypothesis. We first need to compute the average value of $M(r)$ on all simulations, more exactly [24]:

$$\bar{M}(r) = \frac{1}{s-1} \sum_{i=1}^s M_i(r), \quad (9)$$

where s is the number of simulations and $M_i(r)$ the value of $M(r)$ in the i th simulation. Then, the statistic u_i is calculated for the i th simulation by summing on all values of r , where δr

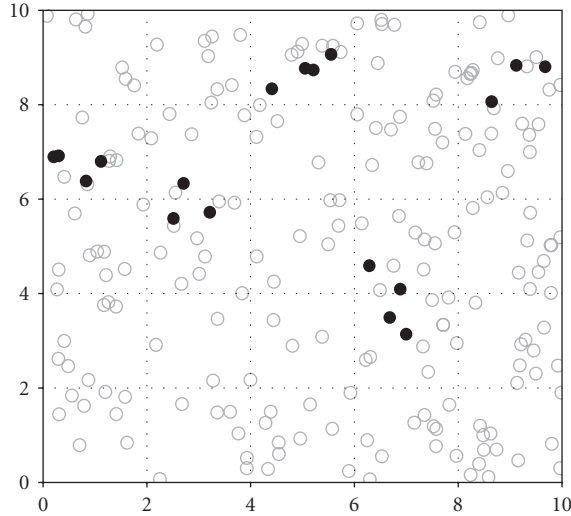


FIGURE 1: Aggregates, Point map. Grey circles are drawn from a homogenous Poisson process. Black disks are generated by a Matérn (radius = 0.5) process.

is the difference between the next value of r and the present one:

$$u_i = \sum_r [M_i(r) - \bar{M}(r)]^2 \delta r. \quad (10)$$

The same statistic for the actual data, denoted u , is compared to the simulated values to get a P -value:

$$P_u \approx \frac{\sum_{i=1}^s \mathbf{1}(u_i > u)}{s}. \quad (11)$$

If u is greater than all simulated values, the P -value to reject the null hypothesis erroneously is around 0. To avoid 0 or 1 P -values, we can assume that another simulation would have given a value of u_i higher or lower than u and write $P_u < 1/s$ or $P_u > 1 - 1/s$.

2.3. Examples. Three theoretical examples are given. Two of them illustrate very simple point patterns on a homogeneous space for a comparison of L and M functions (Sections 3.1.1 and 3.1.2). The third one computes an inhomogeneous Poisson point process to show how the M function controls for the first-order property of point processes (Section 3.1.3). No theoretical example is given with weighted points because they are not so easy to understand visually. Three real point patterns are considered then. They do not allow a classical analysis by the K function because of heterogeneity. Cuzick and Edwards [6] introduced the first formal way to deal with nonhomogeneous point processes: they used a dataset (published with the paper) concerning the location of 62 cases of childhood leukemia between 1974 and 1986 in the North Humberside area, England. A control set of 141 children representing the whole concerned population was chosen from the birth register (all weights are 1). They could conclude that the cases were significantly clumped. We use this data set to go further: we are now able to corroborate

their conclusion and also to precise the size of aggregates. The M function is computed according to the case-control design, (8).

We overall want to provide evidence of the interest of relative spatial structure in ecology. Trees are considered in a 25 ha plot of tropical rainforest in Paracou field station in French Guiana [25]. We investigate the spatial structure of two species, *Vouacapoua americana* Aublet (Caesalpinaceae) and *Qualea rosea* Aublet (Vochysiaceae) in a point set of 11,276 trees above 10 cm diameter at breast height (DBH), excluding flooded zones. All trees above 1 cm diameter have been measured and plotted for a few species, allowing us to study the spatial relations between saplings (up to 10 cm DBH) and possibly reproductive trees (30 cm or more) of *V. americana*. Points are weighted by the basal area of the tree they represent, the reference and neighbor points are mentioned in the results, and all trees of the maps, including references and neighbors, are used as the benchmark.

3. Results

3.1. Theoretical Examples. In what follows, we generate a point pattern (“black points,” represented by closed circles in the figures) to investigate its spatial structure with the L and M univariate functions. Other points (“grey points” represented by grey open circles) are used by M only: grey and black points together constitute the benchmark. Black points may be considered as trees of a species of interest in a forest plot, while grey points are all other trees.

All confidence intervals are computed at 1% risk level generated from 10,000 simulations.

3.1.1. Aggregates. 200 grey points are completely randomly distributed. Black points are generated by a Matérn process [26]: 5 aggregates (radius 0.5) of 5 points. All point weights equal 1. The map is in Figure 1; the curves are in Figure 2.

The M curve shape is similar to L ’s: significant, positive peaks denote concentration. The benchmark points of the M function are distributed almost homogeneously so the number of neighbors around each point is proportional to the area: the relative and the topographic measures are nearly equivalent. Nevertheless, while L peaks approximately correspond to the diameter of aggregates [27], M peaks occur exactly at distances at which the local density is the greatest, that is, approximately the distance between points in the aggregates. The differences are due to the way L is normalized: $M(r)$ peaks occur at the same distance as those of $K(r)/\pi r^2$ (not shown on the figure).

3.1.2. Regularity. 200 grey points are drawn from a homogeneous Poisson process again. 100 black points have a regular distribution around a square, 1-by-1 grid, with a perturbation: each point is randomly moved horizontally and vertically within a 0.4 interval around the grid nodes (Figure 3). All point weights equal 1.

The first part of the univariate M curve (Figure 4) is made of 0 values, showing the absence of neighbors at any distance up to 0.6. Note that the univariate L curve shape is

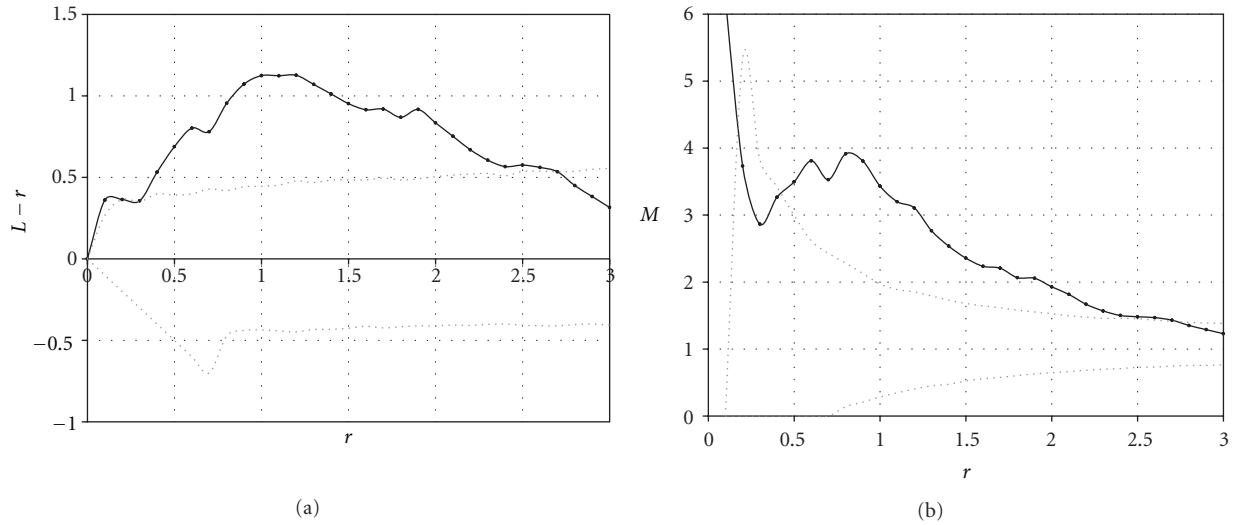


FIGURE 2: Aggregates, univariate L and M functions for the aggregated point set. Solid curves are the L and M function values; dotted lines are the envelope of the confidence interval of the null hypothesis. Both functions detect clumping. $L(r) - r$ is plotted rather than $L(r)$ for convenience.

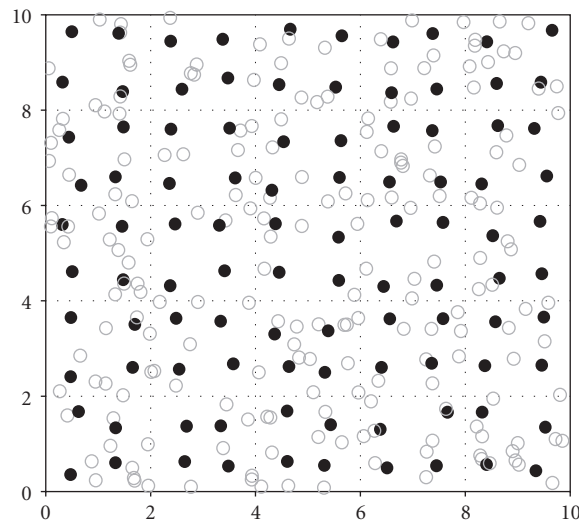


FIGURE 3: Regular point set, Point map. Grey circles are from a Poisson process. Black disks are located close to a 1×1 square grid.

different since its original value is 0 and its minimum slope is -1 by construction.

Negative peaks of both the univariate L and M functions allow detecting the grid size (before $r = 1, 2$, and 3). Maximum values correspond to the diagonal of the grid ($\sqrt{2} \approx 1,44$ is the diagonal length, then $\sqrt{5} \approx 2,24$ and so on).

3.1.3. Inhomogeneous Point Set. We generated two completely random point sets in a 10-by-10 window. Then, we transformed the point coordinates: after having calculated the polar coordinates (r, θ) of each point from the center of the point set, we squared the distance to get (r^2, θ) . The result is a nonhomogeneous Poisson pattern, shown in Figure 5. Both point types have the same random distribution, but the center of the map shows a greater density.

It can be seen (Figure 6) that the L function is not applicable: assuming homogeneity, it interprets the black point distribution as a single big aggregate. This issue is known as “virtual aggregation” [28, 29]. The M function is able to control for density variations: since the pattern of the black points does not differ from that of all points, M values are around 1.

3.2. Cuzick and Edwards Data Set. The case-control M function is used (Figure 7). 0.7 km apart from cases, the average case density is about 70% higher than it would be if the cases followed the control pattern (at this distance, the peak of the M function reaches 1.7).

In the discussion of [6], Diggle (page 101) suggested that the D function, equal to $K_{\text{cases}} - K_{\text{controls}}$, would be appropriate

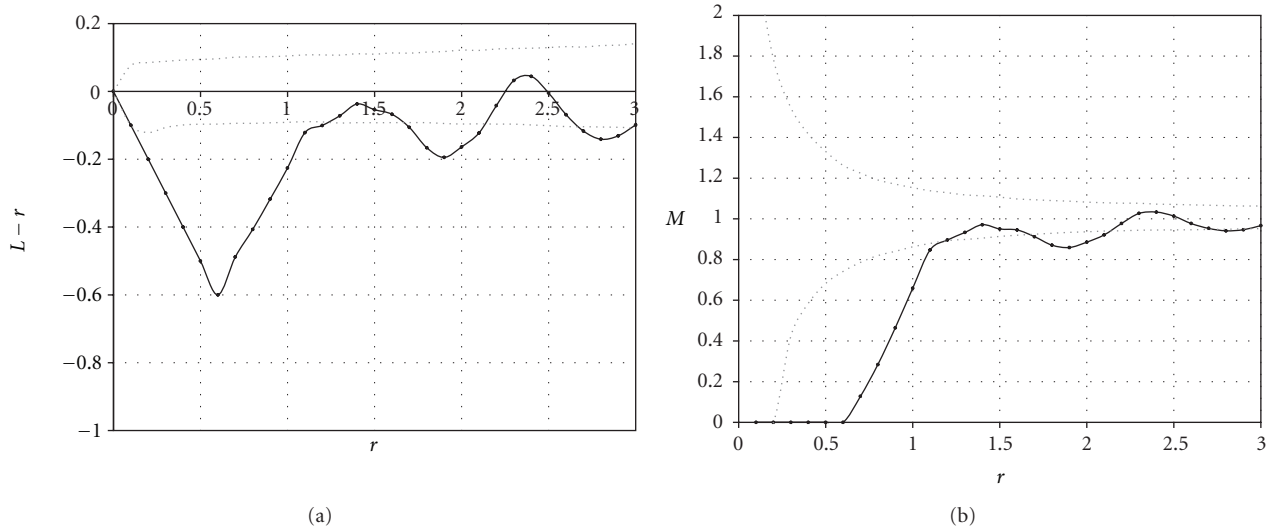


FIGURE 4: Regular point set, univariate L and M functions for the regular point set. Both functions detect dispersion.

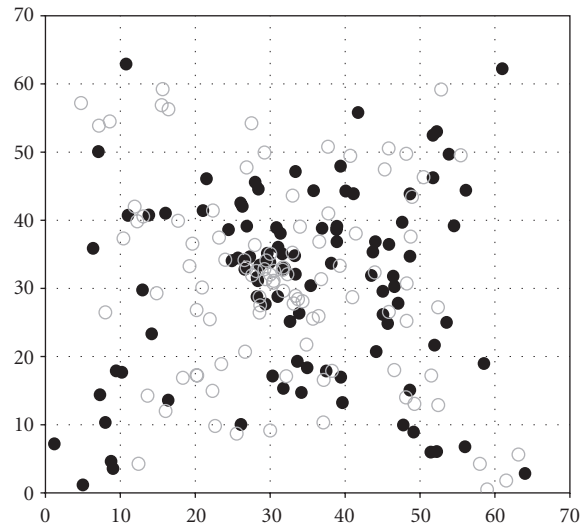


FIGURE 5: Inhomogeneous point set, Point map. All points are drawn from an inhomogeneous Poisson point process.

for this point pattern. The next year, Diggle and Chetwynd [7] published the D function and applied it to the same dataset. We recomputed D considering the rectangle window shown in the figure. This data set has been widely used and gave slightly different results according to the window definition in [7, page 1160] or [30, page 634]. It can be seen that the M and D functions give the same results if points are not weighted. Nevertheless, D values cannot be interpreted easily and cannot be compared across distances.

Both methods suffer here a severe lack of power due to the very little number of controls. The confidence envelopes are computed at 10% levels (from 1000 simulations). The GoF test applied to M returns $P_u = 23\%$. Diggle and Chetwynd obtained a P -value equal to 14% for D . Increasing the number of controls would not have been a real problem if the experimental design had included a distance-based point pattern analysis.

3.3. *V. americana* and *Q. rosea*. The dataset (map in Figure 8) contains 156 *V. americana*, 388 *Q. rosea*, and 10,732 other trees in a 25 ha plot where flooded zones have been excluded, leaving a 20.06 ha, polygonal shape for the study.

Aggregation of both species is detected up from 4–6 meters (Figures 9(a) and 9(d)) by the univariate M function. The species repulse each other (Figures 9(b) and 9(c)). $P_u < 0.1\%$ in all cases, according to the bivariate M function. All trees are used as the benchmark in all analyses.

These results suggest competition if our null hypothesis is correct: we suppose that both species could locate anywhere if the other did not impede it. Of course, it might be wrong so further work is necessary to test alternate hypotheses: the environment may be different and niche preferences may be the reason for segregation, or else the spatial distribution of populations may not be in equilibrium, and we may be observing the contact of two colonization fronts.

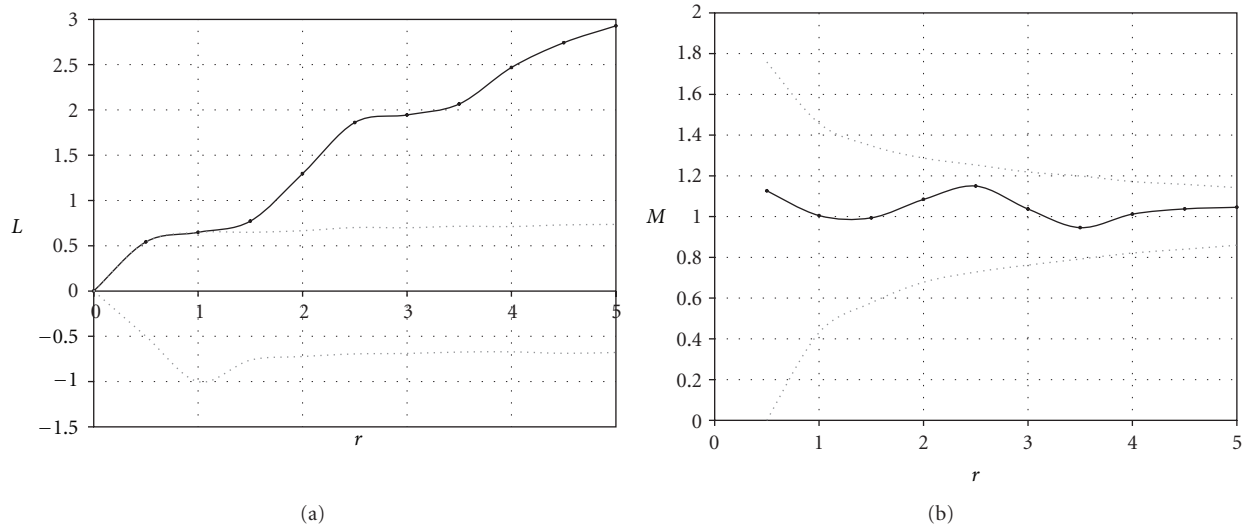


FIGURE 6: Inhomogeneous point set, univariate L and M functions for black points. L is not pertinent contrarily to M that controls for first-order heterogeneity.

3.4. *V. americana* Regeneration. The density of *V. americana* is very variable (Figure 10(a)). The univariate M function is applied to saplings (reference and neighbor points are saplings, the benchmark is all trees; weights are basal areas). Saplings are aggregated (Figure 10(b)) at all distances, $P_u < 0.1\%$. The intertype M function shows that potentially reproducing trees repulse saplings (Figure 10(c)), with significant results up to 15 meters. Actually, $P_u = 6.7\%$ (reference points are potentially reproducing trees, neighbors are saplings, and the benchmark is all trees; weights are basal areas).

Jansen et al. [31] already mentioned the absence of seedlings around adult *V. americana* at short distances (6 meters). Our results show that no sapling can be found less than 3 meters apart from adult trees, but also that the relative abundance of saplings among neighbors is low (these results are significant according to Monte-Carlo simulations), suggesting that *Vouacapoua* regeneration follows the Janzen-Connell hypothesis [32, 33].

4. Discussion

4.1. Using M . The theoretical examples illustrate that $M(r)$ is equivalent to $\hat{K}(r)/\pi r^2$ when applied to a homogenous, nonweighted point pattern. The univariate M function is not affected by virtual aggregation when the point pattern is not homogenous, using all points as its benchmark. This means that the points of interest should be a small enough fraction of the whole point set to allow considering the latter as a valid control set. The case-control approach is meaningful when the points of interest must not be included in the control set. Then, a sufficient number of controls is required, or the test against the null hypothesis of independence of points will not be powerful.

Although the M function requires several point types, it is completely different from a bivariate K or L function. This may be illustrated by the example of Section 3.1.1: the univariate M function characterizes the spatial structure of

black points, exactly like K ; grey points are added to black points to obtain a benchmark for $M(r)$ (the number of points whatever their type less than r apart from each black point), while the disk area πr^2 is the benchmark for $K(r)$. In Section 3.3, the bivariate K function could be computed instead of the bivariate M function: it would only consider the 156 *V. americana* and 388 *Q. rosea* trees (and suppose both are distributed homogeneously), while M also includes the 10,732 other trees to constitute its benchmark.

Figure 9(a) shows a good example of the behavior of the M function. Confidence envelopes are around the expected value equal to 1. At very low distances, they are not defined if no pair of points less than r apart exists, and the confidence interval is very wide then because of stochasticity, amplified by the little number of point pairs. At long distances, all values tend to 1. When point weights are not homogeneously distributed, the envelope is not around 1 (Figure 10). Heuristically, M measures the spatial structure of square centimeters of basal areas of trees: when points are redistributed independently from each other under the null hypothesis, square centimeters are still aggregated. More or less aggregation than under the null hypothesis is detected relatively to its envelope and by the GoF test. As shown by Loosmore and Ford [23], the classical, Monte-Carlo-generated confidence envelope may be too optimistic: while the M curve is clearly out of the 1% envelope, the P -value for *V. americana* regeneration (Figure 10) is only 6.7% (but no power study for the GoF has been conducted as far as we know).

4.2. Relative versus Topographic Measures. Distance-based measures of spatial concentration can be classified into two main categories [34] following Brühlhart and Traeger [11]. Topographic ones compare a number of neighbors to a measure of space (a surface area) while relative ones compare it to another number of neighbors. All indices used in ecology are topographic, except for Diggle and Chetwynd [7] D . On the opposite, economists, who are often interested

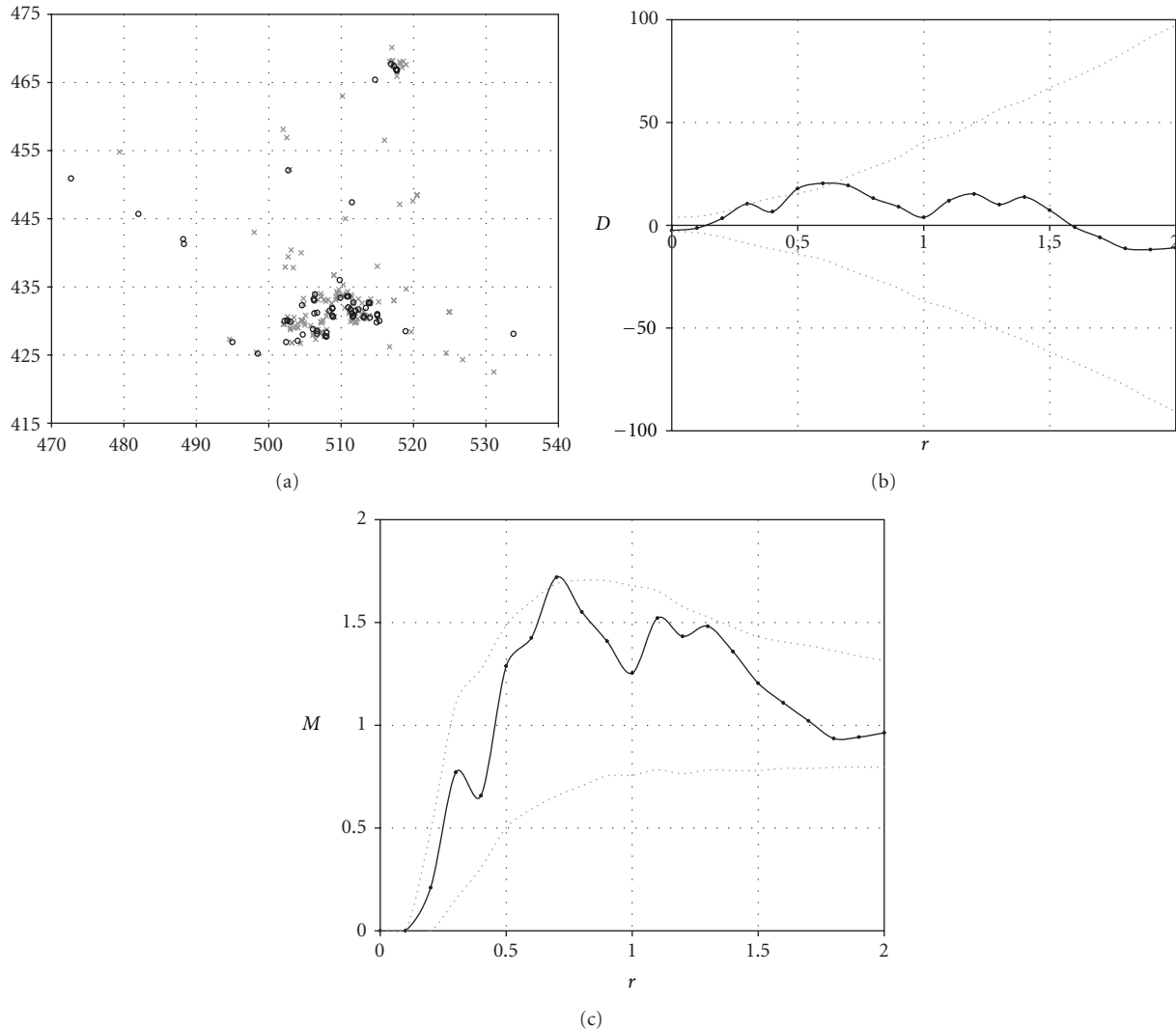


FIGURE 7: Childhood Leukemia epidemiology [6]. Map (a): cases (62 circles) are ill children locations; controls (141 crosses) are a sample of the whole population; distances are in km. Cases are significantly clumped, as shown by both functions D (b) and M (c), drawn as solid lines. M shows that in a 0.7 km radius circle around each case, the case density is 70% higher than expected without aggregation ($M = 1.7$). Confidence intervals for the null hypothesis of independence (dotted lines) are computed by Monte-Carlo simulations at the 10% risk level. The poor significance levels are due to too few controls.

in the spatial distribution of firms on a territory, mostly use relative measures: using the distribution of the whole industry as a benchmark to study the spatial structure of an economic sector is even one of the good properties a measure must respect according to Duranton and Overman [10]. We believe both frameworks can help addressing ecological questions, as they allow different null models to be tested.

The topographic toolbox is already well furnished, with Baddeley et al.'s [8] K_{inhom} and Wiegand and Moloney's [28, 35] O-ring allowing to deal with inhomogeneous point patterns. Diggle et al. [9] separated control points to evaluate intensity and cases to evaluate dependence in K_{inhom} ; that is to say they used K_{inhom} as a relative measure. The M function is designed for this purpose. It is similar to K_{inhom} with a simple box kernel [13] with bandwidth parameter r

used to estimate density around each reference point, but it also allows weighting points. The relative structure of basal areas of trees is more meaningful than that of individuals in many applications (biomass spatial structure, competition for light, etc.): roughly speaking, a big tree is often more similar to many small trees than to a single one.

5. Conclusion

The M function is defined as a generalization of Ripley's K function to allow its application to inhomogeneous point processes and to take into account point weights. From a more theoretical point of view, it is a weighted, relative measure of spatial structure. We believe that relative measures (comparing a point pattern to another) are powerful tools, even though the topographic approach is more used.

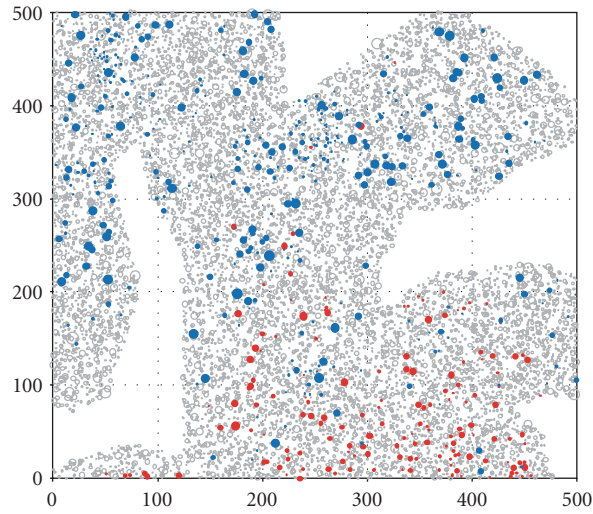


FIGURE 8: Map of trees. *Vouacapoua americana* trees are blue, *Qualea rosea* red and other trees grey. Circle sizes are proportional to those of the trees. Flooded zones are excluded. Distances are in meters.

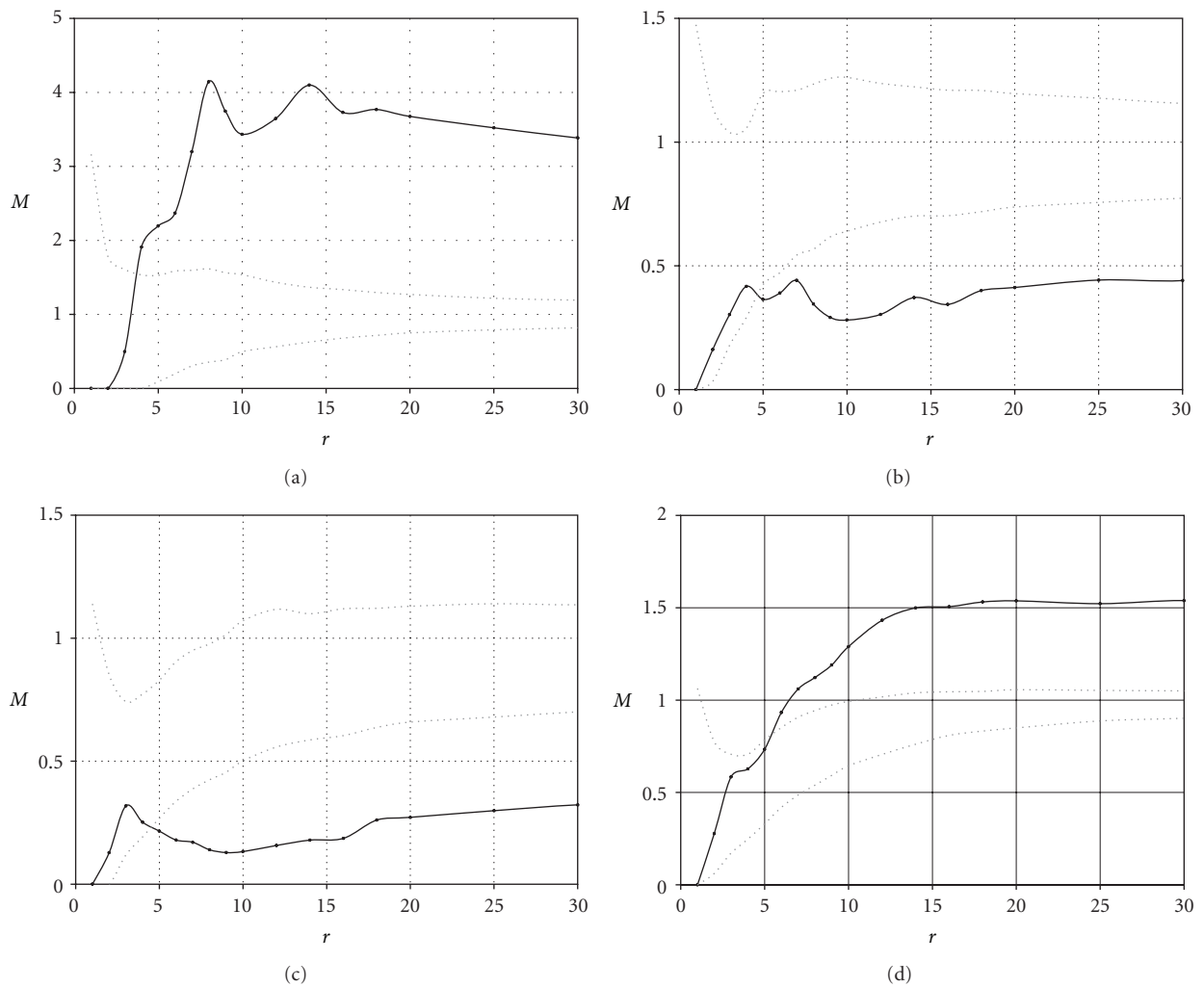


FIGURE 9: Spatial structure of *Vouacapoua americana* ((a) univariate M function) and *Qualea rosea* ((d) univariate M function), and bivariate M functions ((b) M_{V_a, Q_r} and (c) M_{Q_r, V_a}). Confidence intervals are computed at 1% risk level. Both species are aggregated over around 5 meters, and they significantly repulse each other.

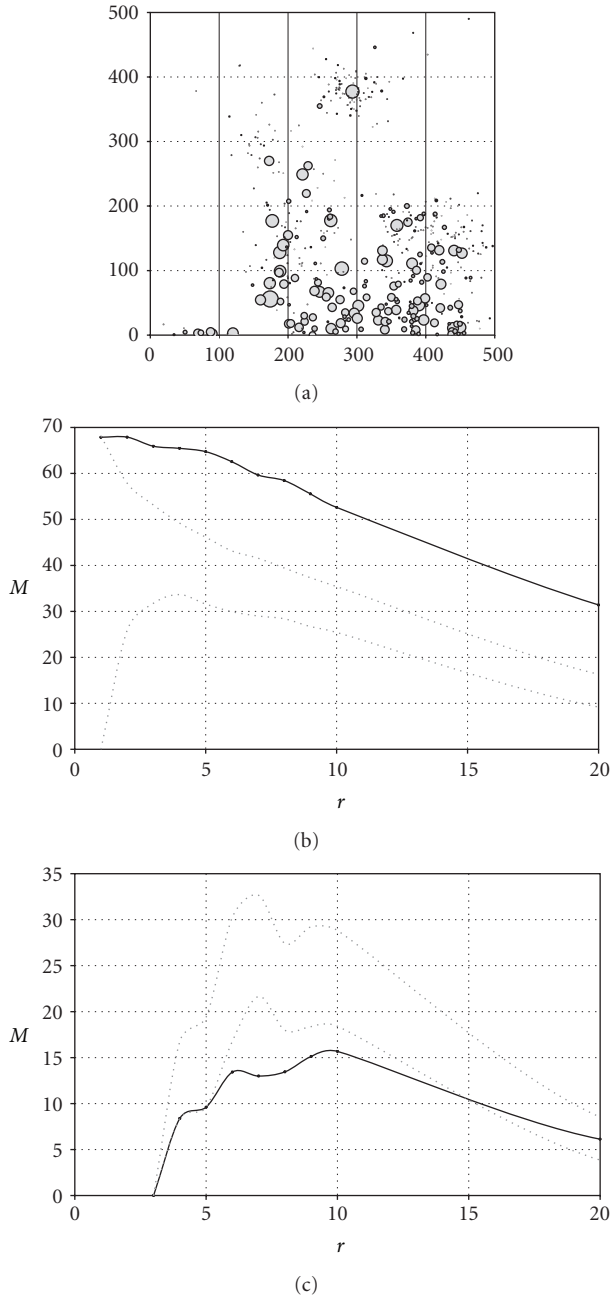


FIGURE 10: Regeneration of *Vouacapoua americana*. (a) The map shows trees with their size. Distances are in meters. (b) Univariate M function applied to saplings: saplings are aggregated at all distances. (c) Bivariate M function applied to saplings around potentially reproducing trees (over 30 cm DBH): a significant lack of saplings is detected between 6 and 9 meters. Confidence intervals are computed at 1% risk level.

To allow the effective use of the M function, we developed the necessary code for R [36], available as a supplementary material available online at doi:10.1155/2012/619281.

Acknowledgments

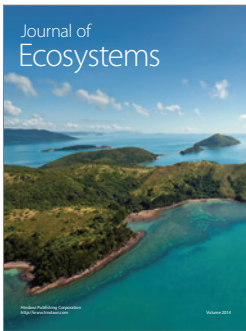
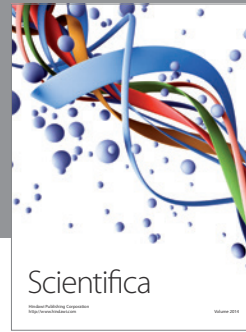
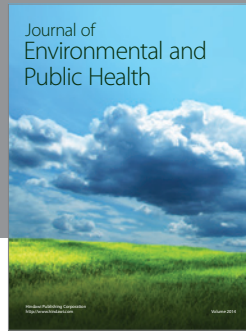
The authors thank the editor and an anonymous referee for useful suggestions. This work has benefited from an

“Investissement d’Avenir” grant managed by Agence Nationale de la Recherche (CEBA, ref. ANR-10-LABX-0025). This paper partially incorporates earlier unpublished work written by E. Marcon and F. Puech (generalizing Ripley’s K function to inhomogeneous populations, halshs-00372631, version 1).

References

- [1] E. C. Pielou, “The use of point-to-plant distances in the study of the pattern of plant populations,” *Journal of Ecology*, vol. 47, no. 3, pp. 607–613, 1959.
- [2] B. W. Silverman, *Density Estimation For Statistics and Data Analysis*, Chapman & Hall, London, UK, 1986.
- [3] B. D. Ripley, “The second-order analysis of stationary point processes,” *Journal of Applied Probability*, vol. 13, pp. 255–266, 1976.
- [4] J. E. Besag, “Comments on Ripley’s paper,” *Journal of the Royal Statistical Society*, vol. B 39, no. 2, pp. 193–195, 1977.
- [5] R. Law, J. Illian, D. F. R. P. Burslem, G. Gratzler, C. V. S. Gunatilleke, and I. A. U. N. Gunatilleke, “Ecological information from satial patterns of plants: insights from point process theory,” *Journal of Ecology*, vol. 97, no. 4, pp. 616–628, 2009.
- [6] J. Cuzick and R. Edwards, “Spatial clustering for inhomogeneous populations,” *Journal of the Royal Statistical Society*, vol. B 52, no. 1, pp. 73–104, 1990.
- [7] P. J. Diggle and A. G. Chetwynd, “Second-order analysis of spatial clustering for inhomogeneous populations,” *Biometrics*, vol. 47, no. 3, pp. 1155–1163, 1991.
- [8] A. J. Baddeley, J. Møller, and R. Waagepetersen, “Non- and semi-parametric estimation of interaction in inhomogeneous point patterns,” *Statistica Neerlandica*, vol. 54, no. 3, pp. 329–350, 2000.
- [9] P. J. Diggle, V. Gómez-Rubio, P. E. Brown, A. G. Chetwynd, and S. Gooding, “Second-order analysis of inhomogeneous spatial point processes using case-control data,” *Biometrics*, vol. 63, no. 2, pp. 550–638, 2007.
- [10] G. Duranton and H. G. Overman, “Testing for localization using micro-geographic data,” *Review of Economic Studies*, vol. 72, no. 4, pp. 1077–1106, 2005.
- [11] M. Brühlhart and R. Traeger, “An account of geographic concentration patterns in Europe,” *Regional Science and Urban Economics*, vol. 35, no. 6, pp. 597–624, 2005.
- [12] E. Marcon and F. Puech, “Measures of the geographic concentration of industries: improving distance-based methods,” *Journal of Economic Geography*, vol. 10, no. 5, pp. 745–762, 2010.
- [13] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan, *Statistical analysis and modelling of spatial point patterns*, Wiley-Interscience, Chichester, UK, 2008.
- [14] G. Lang and E. Marcon, “Testing randomness of spatial point patterns with the Ripley statistic,” *ESAIM: Probability and Statistics*.
- [15] P. Haase, “Spatial pattern analysis in ecology based on Ripley’s K function: Introduction and methods of edge correction,” *Journal of Vegetation Science*, vol. 6, no. 4, pp. 575–582, 1995.
- [16] B. D. Ripley, *Statistical Inference For Spatial Processes*, Cambridge University Press, 1988.
- [17] E. Marcon and F. Puech, “Evaluating the geographic concentration of industries using distance-based methods,” *Journal of Economic Geography*, vol. 3, no. 4, pp. 409–428, 2003.
- [18] S. P. Kingham, A. C. Gatrell, and B. Rowlingson, “Testing for clustering of health events within a geographical information

- system framework,” *Environment & Planning A*, vol. 27, no. 5, pp. 809–821, 1995.
- [19] A. C. Gatrell and T. C. Bailey, “Interactive spatial data analysis in medical geography,” *Social Science and Medicine*, vol. 42, no. 6, pp. 843–855, 1996.
- [20] A. C. Gatrell, T. C. Bailey, P. J. Diggle, and B. S. Rowlingson, “Spatial point pattern analysis and its application in geographical epidemiology,” *Transactions of the Institute of British Geographers*, vol. 21, no. 1, pp. 256–274, 1996.
- [21] N. C. Kenkel, “Pattern of self-thinning in jack pine: testing the random mortality hypothesis,” *Ecology*, vol. 69, no. 4, pp. 1017–1024, 1988.
- [22] F. Goreaud and R. Pélissier, “Avoiding misinterpretation of biotic interactions with the intertype K 12-function: population independence versus random labelling hypotheses,” *Journal of Vegetation Science*, vol. 14, no. 5, pp. 681–692, 2003.
- [23] N. B. Loosmore and E. D. Ford, “Statistical inference using the G or K point pattern spatial statistics,” *Ecology*, vol. 87, no. 8, pp. 1925–1931, 2006.
- [24] P. J. Diggle, *Statistical Analysis of Spatial Point Patterns*, Edward Arnold, London, UK, 2003.
- [25] S. Gourlet-Fleury, J. M. Guehl, and O. Laroussinie, *Ecology & Management of a neotropical rainforest*, Lessons drawn from Paracou, a long-term experimental research site in French Guiana, Elsevier, Paris, France, 2004.
- [26] B. Matérn, “Spatial variation,” *Meddelanden Från Statens Skogsforskningsinstitut*, vol. 49, no. 5, pp. 1–144, 1960.
- [27] F. Goreaud, *Apports de l’analyse de la structure spatiale en forêt tempérée à l’étude de la modélisation des peuplements complexes [Ph.D. thesis]*, ENGREF, Nancy, France, 2000.
- [28] T. Wiegand and K. A. Moloney, “Rings, circles, and null-models for point pattern analysis in ecology,” *Oikos*, vol. 104, no. 2, pp. 209–229, 2004.
- [29] K. Schiffers, F. M. Schurr, K. Tielbörger, C. Urbach, K. Moloney, and F. Jeltsch, “Dealing with virtual aggregation—a new index for analysing heterogeneous point patterns,” *Ecography*, vol. 31, no. 5, pp. 545–555, 2008.
- [30] B. S. Rowlingson and P. J. Diggle, “Splancs: spatial point pattern analysis code in S-plus,” *Computers and Geosciences*, vol. 19, no. 5, pp. 627–655, 1993.
- [31] P. A. Jansen, F. Bongers, and P. J. Van Der Meer, “Is farther seed dispersal better? Spatial patterns of offspring mortality in three rainforest tree species with different dispersal abilities,” *Ecography*, vol. 31, no. 1, pp. 43–52, 2008.
- [32] J. H. Connell, “On the role of natural enemies in preventing competitive exclusion in some marine animals and in forest trees,” in *Dynamics of Populations*, P. J. Den Boer and G. Gradwell, Eds., pp. 298–312, 1971.
- [33] D. H. Janzen, “Herbivores and the number of species in tropical forests,” *The American Naturalist*, vol. 104, no. 940, pp. 501–528, 1970.
- [34] E. Marcon and F. Puech, “A typology of distance-based measures of spatial concentration,” HAL, Working Paper no. halshs-00679993, 2012.
- [35] T. Wiegand, K. A. Moloney, J. Naves, and F. Knauer, “Finding the missing link between landscape structure and population dynamics: a spatially explicit perspective,” *American Naturalist*, vol. 154, no. 6, pp. 605–627, 1999.
- [36] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2012.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

