On explicit formulas of edge effect correction for Ripley's K-function

Goreaud, François1,2 & Pélissier, Raphael3,4*

1ENGREF, Laboratoire de Recherche en Sciences Forestières, 14 rue Girardet, F-54042 Nancy Cedex, France; 2Present address: CEMAGREF, Laboratoire d'Ingénierie pour les Systèmes Complexes, 24 av. des Landais, BP 50085, F-63172 Aubière Cedex 1, France; E-mail: francois.goreaud@clermont.cemagref.fr; 3CNRS-Université Claude Bernard Lyon 1, UMR 5558, Laboratoire de Biométrie et Biologie Évolutive, 43 Bd du 11 novembre 1918, F-69622 Villeurbanne Cedex, France; 4Present address: IRD, Laboratoire ERMES, 5 rue du Carbone, F-45072 Orléans Cedex 2, France; E-mail: raphael.pelissier@orleans.ird.fr

*Corresponding author: Fax +33238499534; E-mail: raphael.pelissier@orleans.ird.fr

Abstract. The analysis of spatial pattern in plant ecology usually implies the solution of some edge effect problems. We present in this paper some explicit formulas of edge effect correction that should enable plant ecologists to analyze a wider range of real field data.

We consider the local correcting factor of edge effect for Ripley's K-function, that can also be used for other statistics of spatial analysis based on the counting of neighbours within a given distance. For both circular and rectangular study areas, we provide a review of explicit formulas and an extension of these formulas for long and narrow plots. In the case of irregular-shaped study plots, we propose a generalization of the method that computes edge effect correction by excluding triangular surfaces from a simple (rectangular or circular) initial shape.

An example in forest ecology, where the soil characteristics determine a study plot of complex shape, illustrates how this edge effect correction can be effective in avoiding misinterpretations.

Keywords: Circular plot; French Guiana; Irregular-shaped plot; Local correcting factor; Qualea rosea; Rectangular plot; Second-order neighbourhood analysis; Spatial point pattern; Tropical rain forest.

Introduction

Ripley's K-function (Ripley 1976, 1977, 1981) and the related functions of second-order neighbourhood analysis of spatial point patterns (e.g. function L in Besag 1977; K in Lotwick & Silverman 1982; L* in Getis & Franklin 1987; K* and g in Stoyan et al. 1987) have been used recently in many plant ecological studies (e.g. Szwarczyk 1990; Duncan 1991; Penttinen et al. 1992; Moer 1993; Gouard et al. 1995; Haase et al. 1996, 1997; Ward et al. 1996; Couteron & Kokou 1997; Martens et al. 1997; Goreaud et al. 1998; Pélissier 1998). Considering plant locations as points in an (x,y) coordinate system, the point pattern can be summarised, under the assumptions of stationarity and isotropy (i.e. invariance by translation and rotation), by its first- and second-order characteristics: its intensity λ corresponding to the expected number of points per unit area; and Ripley's K-function, defined so that λK(t) is the expected number of neighbours in a circle of radius t centred on an arbitrary point of the pattern (Ripley 1977).

Classical estimators of these characteristics are:

\[ \hat{\lambda} = \frac{N}{S} \]  

where N is the number of points in area S;

\[ \hat{K}(t) = \frac{1}{\hat{\lambda} N} \sum_{j=1}^{N} \sum_{i=1}^{N} k_{ij} \]  

where \( k_{ij} = 1 \) if the distance between points \( i \) and \( j \leq t \); and \( k_{ij} = 0 \) if the distance between points \( i \) and \( j > t \).

However, for points located near the boundary of the study area, the real number of neighbours within distance \( t \) can be underestimated because some of them can be located outside of the study area (Fig. 1).

Gignoux et al. (1999) showed that it was not always necessary to take this edge effect into account to test precise point process hypothesis. However, the edge effect correction is highly recommended to compare point patterns or to use \( K(t) \) values for ecological interpretations (neighbourhood, competition, etc.), Ripley (1982) and more recently Haase (1995) made reviews of various methods to correct this edge effect, such as the use of a buffer zone; the toroidal duplication of the study area; a global correcting factor proposed by Osher & Stoyan (1981); and the local correcting factor proposed in Ripley (1977). All these methods can be used with every function based on the counting of neighbours within a given distance. However, Ripley's local correcting factor presents several advantages: (1) it allows unbiased and robust (Kieu & Mora 1996) estimation of \( K(t) \) without drastically reducing the data set under analysis; and (2) it can also be used with individual-point statistics such as \( L*(r) \) (Getis & Franklin 1987). According to this method, for a point \( i \) located closer to the boundary of the study area than to its neighbouring point \( j, k_{ij} \) in Eq. (2) – is...
computed as the inverse of the proportion of the perimeter of the circle \(C_y\) (centred on \(i\) and passing through \(j\)) which is inside the study area (Fig. 1).

Precise formula of this corrected \(k_j\) depends on the shape of the study area and on the location of point \(i\) in relation to the boundaries. Diggle (1983) gave explicit formulas for circular and rectangular study areas. Errors and incomplete transcriptions appeared later (e.g. Getis & Franklin 1987; Haase 1995). Various programs to compute \(K(t)\) can also be found, for instance on Internet, even for study areas of complex shape. But edge effect correction procedures are rarely detailed. We thus propose below: (1) a review of edge effect correcting formulas for both circular and rectangular study areas, with an extension for long and narrow plots; (2) a general method to deal with study areas of complex shape, by excluding triangular surfaces from a simple (rectangular or circular) initial shape. The efficiency of the method is then illustrated with a short example of spatial pattern analysis of trees in an experimental forest plot in French Guiana.

Explicit formulas

Let us call \(t\) the distance between the points \(i\) and \(j\); \(C_y\) the circle centred on \(i\) and passing through \(j\); \(C_{in}\) the part of \(C_y\) which is inside the study area; \(C_{out}\) the part of \(C_y\) which is outside the study area and \(\alpha_{out}\) the corresponding angle (Fig. 1).

The correcting factor \(k_{ij}\) can then be calculated as:

\[
k_{ij} = \frac{2\pi t}{C_{in}} = \frac{2\pi t}{2\pi - C_{out}} = \frac{2\pi}{2\pi - \alpha_{out}}
\]

(3)

The simplest method to compute \(k_{ij}\) is then to calculate first \(\alpha_{out}\) by using classical geometrical properties. Various cases of the relative position of \(C_y\) and the study area boundaries have to be considered.

Circular study area

As far as a circular study area (of radius \(R\) and centred on \(O\)) is concerned, we only have to consider two cases: (1) no intersection between \(C_y\) and the boundary of the circular study area; (2) two intersection points (\(A\) and \(A'\)) between \(C_y\) and the boundary (Fig. 1a). For this case, the simplest way to calculate \(\alpha_{out}\) is to develop:

\[
\vec{OA}^2 = (\vec{Oi} + \vec{IA})^2 = \vec{Oi}^2 + \vec{IA}^2 + 2\vec{Oi} \cdot \vec{IA},
\]

where \(\cdot\) is the scalar product of vectors. As

\[
\vec{OA}^2 = R^2, \quad \vec{Oi}^2 = d_i^2, \quad \vec{IA}^2 = t^2,
\]

and \(\vec{Oi} \cdot \vec{IA} = dt \cos(\alpha_{out} / 2),\)

we can deduce that: \(\cos(\alpha_{out} / 2) = \frac{R^2 - d_i^2 - t^2}{2dt},\)

and thus obtain the explicit formula given in Table 1, which is the same as the one proposed by Diggle (1983).

Rectangular study area

The rectangular shape is a little more complicated, because there are many intersection cases between \(C_y\) and the rectangle boundaries. They can be distinguished by comparing \(t\), the distance between the points \(i\) and \(j\), to: (1) \(d_1, d_2, d_3\) and \(d_4\), the distances between \(i\) and each of the four sides; and (2) the distance from \(i\) to the four corners of the rectangle. For each point \(i, \alpha_{out}\) is estimated by taking into account the global contribution of the four sides and corners. If the circle intersects a side twice its contribution to \(\alpha_{out}\) will be \(2\) Arcosos \((d) / t\), where \(d\) corresponds either to \(d_1, d_2, d_3\) or \(d_4\). Thus, in Fig. 1b, \(\alpha_{out}\) will contribute for \(2\) Arcosos \((d_1 / t) + \alpha_{out}\) for \(2\) Arcosos \((d_2 / t)\), the total \(\alpha_{out}\) value being in that case \(\alpha_{out} + \alpha_{out}\). If the circle intersects a side only once because the corner is inside \(C_y\), the contribution of this side to \(\alpha_{out}\) will be only Arcosos \((d / t)\), the corner itself will contribute for \(\pi / 2\) and the perpendicular side may also contribute. Thus, in Fig. 1c, the total \(\alpha_{out}\) value will be Arcosos \((d_1 / t) + \pi / 2 + \text{Arcosos } (d_2 / t)\).

There are then 27 different configurations that correspond to the eight elementary cases described in the second part of Table 1: no intersection; intersection with one, two or three sides, with corner(s) inside or outside \(C_y\). Cases of intersection with four sides are not considered because they correspond to excessively high values of \(t\).

Diggle (1983) gave equivalent formulas for the first four cases of the rectangular study area in Table 1, which allow computation of \(K(t)\) for \(t\) up to half of the shorter side of the rectangle. The latter four cases allow computation of \(K(t)\) for \(t\) up to half of the longer side of
the rectangle, which is useful to deal with long narrow plots. Other equivalent formulas can of course be obtained, for instance with the Arctan function.

**Generalization to study areas of complex shape**

Simple circular and rectangular study areas are commonly used for experimental plots. But more complex shapes are sometimes designed by natural or artificial obstacles. Moreover, as the K-function is only defined for homogeneous processes (Ripley 1977), it can be necessary to omit some parts of a heterogeneous study area in the computation. The shape of the final study plot can thus be very complex. Therefore we propose here a method to compute \( K(t) \) with edge effect correction in the case of a study area of complex shape.

First, the real shape of the study area must be approximated by removing some polygonal surfaces from a simple (rectangular or circular) initial zone (Fig. 2). These surfaces may be omitted either near the boundary of the initial zone in order to design a polygonal study area, or within the zone itself to remove potential heterogeneous areas. In order to simplify the computation, the omitted surfaces must then be decomposed in triangles, that (1) do not overlap each other; and (2) do not cross the boundary of the initial shape (Fig. 2b).

After deletion of the points initially located within the triangles, the correcting factor \( k_i \) can still be computed through Eq. (3) by considering that \( C_{out} \) the part of the perimeter of \( C_i \) lying outside the study area, is composed of parts lying outside the initial rectangular (or circular) study area \( (C_{init}) \) and of parts inside the removed triangles \( (C_{tri}) \). The global \( \alpha_{out} \) angle is thus computed as the sum of the contributions of the initial shape (according to formulas given above), and of each removed triangle: \( \alpha_{out} = \alpha_{init} + \sum \alpha_{tri} \) (Fig. 2b). The contribution of \( \alpha_{tri} \) depends once more on the relative position of circle \( C_i \) and the triangle \( ABC \). Various intersection cases have thus to be considered (Fig. 3).

They can be analytically distinguished by considering: (1) the distances from the circle \( C_i \) to each vertex of the triangle (this will give the number of vertices inside the circle); and (2) the intersection points between \( C_i \) and the triangle sides. The calculations are made through usual geometrical considerations: coordinates of the intersection points are calculated with line and circle equations, and the excluded angles are calculated with the Arccos function applied to scalar products. The explicit formulas of \( \alpha_{tri} \) corresponding to the cases described in Fig. 3 are given in Table 2 (we do not detail here the fastidious calculation of the intersection points' coordinates and scalar products).

### Table 1. Explicit formulas of \( \alpha_{out} \) for circular and rectangular study areas. Cases of intersection with four sides of a rectangular study area are not considered.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Explicit formula of ( \alpha_{out} ) in Eq. (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circular study area (Fig. 1a) ( t \leq R - d )</td>
<td>( \alpha_{out} = 0 )</td>
</tr>
<tr>
<td>Rectangular study area (Fig. 1b-c) ( r \leq d_1, d_2, d_3, d_4 )</td>
<td>( \alpha_{out} = 2 \arccos \left( \frac{R^2 - d^2 - r^2}{2rd} \right) )</td>
</tr>
<tr>
<td>( r \leq d_1 )</td>
<td>( \alpha_{out} = 0 )</td>
</tr>
<tr>
<td>( r \leq d_2, d_3, d_4 )</td>
<td>( \alpha_{out} = 2 \arccos (d_i/r) )</td>
</tr>
<tr>
<td>( r &gt; d_1, d_2 ) ( r \leq d_3, d_4 )</td>
<td>( \alpha_{out} = 2 \arccos (d_i/r) + 2 \arccos (d_4/r) )</td>
</tr>
<tr>
<td>( r \leq d_3, d_4 )</td>
<td>( \alpha_{out} = \pi/2 + \arccos (d_3/r) + \arccos (d_4/r) )</td>
</tr>
<tr>
<td>( r \geq d_1, d_2 ) ( r \leq d_3, d_4 )</td>
<td>( \alpha_{out} = 2 \arccos (d_3/r) + 2 \arccos (d_4/r) )</td>
</tr>
<tr>
<td>( r \leq d_1, d_3, d_4 ) ( r \geq d_2, d_3 )</td>
<td>( \alpha_{out} = \pi/2 + 2 \arccos (d_1/r) + \arccos (d_2/r) + \arccos (d_3/r) )</td>
</tr>
<tr>
<td>( r \leq d_2, d_3, d_4 )</td>
<td>( \alpha_{out} = \pi/2 + 2 \arccos (d_1/r) + \arccos (d_2/r) + \arccos (d_3/r) )</td>
</tr>
<tr>
<td>( r \leq d_1, d_2, d_3 ) ( r \geq d_4 )</td>
<td>( \alpha_{out} = \pi + \arccos (d_1/r) + \arccos (d_2/r) )</td>
</tr>
</tbody>
</table>
Fig. 2. Spatial pattern of the tree species *Qualea rosea* Aublet (Vochysiaceae) in a 150 m × 250 m experimental plot in Paracou, French Guiana, with three soil categories (from Collinet 1997). a. Real, complex study area; b. Its approximation with a geometrical shape obtained by removing triangular surfaces from an initial rectangular shape.

**Example**

We now illustrate the efficiency of the method of edge effect correction with an example of analysis of spatial pattern of trees in an experimental forest plot in Paracou, French Guiana. The data set was provided by Collinet (1997), who showed that *Qualea rosea* Aublet (Vochysiaceae) avoids hydromorphic soils (i.e. swamps and soils temporary blocked during the rainy season), so that the species distribution is heterogeneous and almost strictly limited to non-hydromorphic areas (Fig. 2a). In order to take into account this heterogeneity, we digitised the soil map and approximated the hydromorphic part of the plot with a polygon secondarily divided into seven contiguous triangles (Fig. 2b). Ripley's function was then computed: (1) on the entire rectangular study area without any edge effect correction; (2) on the entire study area with the edge effect correction for rectangular plots; (3) on the polygonal non-hydromorphic area with the corresponding edge effect correction for complex shapes. The results are presented in Fig. 4, using the linearized

\[ L(t) = \frac{\hat{K}(t)}{\pi t} - t, \]

which is easier to interpret than \( K(t) \) (the expectation of \( L(t) \) under complete spatial random pattern is 0 for all \( t \); it becomes greater than 0 when the pattern is clustered and lower than 0 when it is regular). A 90% confidence interval for the complete spatial randomness hypothesis was obtained by the Monte Carlo method (Besag & Diggle 1977) using 1000 simulated Poisson patterns of same density than the one observed. More details on these methods can be found, for instance, in Goreaud et al. (1998).

Finally, the curves of \( L(t) \) corresponding to cases (1), (2) and (3) are quite different, and only case (3) can be correctly interpreted in terms of spatial structure. When computed without any edge effect correction \( L(t) \) values are highly underestimated (Fig. 4a), and the bias increases with the distance \( t \). In that case the 90% confidence interval for the complete spatial randomness hypothesis was obtained by the Monte Carlo method (Besag & Diggle 1977) using 1000 simulated Poisson patterns of same density than the one observed. More details on these methods can be found, for instance, in Goreaud et al. (1998).

Fig. 3. Intersection cases between \( C_{ij} \) (the circle centred on \( i \) and passing through \( j \)) and one elementary triangle.

![Fig. 4](image_url) Analysis of the spatial pattern of *Qualea rosea* in a 150 m × 250 m experimental plot in Paracou, French Guiana (cf. Fig. 2). The \( L(t) \) function was computed: (a) on the entire rectangular study area without any edge effect correction; (b) on the entire study area with edge effect correction for rectangular plots only; (c) on the non-hydromorphic polygonal area with edge effect correction for complex shapes. The shaded envelopes correspond to the 90% confidence intervals for the complete spatial randomness hypothesis.
Table 2. Explicit formulas of $\alpha_m$ for one elementary triangle. The symbol: $\cap$ means ‘intersection’ of the specified line and circle; $\emptyset$ (empty set) means that there are no intersection points; and $\cdot$ means ‘scalar product’ of the two specified vectors.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Intersection points</th>
<th>Explicit formula of $\alpha_m$ for one triangle (ABC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A, B, C$ inside $C_y$</td>
<td>—</td>
<td>$\alpha_m = 0$ (Fig. 3a)</td>
</tr>
<tr>
<td>$(A, B)$ inside $C_y$</td>
<td>$(AB) \cap C_y = E$</td>
<td>$\alpha_m = \text{Arccos} \left( \frac{iE \cdot IF}{r^2} \right)$ (Fig. 3b)</td>
</tr>
<tr>
<td>$(A, C)$ outside $C_y$</td>
<td>$(AC) \cap C_y = F$</td>
<td>$\alpha_m = \text{Arccos} \left( \frac{iF \cdot iG}{r^2} \right)$ (Fig. 3c)</td>
</tr>
<tr>
<td>$(B, C)$ outside $C_y$</td>
<td>$(BC) \cap C_y = \emptyset$</td>
<td>$\alpha_m = \text{Arccos} \left( \frac{iG \cdot iF}{r^2} \right)$ (Fig. 3d)</td>
</tr>
<tr>
<td>$A, B, C$ outside $C_y$</td>
<td>$(AB) \cap C_y = \emptyset$</td>
<td>$\alpha_m = 0$ (Fig. 3e)</td>
</tr>
<tr>
<td>$(A, C)$ outside $C_y$</td>
<td>$(AC) \cap C_y = \emptyset$</td>
<td>$\alpha_m = \text{Arccos} \left( \frac{iG \cdot iG'}{r^2} \right)$ (Fig. 3f)</td>
</tr>
<tr>
<td>$(B, C)$ outside $C_y$</td>
<td>$(BC) \cap C_y = (G, G')$</td>
<td>$\alpha_m = \text{Arccos} \left( \frac{iE \cdot iF}{r^2} \right)$ (Fig. 3g)</td>
</tr>
<tr>
<td>$A, B, C$ outside $C_y$</td>
<td>$(AB) \cap C_y = (E, E')$</td>
<td>$\alpha_m = \text{Arccos} \left( \frac{iE \cdot iF}{r^2} \right)$ (Fig. 3h)</td>
</tr>
</tbody>
</table>

Confidence interval is also biased, showing a large departure from the theoretical value $L(t) = 0$ under complete spatial randomness. In the second case with the edge effect correction for rectangular study areas (Fig. 4b), the confidence interval is correct, but the $L$-function diverges towards clustering at large distances, indicating clearly that the overall pattern is heterogeneous at that scale, and thus that the spatial structure of $Q. rosea$ at short distances cannot be correctly interpreted from this figure. However, when computed only on the non-hydromorphic polygonal area with edge effect correction for complex shapes (Fig. 4c), the $L$-function remains within the confidence interval at large distances, which means that the pattern can be considered as homogeneous at this scale. The curve can then be interpreted in terms of spatial structure, which shows the existence of significant clusters of various size in the range 10 - 60 m for $Q. rosea$ in non-hydromorphic soil conditions.

**Conclusion**

This paper aims at summarizing explicit formulas to take into account the edge effect in the computation of Ripley’s $K$-function (or of similar functions) for study areas of various shapes. Circular and rectangular shapes are well known (Diggle 1983) and broadly used. More complex study areas can also be considered by approximating the complex shape and removing triangles from a simple (rectangular or circular) initial study area. This general method allows to deal with nearly every polygonal shape by removing triangles near the boundary of a rectangular initial shape. But we also want to point out the suitability of the method to study heterogeneous areas. The use of triangles to remove particular zones within a study area of any shape, allows to define homogeneous point patterns without drastically lowering the number of observations, which would be the case if the initial plot had to be cut into smaller homogeneous subplots. This method allowed Collinet (1997) to analyse the spatial structure of 36 species in Paracou (French Guiana), by removing heterogeneous soil zones from an initial square study area. Our illustration shows that for such complex shape areas, edge effect correction is necessary to avoid misinterpretations.

However, we must be aware of the limits of the method. First, it can be very time-consuming, especially when the studied pattern consists of a large number of points, and also when an excessive number of triangles is defined. As the precision of the shape is related to the number of triangles, it must be balanced with the computation time and the power of the computer. Secondly, the
precision of the results will also depend on the proper definition of the homogeneous zone. If the heterogeneity is the result of a gradient, or if the homogeneous zone is too small, it remains inadvisable to use the K-function.

All these formulas and the computation of $K(t)$ and other classical derived functions have been implemented in C anisi by the authors and can be obtained on request. Programs for Apple Macintosh and PC with Windows are also available with documentation on Internet as ADS in ADE-4 program library (Thioulouse et al. 1997). Computational and graphical display modules can be downloaded from the following Web homepage: http://pbil.univ-lyon1.fr/ADE-4/ADE-4.html

Acknowledgements. The methods presented in this paper are the outcome of joint ideas of researchers and students on the analysis of tree spatial patterns. We thank all members of the group: L. Blanc, C. Gimaret-Carpentier, F. Houllier, F. Mercier, M.-A. Moravie and J.-P. Pascal, and also F. Collinet who provided the data set used for the illustration of this paper. The authors are grateful to D. Chessel and J. Thioulouse for their collaboration in integrating the ADS modules in ADE-4 software, and to two anonymous reviewers, E. van der Maarel and some other colleagues for constructive criticism on the manuscript.

References


Received 17 July 1998;
Revision received 8 November 1998;
Accepted 25 November 1998;
Final version received 28 April 1999.