Handbook of Spatial Point-Pattern Analysis in Ecology

Thorsten Wiegand, Kirk A. Moloney

Estimators and Toolbox

Publication details
Thorsten Wiegand, Kirk A. Moloney
Published online on: 20 Dec 2013

How to cite: Thorsten Wiegand, Kirk A. Moloney. 20 Dec 2013, Estimators and Toolbox from: Handbook of Spatial Point-Pattern Analysis in Ecology CRC Press
Accessed on: 17 Dec 2018

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: https://www.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
This chapter broadens the discussion on point-pattern analysis by introducing various technical aspects concerning the estimation of summary statistics. In Section 3.1, we present methods for estimating the various summary statistics introduced earlier in the book. Estimation, in particular, must explicitly account for the problem of edge correction, a topic briefly touched upon in Section 2.3.2.5. Here, we provide the theoretical foundation of the currently available methods and compile the different approaches into a common framework. In Section 3.2, we present methods for analyzing replicate patterns by combining results into one “master” result. This task is closely associated with the initial development of estimators, since combining the results of replicate patterns comes down to defining rules of how to estimate master summary statistics for replicate plots. Section 3.3 deals with the independent superposition of point processes. If the summary statistics of two-point processes are known, we can calculate, under certain conditions, the summary statistics that result from the independent superposition of the two-point processes. This operation considerably expands the range of point-process models we can use and allows us to describe observed patterns in a more realistic way.

Section 3.4 consists of a toolbox presenting different techniques and methods that are especially useful in a variety of applications, such as an algorithm to detect clusters (Section 3.4.1), methods to conduct point-pattern analysis in observation windows of irregular shape (Section 3.4.2), and methods of pattern reconstruction (Section 3.4.3), a nonparametric approach to generating patterns with predefined statistical properties. In most applications, the predefined statistical properties are the summary statistics of an observed pattern. The underlying simulated annealing algorithm generates patterns that match the (predefined) summary statistics of the observed pattern very closely. Or, in other words, it allows us to generate stochastic, replicate patterns consisting of the same, underlying statistical properties. This algorithm has two important applications: the reconstructed patterns can serve as realizations of the fundamental division for bivariate patterns (i.e., the independence null model we briefly touched upon in Section 2.2.1.2), and it allows us to determine how many (and which) different summary statistics are needed to accurately describe the statistical properties of point patterns.
3.1 Estimators of Summary Statistics

In this chapter, we present the more technical aspects of calculating summary statistics for the different data types by presenting their estimators, which generally contain some form of edge correction. In general, edge correction is used to remove a bias in the estimation of a summary statistic that is caused by the data being confined to a finite observation window \( W \). Corrections are required because points near the border of \( W \) will have fewer neighbors in some directions than others. This is due to the fact that a circle or ring with radius \( r \) centered on a point near the edge is often not completely located within the observation window \( W \). For example, when estimating the \( K \)-function of a univariate pattern, we visit each point \( i \) of the pattern and count the number of additional points \( n_i^c(r) \) lying within distance \( r \) of that point (the superscript \( c \) refers to “circle”) and take the average over all points. This yields an estimate of \( \lambda K(r) \), the number of additional points lying within distance \( r \) of the typical point (see Section 2.3.2.5). A naive estimator of \( \lambda K(r) \) would, therefore, calculate the average of \( n_i^c(r) \) over all points \( i \), that is,

\[
\tilde{\pi}^c(r) = \frac{1}{n} \sum_{i=1}^{n} n_i^c(r)
\]

(3.1)

Figure 3.1

Need for edge correction in estimation of second-order and nearest neighbor summary statistics. As illustrated by circle C3, the \( O \)-ring statistic is determined by calculating the density of points lying within the gray ring of radius \( r \) and width \( dr \) centered on a typical focal point (open circle) of the process. No edge correction is required when the ring lies fully within the observation window, as shown by circle C1. In this example, the distance \( d \) to the nearest neighbor can also be correctly estimated. However, for circle C2 many points within distance \( r \) of the focal point are potentially located outside the observation window \( W \), as is the nearest neighbor (p2).
This approach is fine if the focal point \( i \) is located further than distance \( r \) from the nearest border (e.g., the gray focal point in circle C1 in Figure 3.1). The corresponding estimate of \( \hat{n}_i(r) \) will be correct. However, a problem arises if the circle of radius \( r \) does not completely overlap \( W \) (e.g., circle C2 in Figure 3.1). Under these circumstances some points within distance \( r \) of the focal point \( i \) may be located outside \( W \) and will not be counted, resulting in an estimate of \( \hat{n}_i(r) \) that will be too low. If this effect is not accounted for correctly, the estimator \( \hat{n}(r) \) given in Equation 3.1 will be biased. Clearly, this bias will be aggravated for larger radii \( r \) because the “mean overlap” of circles around the points of the pattern with \( W \) will be correspondingly low.

A similar problem arises for the distribution function \( D(r) \) of the distances \( r \) to the nearest neighbor. Again, we visit each point \( i \) of the pattern in turn and estimate the distance \( d_i \) to the nearest neighbor. The naive estimator of \( D(r) \) enumerates, for each distance \( r \), the number of cases where \( d_i < r \), that is,

\[
\hat{D}(r) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(0 < d_i \leq r)
\]  

The indicator function \( \mathbf{1}(0 < d_i \leq r) \) yields a value of 1 if \( d_i < r \) and 0 otherwise. If the nearest-neighbor distance \( d_i \) of point \( i \) within \( W \) is less than the shortest distance \( d_b \) to the border (e.g., the gray focal point in circle C1 in Figure 3.1), the corresponding estimate of \( d_i \) will be correct. However, if the distance \( d_i \) to the nearest neighbor within \( W \) is larger than the shortest distance \( d_b \) to the border (as with circle C2 in Figure 3.1), a problem arises in that a point outside \( W \) (point p2 in Figure 3.1) may, in fact, be the nearest neighbor. As a consequence, the estimated value of \( d_i \) may be too large, and if this effect is not corrected, the estimate \( \hat{D}(r) \) of \( D(r) \) given in Equation 3.2 may contain a bias.

Note that the effect of not using edge correction in calculating \( D(r) \) is less severe than it is for second-order summary statistics. This can be easily understood. If the circle around a point is only partly located within \( W \), the count \( n_i(r) \) in Equation 3.1 will be, in most cases, too low. In the worst case, where a point is located at the corner of \( W \), the count of \( n_i(r) \) will yield only 1/4 of its expectation (because only 1/4 of the area of the circle overlaps \( W \)). However, for nearest-neighbor statistics there is, even in this worst-case scenario, still a 1/4 probability that the nearest neighbor is indeed located within \( W \) and can be correctly determined. Even if this is not the case, there is a high probability that the second, third, or \( k \)th nearest neighbor is located within \( W \), and will provide a reasonable approximation of the distance to the nearest neighbor. However, this is not guaranteed if the number of points of the pattern is low. We can, therefore, expect larger differences among estimators if the number of points of the pattern is low.

While estimators of summary statistics for unmarked patterns usually require edge correction, this is not necessary for certain summary statistics.
of marked point patterns, such as mark-connection and mark-correlation functions. This is due to the conditional nature of these summary statistics, which are estimated as the quotient of two functions that are subject to the same bias. The biases simply cancel out in the quotient (see Sections 3.1.6 and 3.1.7).

### 3.1.1 General Edge Correction Strategies

Various solutions have been proposed to deal with the problem of edge correction for univariate patterns. In the simplest case, we may ignore the problem of edge correction. Indeed, if the observation window is large compared to the range of distances \( r \) considered, only a few focal points \( i \) would be located close to the border. In this case, even summary statistics of unmarked patterns could be estimated, in a first approximation, without edge correction, using estimators analogous to those given in Equations 3.1 and 3.2 (Figure 2.13a).

Besides ignoring the edge correction problem, there are basically four different strategies to correct for the biases produced through edge effects. First, in minus sampling only the data in \( W \) that contain no bias are used in the analysis (Illian et al. 2008, p. 185f). In this method, only focal points \( i \) that are located at least distance \( r \) from the border of \( W \) are used in estimating the summary statistics (Figure 2.13b). This produces a buffer zone between the outer edge of \( W \) and the data employed in the analysis. In this case, the full circle or ring around point \( i \) is located within \( W \) and the summary statistics can be estimated without bias. Although this is an exact method that completely avoids bias, it has the disadvantage of not fully using the data that are available. Using this approach, the estimators given in Equations 3.1 and 3.2 are modified as follows:

\[
\hat{\pi}_e(r) = \frac{1}{n_{e\theta}} \sum_{i=1}^{n_{e\theta}} n_i^e(r) \tag{3.3}
\]

\[
\hat{D}(r) = \frac{1}{n_{e\theta}} \sum_{i=1}^{n_{e\theta}} 1(0 < d_i \leq r) \tag{3.4}
\]

where \( n_{e\theta} \) is the number of points in the reduced observation window \( W_{e\theta} \) (Stoyan et al. 2001).

Plus sampling is conceptually similar to minus sampling, but in this case missing information outside \( W \) is collected (or reconstructed). Note that this is only different from minus sampling for summary statistics based on nearest-neighbor distances, since second-order summary statistics require that all points be mapped, not only the nearest neighbors. If additional information outside \( W \) cannot be collected, or if the summary statistics require knowledge of all points within distance \( r \), one may also reconstruct the
missing points outside $W$ to ensure that all focal points in $W$ have a full neighborhood. A variety of methods exist for plus sampling. For example, periodic edge correction reconstructs the points outside $W$ by enlarging the point pattern in $W$ by periodic continuation outside $W$. (Periodic edge correction is sometimes referred to as toroidal edge correction, since, by attaching a copy of $W$ to each edge of $W$, ad infinitum, we effectively turn the plane defined by $W$ into a torus; Figure 2.13c.) In this case, the estimators given in Equations 3.1 and 3.2 can be used after redefining the distances between points, utilizing toroidal geometry. Conceptually, similar to this approach is the reflection method, where the points outside $W$ are not reconstructed by periodic continuation outside $W$, but by reflection at the border. However, these latter two methods are somewhat questionable, since points outside $W$ are not actually sampled, but are inferred. Plus sampling can be used, but only as a crude approximation for analyzing point patterns across a broader range of scales. The interested reader can refer to Illian et al. (2008, p. 183ff) for more details regarding the different plus-sampling methods.

A third strategy of edge correction does not attempt to reconstruct the pattern outside the observation window, but is based on individual assessment of each point or point pair. In the case of nearest-neighbor statistics, such an estimator considers only points $i$ for which the nearest neighbor is located within $W$ (i.e., $d_i \leq d_{i*}$; Figure 3.1) and uses a correction to compensate for the skipped points. This strategy for estimating the nearest-neighbor distribution functions goes back to Hanisch (1984) and will be called in the following Hanisch edge correction (Stoyan et al. 2001). For second-order statistics, these more sophisticated methods basically extrapolate the number of points contained within incomplete disks or rings to account for what is expected for complete disks or rings. In practice, this is accomplished by weighting each point pair with an appropriate correction factor. Differences among methods arise through the use of different weights. One example for this approach is a weight based on dividing the circumference of the full disk by the circumference of the disk lying inside $W$ (i.e., Ripley edge correction). In this case, if one-third of the disk lies within $W$, the weight would be 3.

Finally, in a fourth strategy, the estimator itself is adapted to compensate for the bias by multiplying the naive estimator given in Equation 3.1 with a correction factor that depends only on the distance $r$, but not on a given point pair. In this case, the geometry of $W$ is used to calculate the expected bias. This strategy for estimating second-order statistics can be traced to Ohser and Mücklich (2000) and will be called in the following Ohser edge correction. The first two strategies (i.e., minus and plus sampling) are thoroughly described in the literature (e.g., Haase 1995; Yamada and Rogerson 2003; Illian et al. 2008, pp. 183–186), but are only rarely used in modern applications of point-pattern analysis. We therefore focus here on the more sophisticated estimators in which edge correction weights and adapted estimators are used.

The selection of the appropriate edge correction method depends critically on the purpose of the analysis. If the absolute value of the summary statistics
matters, for example, when fitting a specific point-process model to the data, an edge correction method is required that minimizes the bias. However, if the purpose is to detect spatial patterns in the data, for example, by comparing the observed summary statistics with those of null models, the bias in the summary statistic may cancel out because it appears in both the summary statistics of the observed and simulated data.

3.1.2 Second-Order Summary Statistics for Univariate Patterns (Data Type 1)

3.1.2.1 Different Strategies for Deriving Edge Correction Weights

Four different strategies for deriving edge correction weights \( w_{ij} \) for second-order summary statistics are useful in most practical applications. They all attempt to compensate for the missing area of rings, which are located outside the observation window. Differences among the second-order estimators discussed in this book arise only due to different strategies for deriving the weights \( w_{ij} \) that accomplish edge correction.

The first strategy for deriving individual weighting factors \( w_{ij} \) goes back to Ripley (1976). *Ripley edge correction* has an appealingly intuitive geometric interpretation. The Ripley weighting factor \( w^R_{ij} \) basically scales the area of an incomplete circle, with an origin at point \( i \) that passes through point \( j \), by the area of the complete circle. For example, if only half of a circle is located within \( W \), the weight yields \( w^R_{ij} = 2 \). Thus, the estimator evaluates all pairs of points (with focal point \( i \) and a second point \( j \)) that are located approximately at distance \( r \) apart and weights each observation by \( w^R_{ij} \). If a circle with radius \( r \) around a focal point \( i \) is completely inside the observation window, there is no need for edge correction, that is, \( w^R_{ij} = 1 \). However, if the circle is located only partly inside the observation window, as with circle \( C_2 \) in Figure 3.1, edge correction is necessary to compensate for points potentially located on the circle outside the observation window \( W \) (such as point \( p_3 \) in Figure 3.1) and weights of \( w^R_{ij} > 1 \) are applied.

Ripley edge correction works by first identifying all the points of intersection between a circle of radius \( r \) centered on focal point \( i \) and the observation window \( W \) (i.e., the small open circles in Figure 3.2). Next, the angles of the “pieces of cake” that lie partly outside of the observation window are determined (e.g., \( \phi_1 \) and \( \phi_2 \) in Figure 3.2a and \( \phi_1 \) in Figure 3.2b). The arcs of these angles, which represent the arc length of the circle that is located outside the observation window (fine lines in Figure 3.2), are then summed up. The Ripley weights are then obtained by dividing the arc length of the full circle (i.e., \( 2\pi r \)) by the arc length of the circle located within \( W \) (\( \phi_{ij} \), bold lines in Figure 3.2). For example \( \phi_{ij} = 2\pi - \phi_1 - \phi_2 \) in Figure 3.2a. The resulting weight is

\[
\phi^R_{ij} = \frac{2\pi}{\phi_{ij}}
\]  

(3.5)
where \( \varphi_{ij} \) is the sum of all the central angles of the arcs lying entirely within \( W \) of the circle with center \( x_i \) and radius \( r_{ij} = \|x_i - x_j\| \). If the circle is completely inside the rectangular observation window \( W \), we find that \( w_{ij}^R = 1 \). In the most extreme case, where point \( i \) is located on one of the four corners of the rectangle \( W \), we find that \( w_{ij}^R = 4 \).

The Ripley edge correction method is somewhat sensitive to clusters close to the corners (or boundary) of \( W \), because the points of these clusters receive high weights when they are treated as focal points \( i \). This may lead to overestimation of the degree of clustering in the pattern. Calculation of the weights \( w_{ij}^R \) involves basic trigonometric calculations, but is somewhat complicated by the various cases in which a circle can intersect \( W \). Appendices B.4 and B.5 of Illian et al. (2008, p. 486f) provide elegant computer code for calculating the weights for \( w_{ij}^R \). Goreaud and Pelissier (1999) also provide explicit formulas, even for irregularly shaped observation windows.

An alternative strategy for deriving edge correction weights \( w_{ij} \) was proposed by Wiegand and Moloney (2004). Instead of assigning each point pair \( i-j \) an individual weight, as in the Ripley method, the \( WM \) edge correction addresses bias due to edge effects globally by using the same weight for all point pairs separated by a given distance \( r \). This is done by dividing the area of a complete ring with radius \( r \) and width \( dr \) by the mean area \( \bar{v}(r) \) of all the rings (inside \( W \)) with radius \( r \) (and width \( dr \)) centered on the points of the pattern, that is, \( w_{ij}^{WM} = 2\pi dr / \bar{v}(r) \) (Figure 3.1). (A more formal approach to defining \( \bar{v}(r) \) will be provided in Section 3.1.2.2 below.) Note that \( w_{ij}^{WM} \) belongs to the “Ohser” group of estimators that do not correct for the contribution of individual point pairs \( i \) and \( j \) to the estimator, but correct for the bias in the naive estimator with a factor that only depends on the distance \( r \) between points and the locations of all points of the pattern.

![Figure 3.1](image_url)
The Ohser edge correction goes back to Ohser and Mucklich (2000) and is related to the WM edge correction. As with WM edge correction, the Ohser weight corrects only the expected final bias. This is accomplished by using a weighting factor $w_{ij}^O$, which is only a function of distance $r$, not of the specific locations of points. Recall that the denominator of the WM weight is the mean area of all the rings (inside $W$) with radius $r$ (and width $dr$) centered on the points of the pattern. By dividing with the ring width $dr$ we obtain the mean length of the circumferences of circles with radius $r$ around the points $i$ of the pattern that lie inside $W$. The Ohser weight is based on an idea similar to the WM weights, but instead of using the actual points of the pattern, it calculates the expected length of the circumference of a randomly placed circle with radius $r$ lying within $W$. This is accomplished with the weight:

$$w^O(r) = A/\overline{\gamma}_W(r)$$

where $\overline{\gamma}_W(r)$ is the so-called isotropized set covariance, and $A$ the area of the observation window.

The isotropized set covariance $\overline{\gamma}_W(r)$ is a quantity developed in set theoretical analysis (e.g., Stoyan and Stoyan 1994; Chapter 8). It is based on the set covariance $\gamma_W(r)$ for a rectangular observation window $W$ with sides of length $a_x$ and $b_y$ (and area $A = a_x b_y$). Essentially, $\gamma_W(r)$ is the area of the intersection of $W$ with $W_r$ (the hatched area in Figure 3.3), where $W_r$ is the rectangle $W$ shifted in space by a vector $r = (r_x, r_y)$ and $\gamma_W(r) = (a_x - r_x)(b_y - r_y)$. Thus, $\gamma_W(r)$ is the area of overlap between the original rectangle and the rectangle shifted by $r$. If $r = 0$, then $\gamma_W(r) = A$. Using polar coordinates, the vector $r$ can be rewritten as $r = (r, \phi)$. The isotropized set covariance function $\overline{\gamma}_W(r)$ for rectangle $W$ is then the average set covariance $\overline{\gamma}_W(r)$ taken over all angles $\phi$ (Figure 3.4).

**FIGURE 3.3**

Isotropized set covariance. (a) Definition of the set covariance. The set covariance is the area of the intersection of window $W$ and the translated window $W_r$ where $r = (r, \phi)$ is a vector. (b) Calculation of the isotropized set covariance. A point $x$ on $W$ is shifted by vector $r$ and counted if it is located within $W$. The result is that all rotations $\phi$ of $r$ located in $W$ are counted, which produces an estimate of the boundary length of the portion of the circle with center $x$ and radius $r$ lying in $W$. This is repeated for all $x$ in $W$, thereby yielding the isotropized set covariance.
It is, therefore, the expected intersection of $W$ and the translated window $W_r$ without specifying the angle $\phi$.

For a rectangle with sides $a_x \leq b_y$, the isotropized set covariance can be calculated analytically (Stoyan and Stoyan 1994, p. 123). For the range $r \leq a_x$, which is relevant for point-pattern analysis, this yields

$$\gamma_W(r) = \pi \left( a_x + b_y \right) + \frac{r^2}{\pi}$$

Thus, $\gamma_W(r)$ decreases almost linearly with distance $r$.

Initially, it might be surprising that calculating the overlap between the original rectangle $W$ and the set of shifted and rotated rectangles $W_r$ might be an appropriate weight for edge correction, that is, $w^O(r) = A/\gamma_W(r)$. However, there is a relatively simple geometric explanation. As shown in Figure 3.3, calculation of the set covariance $\gamma_W(r)$ involves shifting each point $x$ within $W$ by vector $r$ and integrating over all shifted points $x + r$ that are still located within $W$ (Figure 3.3a). The isotropized set covariance $\gamma_{W}(r)$ is then calculated by rotating $W_r$ over all angles $\phi$ and integrating the resulting values of $\gamma_W(r)$ for all rotations $\phi$ (Figure 3.4). If we switch the order of integration, we can obtain a quantity related to the Ohser weight as follows: First, we integrate over all points lying inside $W$ that result from rotating the vector $r = (r, \phi)$ for $\phi \in (0, 2\pi)$ around a point $x$ that also lies in $W$. This yields the length of the circumference of the circle with radius $r$ around point $x$ that lies within $W$ (Figure 3.3b). In the second step, we integrate the step over all points $x$ that lie within $W$. The result is then divided by $A$, the area of $W$. This gives us the expected length of the circumference of a randomly placed
circle, with radius \( r \), lying inside \( W \), that is, \( 2\pi \gamma_W(r)/A \). Taking the inverse and multiplying by \( 2\pi \) yields the Ohser weight.

For a random pattern, the WM weight approximates the Ohser weight, as can be seen in Figure 3.5b. The expected area of incomplete rings with radius \( r \) and random location within \( W \) captured by the Ohser weight (solid line in Figure 3.5b) is closely approximated by the actual area of incomplete rings with radius \( r \), centered on the points of the pattern that are used in calculating the WM weight (closed circles in Figure 3.5b). Interestingly, there are two small systematic departures between the WM and Ohser weights. These arise due to the fact that the WM weights are based on a single realization of a random process. The mean area of the “real” rings is somewhat less than expected up to a radius of 2 m. This indicates that a few more points than expected are located close to the border of \( W \). In contrast, the area of the real rings is slightly greater than expected at distances of 6–8 m. Thus, the WM weights are “adapted” to the small irregularities of the pattern and may provide slightly better results than the Ohser weights. However, the Ohser weights do not depend on the actual locations of the points of the pattern and, therefore, have the advantage of allowing rapid computation. In a later section we will see that this difference is related to the so-called “adapted” estimator of the intensity \( \lambda \) used for estimation of the pair-correlation function (see Section 3.1.2.4).

A fourth form of edge correction, the Stoyan edge correction, is the one favored by Stoyan and Stoyan (1994) and Illian et al. (2008). It is based on the set covariance \( \gamma_W(r) \), which for a rectangle with sides of lengths \( a_x \) and \( b_y \) yields

\[
\omega_{i,j}(r) = A / \gamma_W(x_i - x_j)
\]

(3.8)

**Figure 3.5**
Ohser correction factor. (a) Random pattern of 100 points within a 20 × 20 m² observation window. (b) The inverse of the Ohser edge weight \( \frac{\gamma_W(r)}{A} \) (solid gray line) and inverse of the WM edge correction weight (closed disks) for the pattern in panel (a).
where $\gamma_W(r) = |(a_x - r_x)(b_y - r_y)|$, as before. Thus, the Stoyan weight follows the first step of the calculation of the Ohser weights. It shifts each point $x$ within $W$ by vector $r$ and integrates over all shifted points $x + r$ that are still located within $W$, resulting in a calculation of the set covariance $\gamma_W(r)$. It then uses the set covariance $\gamma_W(r)$ as a weight, thereby summing up and weighting all point pairs in $W$ that have distance $r$ separating them. Because the set covariance yields only a value of $A$ (i.e., the area of $W$) when $r = 0$, all pairs of points separated by a distance $r > 0$ receive a weight greater than 1. Point pairs with small separation distances $r$ receive, in general, small weights, whereas pairs of points separated by larger distances $r$ receive larger weights. In contrast, the Ripley edge correction method weights only point pairs defining incomplete circles within $W$. In the latter case, only points close to the border are given weights greater than 1. The Stoyan weights, therefore, also consider the explicit locations of the points and correct for smaller amounts of stochastic irregularity in the pattern.

The formulas for the four edge correction weights are given below in Equation 3.9. The Ripley and Stoyan weights are applied as individual weights to each $i$–$j$ point pair, whereas the Ohser weights are completely independent of the pattern being analyzed and depend only on the geometry of the observation window $W$ and distance $r$. The WM weights are intermediate to these two extremes as they are specific to the given pattern analyzed, but are not applied to the specific point pairs $i$–$j$, depending only on the geometry of $W$ and distance $r$, like the Ohser weights.

\[
w_{\text{WM}}(r) = \frac{2\pi rdr}{\delta(r)} \quad \text{WM}
\]
\[
w_{\text{R}}(r) = \frac{2\pi}{\varphi_{ij}} \quad \text{Ripley}
\]
\[
w_{\text{S}}(r) = \frac{A}{\gamma_W(x_i - x_j)} \quad \text{Stoyan}
\]
\[
w_{\text{O}}(r) = \frac{A}{\gamma_W(r)} \quad \text{Ohser}
\]

In the next sections, we will combine the edge correction weights with the estimators for the different second-order statistics and explore the implications of using the different weighting schemes.

### 3.1.2.2 O-Ring Statistic

Now that we have introduced several approaches to edge correction, we are in a position to present the methods for estimating the various second-order point-pattern statistics. The first estimator we consider is the O-ring statistic.
O(r). O(r) can be defined as the density of points within a ring of radius r and width dr around the typical point of the pattern (gray ring in Figure 3.1). Note that O(r) is a conditional summary statistic; meaning that it yields the expected density of points within a ring, given the condition that there is a point at the center of the ring. This characteristic also means that the O-ring statistic is a neighborhood density function and, in some sense, a point-related analog to the location-related intensity function λ(x). The naive estimator of O(r) is also analogous to the estimator of λK(r) given in Equation 3.1 and is defined as follows:

\[ \hat{O}(r) = \frac{n^R(r)}{2\pi dr} = \frac{1}{n} \sum_{i=1}^{n} n_i^R(r) \]  

where \( n_i^R(r) \) is the number of points located in a ring with radius r and width dr around point i (Figure 3.1). The superscript \( R \) refers to “ring.” This is summed over all n points i in the pattern and divided by the area \( 2\pi dr \) of a ring of radius r to obtain a density. Clearly, this estimator does not correct for edge effects. A simple way to modify Equation 3.10, when some rings are not completely located within \( W \), was proposed by Wiegand and Moloney (2004). This results in the WM edge correction factor described above. The basic idea is to estimate the mean number of points within complete and incomplete rings around points i, that is, \( \sum n_i^R(r)/n \), and divide by the mean area of these rings. More technically, the neighborhood density is estimated as a point-related analog to the estimation of the location-related intensity \( \lambda \), based on several plots (see Section 3.3.1). In essence, this means that the number of points \( n_i \) is counted within \( m \) subwindows \( W_i \) with area \( v_i \) located within a larger area \( W \) (Figure 3.6a). Following Illian et al. (2008, p. 261), the intensity of points within \( W \) can be estimated based on the counts within the \( m \) subwindows \( W_i \):

\[ \hat{\lambda} = \sum_{i=1}^{m} \hat{\lambda}_i \frac{v_i}{v} \]  

where \( \hat{\lambda}_i \) is the estimated intensity of points within the \( i \)th subwindow \( W_i \) (=\( n_i/v_i \)), and \( v = \sum_{i=1}^{m} v_i \). We can rearrange Equation 3.11 to produce

\[ \hat{\lambda} = \frac{\sum_{i=1}^{m} \hat{\lambda}_i v_i}{\sum_{i=1}^{m} v_i} = \frac{1}{m} \sum_{i=1}^{m} \frac{n_i}{v_i} \]  

(3.12)
which shows that this indeed yields the mean number of points within the $m$ subwindows $W_i$ divided by the mean area of these subwindows. This idea can be translated into an approach for estimating the neighborhood density function $O(r)$. In this case, the different subwindows are the rings $R_i$ around the points $i$ of the pattern (Figure 3.6b) and the resulting estimator is given by

$$\hat{O}^{WM}(r) = \frac{1}{n} \sum_{i=1}^{n} \frac{n_i^R(r)}{v_i} = \frac{\overline{n_i^R(r)}}{\overline{v}(r)},$$

(3.13)

where $v_i$ is the area of a potentially incomplete ring in $W$ around point $i$ and $n_i^R(r)$ is the number of points within that ring. Thus, the neighborhood density is estimated by dividing the average number of points in the rings surrounding the points of the pattern by the average area within each ring that lies in $W$.

Comparison of the naive estimator of the $O$-ring statistic given in Equation 3.10 with the method proposed in Wiegand and Moloney (2004) reveals that the area $2\pi r dr$ of the complete ring is replaced by the mean area of the rings within $W$ for the given point pattern. Basically, Equation 3.13 is obtained by multiplying Equation 3.10 by the WM edge correction factor $\omega_{i,j}^{WM} = 2\pi r dr / \overline{v}(r)$.
In essence, this estimator is “adapted” to the point pattern because the same estimation rule is used in the numerator and denominator; in both cases, all points \( i \) are visited in turn and some property of the ring around point \( i \) is evaluated. It can, therefore, be expected that this estimator will compensate for stochastic irregularities in the distribution of the points and will be robust to small departures from heterogeneity. Note also that this estimator belongs to the fourth strategy of edge correction methods presented above (i.e., Ohser edge correction), which corrects for edge effects by multiplying the naive estimator (Equation 3.10) by a correction factor, yielding in this case \( 2\pi dr / \sigma(r) \).

The correction factor depends only on the distance \( r \) (and the locations of all points of the pattern), but not on a given point pair. Figure 3.5b shows the bias in \( \sigma(r)/2\pi dr \) (i.e., the mean area of the incomplete rings relative to that of a complete ring) resulting from the pattern shown in Figure 3.5a. This pattern comprises 100 points within a \( 20 \times 20 \) m\(^2 \) plot. Interestingly, the bias increases almost linearly with distance \( r \) and reaches a value of 0.45 for a distance that corresponds to half of the width of the plot (i.e., \( r = 10 \) m). This shows that the bias cannot be ignored in most practical applications of the neighborhood density function and that edge correction is required.

Although the approach to estimation discussed above is geometrically intuitive, we need to introduce a more general approach for incorporating edge correction into our estimators. With this in mind, a modified approach based on the notation provided by Illian et al. (2008) will be first used to redefine the estimator presented in Equation 3.13. To do this, we need to introduce a general class of functions known as the kernel function \( k(x) \). A kernel function essentially evaluates the distance between two points and returns a value based on that distance. Using this, we rearrange the numerator of Equation 3.13 by incorporating a kernel function in its definition. This is used to determine the number of points falling at distance \( r \) from a focal point \( i \), which is given by \( n_i^K(r) \) as follows:

\[
n_i^K(r) = \sum_{j=1, j\neq i}^{n} k(\|x_i - x_j\| - r)
\]  

The sum on the right-hand side of Equation 3.14, essentially compares all points of the pattern \( j \) to point \( i \), where \( x_i \) and \( x_j \) are the coordinates of points \( i \) and \( j \), respectively, and \( d_{ij} = \|x_i - x_j\| \) is the distance separating the two points. The kernel function evaluates to 1 if the distance between \( i \) and \( j \) is close to the distance \( r \), otherwise it evaluates to 0. Thus, what we described more loosely as a ring (e.g., R1–R4 in Figure 3.6b) can be described more formally with a kernel function \( k(x) \) which determines whether or not two points \( i \) and \( j \) are located distance \( r \) apart. The kernel function introduces a small “tolerance” interval \((r - dr/2, r + dr/2)\) for the distance \( r \) (called the bandwidth \( h = dr/2 \)) within which two points are regarded as being located distance \( r \) apart. In this way, the kernel function defines rings with width \( dr \) and radius \( r \) around
the focal point \(i\) (Figure 3.6b). Any point falling within the ring causes the kernel function to evaluate to a positive value, otherwise it returns zero.

Note that the bandwidth should not be too small, otherwise the number of points \(n_r^{\hat{K}}(r)\) within the ring will be low and cause spurious stochastic fluctuations in the estimator. However, bandwidths that are too large will produce smooth estimates, but important details regarding the pattern may be lost. Illian et al. (2008, p. 236) recommend, as a rule of thumb, an initial value of \(h \approx 0.1/\lambda^{0.5}\). Thus, a pattern with 100 points within a 100 × 100 m\(^2\) observation window will require a bandwidth of \(h = 1\) m, and 100 points within a 500 × 500 m\(^2\) observation window will require a bandwidth of \(h \approx 5\) m. If the pattern is clustered, a somewhat smaller bandwidth could be selected, since the neighborhood density at small distances \(r\) is larger and more points are located within the rings.

Figure 3.7 shows two popular kernel functions. The box kernel (Figure 3.7a) is the simplest kernel function and evaluates as

\[
k^B_h(d_{ij} - r) = \begin{cases} 
\frac{1}{2h} & |d_{ij} - r| \leq h \\
0 & \text{otherwise}
\end{cases}
\]  

(3.15)

where \(h\) is the bandwidth, that is, \(dr/2\), and \(d_{ij}\) is the distance between points \(i\) and \(j\) (Figure 3.1). Illian et al. (2008) recommend use of this kernel function in the estimation of the product density. Note that the kernel function requires a small correction if \(r - h/2 < 0\). Another popular kernel function is the Epanechnikov kernel, which was already presented in Box 2.4 in the context of nonparametric intensity estimators (Figure 3.7b):

\[
k^E_h(d_{ij} - r) = \begin{cases} 
\frac{3}{4h} \left(1 - \frac{(d_{ij} - r)^2}{h^2}\right) & |d_{ij} - r| \leq h/2 \\
0 & \text{otherwise}
\end{cases}
\]  

(3.16)

FIGURE 3.7
Kernel functions for selecting points separated by a distance of approximately 4 units (i.e., \(r = 4\)), using a bandwidth \(h\) of one unit. (a) The box kernel. (b) The Epanechnikov kernel.
Both kernel functions yield values of zero if \( d_{ij} < r - h \) or \( d_{ij} > r + h \) (i.e., they consider only points \( j \) within the ring with radius \( r \) and width \( 2h \) around the focal point \( i \)), but the Epanechnikov kernel weights point pairs with a separation distance \( d \) closer to \( r \) higher than those that are less close to \( r \) (Figure 3.7b).

A second modification, which formalizes the notation of the denominator of Equation 3.13, is more complicated. First, a disk with radius \( r \) and center \( x_i \) for point \( i \) is written as \( b(x_i, r) \) and the circumference of \( b(x_i, r) \) is written as \( \partial b(x_i, r) \). Next, the symbol \( \cap \) is used to define the intersection of two geometric elements, which in our case is the intersection of the circumference of a disk centered on \( x_i \) with the observation window \( W \), that is, \( W \cap \partial b(x_i, r) \). To express the length of the circumference located in \( W \), the function \( v_{d-1} \) is used which gives the length of a one-dimensional geometric object (i.e., a line; \( d = 2 \)), thus \( v_{d-1}(W \cap \partial b(x_i, r)) \) in formal notation. With these definitions the translation of the estimator given in Equation 3.13 into formal language yields:

\[
\hat{O}_{WM}(r) = \frac{1/n \sum_{i=1}^{n} \left( \sum_{j=1}^{n} k_i^j \left( \|x_i - x_j\| - r \right) \right)}{1/n \sum_{i=1}^{n} v_{d-1}(W \cap \partial b(x_i, r))}
\] (3.17)

A natural estimator of the pair-correlation function \( g(r) \) based on 3.17 is to divide 3.17 by an estimate of the intensity, since \( O(r) = \lambda g(r) \). The simplest way to accomplish this is to use the “natural” intensity estimator for this purpose, which is given by the number of points within \( W \) divided by the area of \( W \); that is, \( \lambda_n(r) = n/v_n(W) = n/A \). In later sections we discuss other options for estimating the intensity using “adapted intensity estimators” (see Section 3.1.2.4).

### 3.1.2.3 The Product Density

A more formal method of estimating second-order summary statistics is based on the so-called “product density” \( \rho^{(2)}(x_1, x_2) \) briefly introduced in Sections 2.3.1.1 and 2.3.2 (see also Figure 2.6b). This method is used in Illian et al. (2008). For greater clarity in notation we will omit the superscript (2) in the following sections.

For homogeneous patterns, the product density \( \rho(r) \) is the conditional probability that one point of a given point process occurs within a disk with area \( dx_2 \) located at \( x_2 \) given that another point is located at distance \( r \) within a disk with area \( dx_1 \) located at \( x_1 \). This is, in fact, related to the neighborhood density function \( O(r) \) as follows:
The general form of the estimators of the product density for univariate patterns is given by

\[ \hat{\rho}(r) = \hat{\lambda} n n A \]

\[ \hat{\rho}(r) = \frac{1}{2\pi r} \frac{1}{A} \sum_{i=1}^{n} \sum_{j=1}^{n} k(\|x_i - x_j\| - r)w_{i,j} \]

\[ = \frac{1}{2\pi r} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} k(\|x_i - x_j\| - r)w_{i,j}dr \]

(3.19)

The second expression for the function shown in Equation 3.19, incorporating the factor \( \hat{\lambda}_n \), follows simply from the first expression, if we multiply the numerator and denominator by \( n \) and \( dr \). The modified double sum can now be interpreted as the mean number of points located at a distance interval \((r - dr/2, r + dr/2)\) from the observed points of the pattern, with \( n/A \) yielding the natural estimator \( \hat{\lambda}_n \) of the intensity. Written in this form, the equation can be interpreted as the following: the intensity \( \hat{\lambda}_n \) represents the probability of having a point at a given focal location, and \( 1/n \sum_{i=1}^{n} \sum_{j=1}^{n} k(\|x_i - x_j\| - r)w_{i,j}dr \) is an estimate of the expected number of points located at a distance interval \((r - dr/2, r + dr/2)\) away from the focal location. Note that each point pair in the double sum is adjusted by an edge correction weight \( w_{i,j} \), which compensates for rings that are partly located outside the observation window (Figure 3.1). Finally, the denominator in the second version of Equation 3.19 yields the area of a ring with radius \( r \) and width \( dr \). Thus, we divide the expected number of points located within the ring \((r - dr/2, r + dr/2)\) centered on the focal location by the area of the ring, yielding the probability that there is an additional point at any location distance \( r \) away from the focal point.

Slight rearrangement of Equation 3.19 shows that this estimator basically multiplies the natural estimate \( \hat{\lambda}_n = n/A \) of the intensity by the estimated mean density of points within rings with radius \( r \) around the points \( i \) of the pattern:

\[ \hat{\rho}(r) = \left( \frac{n}{A} \hat{\lambda}_n \right) \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\sum_{j=1}^{n} k(\|x_i - x_j\| - r)w_{i,j}}{2\pi r} \right) \]

(3.20)

This is reasonable since \( \rho(r) = \lambda O(r) \) (Equation 3.18).
We can incorporate the four different weights adjusting for edge effects into the univariate product density estimator as follows:

\[
\hat{\rho}^k(r) = \frac{1}{2\pi r A} \sum_{i=1}^{n} \sum_{j=1}^{n} n_{ij} k(\|x_i - x_j\| - r) \left[ \frac{2\pi}{\varphi_{ij}} \right]
\]
Ripley

\[
\hat{\rho}^\gamma(r) = \frac{1}{2\pi r A} \sum_{i=1}^{n} \sum_{j=1}^{n} n_{ij} k(\|x_i - x_j\| - r) \left[ \frac{A}{\gamma_W(r)} \right]
\]
Stoyan

\[
\hat{\rho}^\Omega(r) = \frac{1}{2\pi r A} \sum_{i=1}^{n} \sum_{j=1}^{n} n_{ij} k(\|x_i - x_j\| - r) \left[ \frac{A}{\gamma_W(r)} \right]
\]
Ohser

\[
\hat{\rho}^{WM}(r) = \frac{1}{2\pi r A} \sum_{i=1}^{n} \sum_{j=1}^{n} n_{ij} k(\|x_i - x_j\| - r) \left[ \frac{2\pi r}{(1/n) \sum_{i=1}^{n} v_{d-1}(W \cap \partial b(x_i, r))} \right]
\]
WM

(3.21)

A direct comparison of the weights resulting from the different edge correction methods for a random pattern is illustrated in Figure 3.8. Since the Ohser and WM weights depend only on the distance between the two points i and j, not on their location, they are approximately equal in value and cannot be distinguished in Figure 3.8. Interestingly, there are in general only small differences between the weights calculated by the Ohser and the Stoyan methods. In contrast, the Ripley weights may be much larger than the other weights, if the focal point is located close to the boundary of W. On average, the Ripley weights correspond, in good approximation, to the other weights, as can be seen by the Ripley regression lines shown in Figure 3.8. However, in the example shown in Figure 3.8, there is some departure for estimates at a distance of 150 m (Figure 3.8c). Thus, in general, we can expect that the four alternative estimators of the product density will yield similar results, since the noise of the Ripley and Stoyan weights will average out, even though the noise in the Ripley weights is somewhat greater.

Estimates of the product density utilizing the different edge correction techniques for a CSR pattern of 500 points (Figure 3.9a) within a 100 x 100 m² observation window are given in Figure 3.9b. We find that the estimates of the product density that consider edge correction for a random pattern are essentially the same for the four edge correction methods but that the naive estimator yields a considerable bias (Figure 3.9b). In contrast, Figure 3.9d explores how the estimators behave in the case of a strongly clustered pattern shown in Figure 3.9c, where some clusters are accidently located at the corners of the observation window W. The Ohser and Stoyan estimators cannot be easily distinguished (their curves overlap), but the Ripley estimator,
as expected, indicates somewhat larger clustering at smaller distances (i.e., \( r < 15 \)) (Figure 3.9d). This is consistent with the scale of the clusters in the lower and top left corners and occurs since the points located in the clusters receive greater weight, due to their proximity to the border. The result is a slightly higher estimate of the product density over a range of scales approximating the cluster size. The WM estimator ranges somewhat between the Ohser and the Ripley estimator and exhibits smaller fluctuations at larger scales (i.e., \( r > 15 \)). However, note that the difference among estimators will appear smaller when the product density is plotted on a linear scale.

### 3.1.2.4 Estimators of the Pair-Correlation Function

A simple estimator of the pair-correlation function \( \hat{g}(r) \), based on the different estimators of the product density shown in Equation 3.21, would be
to use the “natural” estimator of \( \hat{\lambda}_m^2 = n(n-1)/A^2 \), that is, \( \hat{g}(r) = \hat{\rho}(r)/\hat{\lambda}_m^2 \), since \( \rho(r) = \lambda^2 g(r) \). Note that \( n-1 \) is required in the numerator of the natural estimator, since the double sum of the estimator in Equation 3.21 considers only points \( i \neq j \). Although these estimators were used in earlier studies (e.g., Stoyan and Stoyan 1994), Stoyan and Stoyan (2000) and Illian et al. (2008) now recommend use of an estimator of \( \lambda \) that is especially adapted to the estimator of the product density. The objective of this is to reduce variability in the estimator.

The need to use a correction in estimating the intensity \( \lambda \) for the pair-correlation function can be understood from Figure 3.5b. The figure compares the mean area of the partly incomplete rings around the points of the pattern (dots) with the expectation for randomly distributed points provided by the isotropized set covariance (line). Small irregularities may cause slightly more (or fewer) points than expected being located close to the border. This decreases (increases) the mean area of the rings (as shown

---

**FIGURE 3.9**
Comparison of estimators of the product density. (a) Random pattern. (b) Estimators for the random pattern in panel (a). (c) Clustered pattern. (d) Estimators for the clustered pattern in panel (c). Note the logarithmic scale used in panel (d).

© 2014 Taylor & Francis Group, LLC
in Figure 3.5b), whereas the expectation of the isotropized set covariance produces no such irregularities. To compensate for stochastic variation in estimating the pair-correlation function, Stoyan (2006a; his Equation 19) proposed an adapted estimator of the intensity $\hat{\lambda}$ given by

$$
\hat{\lambda}_S(r) = \frac{\sum_i^n v_{d-1}(W \cap \partial b(x_i, r))}{2\pi r \bar{\gamma}_W(r)}
$$

(3.22)

where $v_{d-1}(W \cap \partial b(x_i, r))$ is the length of the circumference of a circle of radius $r$ around point $i$ in $W$. This formal notation was introduced in Section 3.1.2.2. The length of the circumference of a circle of radius $r$ around point $i$ in $W$ yields $2\pi r$, if the circle is fully located within $W$, but it may be smaller if the circle only partly overlaps $W$. By factoring out the natural estimator $\hat{\lambda}_n(r) = n/A$ of lambda and slightly rearranging Equation 3.22 we find that

$$
\hat{\lambda}_S(r) = \frac{\sum_i^n v_{d-1}(W \cap \partial b(x_i, r))}{2\pi r \bar{\gamma}_W(r)}
$$

$$
= \hat{\lambda}_n \frac{(1/n) \sum_i^n v_{d-1}(W \cap \partial b(x_i, r))}{2\pi} \frac{A}{\bar{\gamma}_W(r)}
$$

$$
= \hat{\lambda}_n \frac{w_{ij}^O(r)}{w_{ij}^{WM}(r)} = \hat{\lambda}_n \times \text{corr}(r)
$$

(3.23)

Thus, the adapted estimator of lambda $\hat{\lambda}_S(r)$ is basically the natural estimator $\hat{\lambda}_n(r)$ multiplied by the ratio of the Ohser and the WM edge correction weights, that is, corr(r). The correction term corr(r), itself, is basically the ratio of the average length of the circumference of circles centered on points $i$ lying inside of $W$ to the circumference expected for random points in $W$.

Further consideration also shows that the WM estimator of the neighborhood density function $O(r)$ given in Equation 3.17 is identical with the estimator that results from combining the estimator of the product density based on Ohser edge correction (Equation 3.21) with the adapted intensity estimator in Equation 3.22, that is, $\hat{O}^{WM}(r) = \hat{\rho}^O(r)/\hat{\lambda}_S(r)$. This integrates the estimator of the $O$-ring statistic suggested in Wiegand and Moloney (2004) into the more general framework of estimators presented in Illian et al. (2008).

The main effect of using $\hat{\lambda}_S(r)$ to calculate the pair-correlation function is that the mean-squared error (mse) of the estimate is smaller than it would be if we used the natural estimator of $\lambda$ (Illian et al. 2008, p. 232). Figure 3.10 illustrates the correction factor applied to different patterns. For the random
pattern shown in Figure 3.10a, the adapted estimator departs somewhat from the expectation, but only a small correction is required. The points of the pattern exhibit only a slight tendency to be located away from the border (i.e., the area of the corresponding rings is slightly larger than expected) yielding a correction \(>1\) at smaller distances (Figure 3.10d). However, in the case of the clustered pattern, where some clusters with many points are located close to the (lower and upper left) corners (Figure 3.10b), we find that the rings around the points of the pattern have, on average, a smaller area than expected inside \(W\). Consequently, the Ohser weight yields an underestimation that must be corrected for by a value of \(\hat{\lambda}_S(r) < \hat{\lambda}_n\) (Figure 3.10e). For illustrative purposes, we also show the intensity estimates for an extreme pattern, as shown in Figure 3.10c that has no points close to the border (Figure 3.10f). In this case, the rings around the points of the pattern have, on average, a larger area than expected inside \(W\) (because they are mostly inside \(W\)) leading to a correction with \(\hat{\lambda}_S(r) > \hat{\lambda}_n\) (Figure 3.10f).

Following the suggestion by Stoyan (2006a) and Illian et al. (2008), we therefore obtain three alternative estimators for the pair-correlation function based on the different edge correction methods (i.e., the Ripley, Stoyan, and Ohser weights shown in Equation 3.21) that divide the estimators for the product by the square of the adapted intensity estimator (Equation 3.22):
\[ \hat{g}_R(r) = \frac{\hat{\rho}_R(r)}{(\hat{\lambda}_S(r))^2} \quad \text{Ripley} \]
\[ \hat{g}_S(r) = \frac{\hat{\rho}_S(r)}{(\hat{\lambda}_S(r))^2} \quad \text{Stoyan} \]
\[ \hat{g}_O(r) = \frac{\hat{\rho}_O(r)}{(\hat{\lambda}_S(r))^2} \quad \text{Ohser} \]
\[ \hat{g}_{\text{WM}}(r) = \frac{\hat{\rho}_{\text{WM}}(r)}{\hat{\lambda}_n^2} = \frac{\hat{\rho}_O(r)}{\hat{\lambda}_S(r) \hat{\lambda}_n} \quad \text{WM} \]

(3.24)

For completeness, we also show the estimator \( \hat{g}_{\text{WM}}(r) \) of the pair-correlation function presented in Wiegand and Moloney (2004), which divides the estimator of the O-ring statistic by the natural estimator of the intensity (i.e., \( n/A \)). Equation 3.20, however, shows that the estimator of the product density contains the naive estimator of the intensity. This suggests that it may be more appropriate to divide the estimator of the product density (Equation 3.21), one time, by the naive estimator of the intensity and the second time by the adapted estimator, as is done with the WM estimator.

We apply the different estimators to an extreme case of a heterogeneous pattern to better understand their behavior. We do this using a random pattern within a 100 × 100 m² area (denoted \( W_i \)) with area \( A_i \), which is enlarged by an empty buffer of 50 m, thus yielding an observation window \( W \) of 200 × 200 m² (Figure 3.11). The expected average neighborhood density of this pattern can be calculated analytically yielding \( O_{\exp}(r) = (n/A_i)(\bar{f}_{W_i}(r)/A_i) \) (Figure 3.11c). The expected average neighborhood density of the points of this pattern is always smaller than \( n/A_i \) because rings close to the border of the patch overlap the patch only partly. The product densities calculated with Equation 3.21 differ among estimators. The WM, Ohser, and Stoyan estimators cannot be distinguished, but differ quite a bit from the Ripley estimator (Figure 3.11b). This is because the weights for the Ripley edge correction always yield \( w_{ij}^R = 1 \) (since all the rings around the points of the pattern are fully within \( W \)), but the Ohser and Stoyan weights are always greater than 1. When calculating the neighborhood density \( O(r) \) we find that the WM estimator (Equation 3.17) and the Ohser and Stoyan estimators (based on Equations 3.21 and the adapted intensity estimate Equation 3.22) agree with the expected neighborhood density, but the estimator based on Equations 3.21 and 3.22, for the Ripley edge correction method, underestimate the neighborhood density at greater distances (Figure 3.11c).

This result illustrates quite well the problem of Ripley edge correction. In particular, if more points (than expected) are located close to the border, Ripley edge correction will overestimate the neighborhood density. In contrast, if fewer points are, by accident, located close to the border, the neighborhood density will be underestimated. The Ohser and Stoyan estimators are
robust against such effects and are therefore preferable. Finally, if we calculate the pair-correlation function with the estimators given in Equation 3.24, the difference between the Ohser and Ripley methods persists, with lower estimates given by the Ripley method (Figure 3.11d). At larger distances, the WM estimator, which uses the adapted intensity estimator (3.23) only once in the calculation of the pair-correlation function, produces somewhat higher estimates than the Ohser estimator (Figure 3.11d).

The example shown in Figure 3.11, which contains an extreme case of heterogeneity, was only used to develop a better understanding of the behavior of the different estimators. We see that, even in this extreme case, differences among the four estimators of the pair-correlation function provided in Equation 3.24 are relatively small. Thus, we can expect that all four estimators should produce robust results, with only small differences among them, especially for patterns that only depart slightly from homogeneity. However, an important lesson is that the neighborhood density $O(r)$ provides, even in the case of an inhomogeneous pattern, the correct result, if the Ohser or Stoyan estimator is used. However, it produces a notable bias if the Ripley estimator is used. Clearly, in the case of a heterogeneous pattern, we will need more precise methods to factor out the effects of heterogeneity. These will be discussed in the next section.

**FIGURE 3.11**
Comparison of estimators for a heterogeneous pattern. (a) Heterogeneous pattern used for comparing estimators. (b) Product density. (c) Neighborhood density. (d) Pair-correlation function.
3.1.2.5 Inhomogeneous Pair-Correlation Function Using Reweighting

For a homogeneous pattern (i.e., one that has the same statistical properties throughout the observation window), the estimators of the product density, the neighborhood density, and the pair-correlation function have an intuitive interpretation, as they represent the (neighborhood) properties of the typical point of the pattern. However, each point may have somewhat different neighborhood properties, meaning that the pattern is not homogeneous. In this case, we can still apply the estimators for homogeneous patterns (Illian et al. 2008, p. 280), but the interpretation is more complicated, because no typical point exists. Under these circumstances we can only estimate “average” neighborhood properties. It would, therefore, be desirable to derive estimators for nonstationary patterns that would be able to factor out the effects of the heterogeneity and reveal the “pure” underlying neighborhood (second-order) properties. Indeed, this is possible for certain types of heterogeneous patterns.

Baddeley et al. (2000) generalized the product density estimator (Equation 3.21) for certain types of heterogeneous patterns using the Ripley or Stoyan edge correction method. The basic assumption of their approach was that the observed pattern $\phi_o$ is only heterogeneous with respect to the intensity function $\lambda(x)$, but that the other properties of the pattern are not dependent on location $x$ (i.e., homogeneous point interactions are “modulated” by the intensity function). A pattern like this can be conceptualized as resulting from a two-step process: Initially a homogeneous point pattern $\phi_h$ is independently thinned using a thinning function $p(x) = \lambda(x)/\lambda^*$, where $\lambda^*$ is the maximal value of the intensity function $\lambda(x)$ within the observation window. In the thinning operation, all points of the initially homogeneous pattern $\phi_h$ are visited in turn and are retained in the pattern with probability $p(x)$. This results in a heterogeneous pattern with intensity function $\lambda(x) = \lambda p(x)$. The idea of the estimator introduced by Baddeley et al. (2000) is to reconstruct the pair-correlation function of the initial homogeneous pattern $\phi_h$, based on the information provided by the points of the observed heterogeneous pattern $\phi_o$ and an estimate of the intensity function. To do this, we first look at the probability that a point of $\phi_h$ is located within an infinitesimal disk of area $dx$ centered on location $x$. This probability yields $\lambda dx$. Consequently, the probability that a point of $\phi_o$ is located within an infinitesimal disk with area $dx$ centered on location $x$ yields $\lambda p(x)dx = \lambda(x)dx$. Thus, to recover the properties of $\phi_h$, an estimator of the “inhomogeneous” pair-correlation function must weight a point at location $x$ with the inverse of $p(x)$. The reason for this is that the location contained a point with a probability elevated, on average, by a factor of $1/p(x)$ before thinning was applied. Statisticians call this “re-weighting” (Baddeley et al. 2000). The corresponding estimator of the pair-correlation function therefore yields:

$$\hat{g}_{BMW}^{BMW}(r, \hat{\lambda}(x)) = \frac{1}{2\pi rA} \sum_{i=1}^{n} \sum_{j=1}^{w_{ij}} k(\|x_i - x_j\| - r) \frac{w_{ij}(r)}{\hat{\lambda}(x_i)\hat{\lambda}(x_j)}$$  \hspace{1cm} (3.25)
The superscript BMW stands for the three authors Baddeley, Møller, and Waagepeterson of the seminal 2000 article describing this approach.

The inhomogeneous pair-correlation function reveals the pure second-order properties of the pattern and has properties analogous to a homogeneous pair-correlation function, but only if the underlying intensity function \( \lambda(x) \) has been correctly identified and fulfills certain conditions (i.e., the observed pattern \( \phi_\nu \) can indeed be obtained by thinning of a homogeneous pattern \( \phi_\nu \), and the thinning function \( p(x) \) is a smooth function that does not contain very low values). Note that for each individual point pair \( i-j \), a common weighting term \( w_i(\nu)/[\lambda(x_i)\lambda(x_j)] \) in the estimator is used for both edge correction and compensating for the heterogeneity. Therefore, it makes sense to apply this together with the Ripley and Stoyan edge correction methods, which weight point pairs individually (Equation 3.21).

Although it has some useful properties, the BMW estimator 3.25 has two problems. First, it cannot be applied for all intensity functions \( \lambda(x) \) because \( \lambda(x) > 0 \) is required for all \( x \) in \( W \). Second, the estimator is highly unstable if the value of \( p(x) \) and/or \( p(x) \) is small in some locations, because the weights can become very large (see Section 2.6.3.5). Because of these problems, estimator 3.25 cannot be applied for a wide range of situations that are however highly relevant in practical applications. We therefore present, in the following, an alternative estimator of the inhomogeneous pair-correlation function that is not subject to these restrictions. This estimator generalizes the WM and Ohser edge correction methods.

### 3.1.2.6 Alternative Estimators of Inhomogeneous Pair-Correlation Functions

A naive estimator of the mean number of points at distance \( r \) from the points of the pattern is given by

\[
\hat{Z}_n^g(r) = \frac{1}{n} \sum_{i=1}^{n} \left( \sum_{j=1, j \neq i}^{n} k(\|x_i - x_j\| - r) \right) dr
\]

and a naive estimator of the expected number of points within a ring with radius \( r \) and width \( dr \) is given by

\[
\hat{N}_n^g(r) = 2\pi r dr \lambda.
\]

Thus, a naive estimator of the pair-correlation function yields \( \hat{g}_n(r) = \hat{Z}_n^g(r)/\hat{N}_n^g(r) \). The bias due to edge effects can be corrected in the numerator or the denominator. The Ripley and Stoyan edge correction methods correct the numerator, leaving the denominator proportional to \( 2\pi r \lambda \).
Consequently, the corresponding estimator of the inhomogeneous pair-correlation function yields the BMW estimator (Equation 3.25), which is based on individual reweighting of point pairs.

Instead of altering the numerator (Equation 3.26), estimation of the inhomogeneous pair-correlation function using the Ohser and WM methods of edge correction modifies the denominator \( \hat{N}_g \), which is the expected number of points in rings with radius \( r \) and width \( dr \). For homogeneous patterns, the expectation is given by the homogeneous Poisson process. However, for heterogeneous patterns, where \( \varphi_h \) results from independent thinning of a homogeneous pattern \( \varphi_h \), the expectation is given by a heterogeneous Poisson process based on the intensity function \( \lambda(x) \) used in thinning. In contrast to the BMW estimator, no further assumptions need to be made on the intensity function \( \lambda(x) \).

Modification of the denominator can be accomplished in two ways, depending on whether or not we use an adapted estimator. The nonadapted estimator of the expected number of points based on the Ohser edge correction method yields:

\[
\hat{N}_g (r) = \frac{1}{n} \int_W \lambda(x) \left[ \int_W \lambda(a) k(\|x - a\| - r) da \right] dx
\]

To understand this, consider that under the heterogeneous Poisson process, the probability that a point is located within a small disk with area \( da \) centered at location \( a \) yields \( \lambda(a) da \). Thus, to determine the number of points \( w(x, r) \) expected to be located within \( W \) in a ring of radius \( r \) centered on \( x \), we need to integrate \( \lambda(a) da \) over all locations \( a \) in \( W \) that are located within the ring, that is, \( w(x, r) = \int_W \lambda(a) k(\|x - a\| - r) da \). Note that the kernel function \( k() \) yields a value of 1 if a point is located within the ring and 0 otherwise. Finally, we need to average over all potential point locations \( x \). This is done by multiplying \( w(x, r) \) by \( \lambda(x) dx \), the probability that there is a point within disk \( dx \) centered at \( x \), and integrate over all locations in \( W \) to end up with Equation 3.28. The corresponding estimator of the pair-correlation function thus yields

\[
\hat{\gamma}_{inhom}^{O}(r, \lambda(x)) = \frac{(1/n) \sum_{i=1}^{n} \left( \sum_{j=1}^{n} k(\|x_i - x_j\| - r) \right)}{(1/n) \int_W \lambda(x) \left[ \int_W \lambda(a) k(\|x - a\| - r) da \right] dx}

= \frac{1}{\lambda^2} \frac{1}{2\pi r A} \sum_{i=1}^{n} \left( \sum_{j=1}^{n} k(\|x_i - x_j\| - r) \right) \frac{A}{\gamma_W(r, \lambda(x))}
\]
where \( \hat{\lambda}_n = n/A \) is the natural estimator of the intensity. The denominator contains a generalized, isotropized set covariance

\[
\bar{N}_{W}(r, \lambda(x)) = \frac{1}{2\pi r} \frac{1}{\hat{\lambda}_n^2} \int_W \hat{\lambda}(x) \left[ \int_W \lambda(a) k(\|x - a - r\|) da \right] dx
\]

(3.30)

which needs to be estimated numerically.

The analogous adapted estimator that modifies the estimator \( \hat{N}_n^{\text{inhom}}(r) \) in Equation 3.27 does not average over all focal points \( x \) within \( W \), but only over the points \( x_i \) of the pattern

\[
\hat{N}_{W}^{\text{inhom}}(r, \lambda(x))
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \left( \sum_{j=1}^{n,x} k(\|x_i - x_j\| - r) \right)
\]

contains intensity function \( \lambda(x) \)

\[
= \frac{1}{n-1} \sum_{i=1}^{n} \left( \sum_{j=1}^{n,x} k(\|x_i - x_j\| - r) \right)
\]

approximates \( \lambda(x) \) by points \( a_i \) of an auxiliary pattern

(3.31)

Note that this estimator does not need to weight the points \( x_i \) with the intensity function \( \lambda(x_i) \) as required in the estimator using the generalized Ohser edge correction (i.e., 3.29 and 3.30). This is because the points \( x_i \) of the pattern are assumed to be based on the intensity function \( \lambda(x) \). This becomes clear when comparing the denominators of Equations 3.29 and 3.31, where the integration \( \left(1/n \right) \int_W \hat{\lambda}(x) \) over the entire observation window \( W \) is replaced in Equation 3.31 by the sum \( (1/n) \sum_{i=1}^{n} \) over the points \( x_i \) of the observed pattern in Equation 3.29.

In practice, the denominator of Equation 3.31 can be approximated numerically by generating \( m \) points \( a_i \) based on an auxiliary pattern that follows a heterogeneous Poisson process with intensity function \( \lambda(x) \). If \( m \) is large (e.g., \( m = 20,000 \)) it will approximate the continuous integral, but may somewhat slow down the computation. Because estimator 3.31 is composed of the ratio of two quantities that are estimated in exactly the same way around the points \( x_i \) of the pattern (i.e., it is an adapted estimator) we expect that effects of small irregularities in the pattern will cancel out.

Estimators 3.29 and 3.31 of the inhomogeneous pair-correlation function have three principal advantages over the BMW estimator 3.25. First, they allow estimation of the inhomogeneous pair-correlation function for all intensity functions \( \lambda(x) \), including those with zero values at some locations \( x \).
Second, because they do not weight each point pair $i-j$ individually, they do not produce very large weights if the intensity function contains very low values. Thus, it can be expected that these estimators yield robust estimates of the pair-correlation function, even for rugged intensity functions that include zero values. These properties open up many possibilities for practical applications. Third, estimators 3.29 and 3.31 reconcile edge correction and heterogeneity and allow for a simple approach to edge correction for irregular observation windows (see Section 3.4.2.2). For example, the pair-correlation function of a homogeneous pattern within an irregular observation window can be estimated with Equations 3.29 and 3.31, if the intensity function is defined as being zero outside the observation window and $\lambda$ inside. The generalized isotropized set covariance in Equation 3.30 will then collapse to the “standard” isotropized set covariance for the irregularly shaped observation window $W$. However, these estimators have the disadvantage that they must be approximated numerically.

A simple example illustrating how the inhomogeneous pair-correlation function works is given in Figure 3.12. The pattern arises from a random

![Figure 3.12](image)

**FIGURE 3.12**
Application of inhomogeneous pair-correlation functions to a pattern with a gradient heterogeneity. (a) Original homogeneous random pattern $\Phi_h$ with 961 points within a $100 \times 100$ m$^2$ observation window. (b) Thinning function $p(x)$ used to generate the heterogeneous pattern $\Phi$, shown in panel (c), darkness of shading is proportional to intensity. (c) Thinned pattern. (d) Analysis of the inhomogeneous pattern $\Phi_o$ in panel (c). The expected homogeneous pair-correlation function for the original pattern in panel (a) is given as a gray line.
pattern \( \phi_h \) (Figure 3.12a) that was independently thinned with a gradient intensity function \( \lambda(x) \) (Figure 3.12b), yielding an inhomogeneous pattern \( \phi_o \) (Figure 3.12c). Figure 3.12d shows the homogeneous pair-correlation function of the original homogeneous pattern \( \phi_h \) (gray line), and the inhomogeneous pair-correlation function estimated with the WM method (Equation 3.31) from the inhomogeneous pattern \( \phi_o \) and the intensity function \( \lambda(x) \) (dots). The two estimates are in excellent agreement. Figure 3.12d also shows that the corresponding inhomogeneous pair-correlation function estimated with the BMW method yields essentially the same results. However, it also includes some erratic fluctuations, which are caused by the two points located in the bottom left at locations with low values of the intensity function (Figure 3.12c).

The inhomogeneous pair-correlation function also works very well for more complex patterns produced by complex intensity functions, as we show in Figure 3.13. This provides an example of a homogeneous pattern generated by a double-cluster Thomas process, where 468 small clusters with an approximate radius of 8.8 m and an average of 27 points were nested within 183 large clusters with approximate radii of 36 m (Figure 3.13a). This rather complex pattern was then independently thinned with the thinning function shown in Figure 3.13b. Most areas in the pattern have very low or zero values (light gray) for the intensity. However, there are also high values of the intensity function in a continuous band extending from the lower part of the plot toward the eastern border (dark gray). Independent thinning with this intensity function removed almost 90% of the points of the original pattern yielding the pattern shown in Figure 3.13c. As a result, most of the very dense small clusters were removed and only the pattern at the lower middle of the plot was retained. Thus, much of the structure of the original homogeneous pattern was destroyed and the question is whether or not the inhomogeneous pair-correlation function is able to recover the original second-order structure.

Indeed, Figure 3.13d demonstrates that the inhomogeneous pair-correlation function estimated with the WM method recovers the pair-correlation function of the original homogeneous pattern very well (cf. dark gray line with dotted line). Figure 3.13d shows that the corresponding inhomogeneous pair-correlation function estimated with the BMW method yields essentially the same results, but shows some erratic fluctuations, which are caused by points located at locations with low values of the intensity function. Thus, the two examples of Figures 3.12 and 3.13 show that the inhomogeneous pair-correlation function is able to recover the “pure” second-order structure of an inhomogeneous pattern that resulted from thinning of a homogeneous pattern, if the underlying intensity function is known. The examples also suggest that the BMW estimation method is, as expected, somewhat sensitive to intensity functions with near-zero values, whereas this problem does not arise with the WM estimation method.
Application of inhomogeneous pair-correlation functions to a pattern with complex heterogeneity. (a) Original homogeneous pattern $\varphi_h$ consisting of 11,519 points within a $1000 \times 500$ m$^2$ observation window. The pattern was generated by a double-cluster Thomas process containing 468 small clusters, with an approximate radius of 8.8 m (and an average of 27 points), nested within 183 large clusters with approximate radii of 36 m. (b) Thinning function $p(x)$ used to generate the pattern shown in panel (c). The gray scale is proportional to intensity, ranging from black, with $p(x) = 1$, to white, with $p(x) = 0$. (c) Pattern resulting from thinning. (d) Analysis of the pattern in panel (c). The expected homogeneous pair-correlation function for the original pattern in panel (a) is given as a gray line.
3.1.2.7 Ripley’s K-Function

The estimators for the K-function are fully analogous to those of the pair-correlation function. As a consequence, the same principles for edge correction and generalization for inhomogeneous patterns can be applied. For homogeneous patterns, the quantity $\lambda K(r)$ can be interpreted as the number of points lying within distance $r$ of the typical point of the pattern. Thus, $\lambda K(r)$ counts the number of further points that are expected within a neighborhood $r$ of a given point. Following Illian et al. (2008), the estimators of the K-function are based on the quantity $\kappa = \lambda^2 K(r)$ which is analogous to the product density $\rho(r) = \lambda^2 g(r)$. We estimate the quantity $\kappa$, with an estimator analogous to Equation 3.19, based on pairs of points lying within distances $d$ that are less than $r$:

$$
\hat{\kappa}(r) = \frac{1}{A} \sum_{i=1}^{n} \sum_{j=1}^{n} 1(\|x_i - x_j\|, r) \times w_{i,j}
$$

where $\lambda$ is the mean number of points within distance $r$ of points $i$.

The indicator function $1(d, r)$ plays the role of the kernel function used in estimating the product density and has a value of 1 if point $j$ is located within distance $r$ or less of point $i$ and zero otherwise. By slight rearrangement of the terms, we see that indeed $\kappa = \lambda_n [\lambda K(r)]$ (Equation 3.32).

The edge correction terms $w_{ij}$ in Equation 3.32 are the same as in case of the product density (i.e., Equation 3.9, 3.21), although the Ohser edge correction term $w_{ij}^O = A/\pi W(r)$ now depends on the distance $d$ between the points $i$ and $j$, because the isotropized set covariance depends on distance $r$. As a consequence, it cannot be factored out in the corresponding estimator. To yield an edge correction weight $w_{ij}$ that is independent of the locations of the two points $i$ and $j$, and only depends on their distance $d = \|x_i - x_j\|$, Wiegand and Moloney (2004) used the edge correction weight

$$
w_{ij}^{WM} (\|x_i - x_j\|, r) = \frac{\pi r^2}{(1/A)2\pi \int_0^r t\pi W(t)dt} \text{ if } \|x_i - x_j\| \leq r.
$$

This weighting factor is composed of the area of the full circle (numerator) divided by the average area (within $W$) of disks with radius $r$ located in $W$. 

© 2014 Taylor & Francis Group, LLC
Estimators and Toolbox

The WM weight should provide a good approximation of the Ohser weights for shorter distances of \( r \). Ward and Ferrandino (1999) derived basically the same edge correction term as Equation 3.33. When inserting the formula for \( \gamma_W(t) \) given in Equation 3.33 into Equation 3.33 we can derive an expression analogous to Equation 3.7 by integration of

\[
\int_0^r \gamma_W(t)2\pi dt = \left( A - \frac{4r}{3\pi} (a_x + b_y) + \frac{r^2}{2\pi} \right) (\pi r^2).
\]

Given the preceding considerations, the four estimators of the univariate quantity \( \hat{\kappa}(r) \) yield

\[
\begin{align*}
\hat{\kappa}^R(r) &= \frac{1}{A} \sum_{i=1}^{n} \sum_{j=1}^{n \neq i} \mathbf{1}(\|x_i - x_j\|, r) \frac{2\pi}{\varphi_{ij}} & \text{Ripley} \\
\hat{\kappa}^S(r) &= \frac{1}{A} \sum_{i=1}^{n} \sum_{j=1}^{n \neq i} \mathbf{1}(\|x_i - x_j\|, r) \frac{A}{\gamma_W(x_i - x_j)} & \text{Stoyan} \\
\hat{\kappa}^O(r) &= \frac{1}{A} \sum_{i=1}^{n} \sum_{j=1}^{n \neq i} \mathbf{1}(\|x_i - x_j\|, r) \frac{A}{\gamma_W(\|x_i - x_j\|)} & \text{Ohser} \\
\hat{\kappa}^{WM}(r) &= \frac{1}{A} \sum_{i=1}^{n} \sum_{j=1}^{n \neq i} \mathbf{1}(\|x_i - x_j\|, r) \frac{A\pi r^2}{2\pi \int_0^r t\gamma_W(t)dt} & \text{WM}
\end{align*}
\]

Because the expectation of \( \hat{\kappa}(r) \) increases at a rate of \( \pi r^2 \), we compare the different estimators based on the quantity \( \hat{\kappa}(r)/\pi r^2 \). Figure 3.14a shows the values of the four estimators of \( \hat{\kappa}(r)/\pi r^2 \) for a random pattern with 1000 points within a 100 \( \times \) 100 m\(^2\) observation area (Figure 3.14a). In contrast, Figure 3.14b shows the estimators for a clustered pattern shown in Figure 3.10b. The WM, Ohser, and Stoyan estimators yield virtually identical results, whereas the Ripley estimator yields slightly lower values for the random pattern and somewhat larger values for the clustered pattern (Figure 3.14a,b). This result parallels the results for the product density function (e.g., Figure 3.11b).

Analogous to the estimation of the pair-correlation function, where the adapted intensity estimator \( \hat{\lambda}\_S(r) \) was used (Equations 3.22 and 3.24), one can also define \( \hat{\lambda}\_V(r) \), an estimator of the intensity that is adapted to the \( K \)-function (Illian et al. 2008, p. 194). While the adapted intensity estimator \( \hat{\lambda}\_S(r) \) used for the pair-correlation function considered the length of the
circumference of the circles of radius $r$ around the points $i$ of the pattern that lay inside the observation window $W$, the intensity estimator adapted for the $K$-function considers the areas of disks with radius $r$ around the points $i$ of the pattern that lie inside $W$. As a consequence, there are two small modifications required in the intensity estimator:

1. The length $v_d(W \cap \partial b(x_i, r))$ of the circumference of a circle of radius $r$ around point $i$ in $W$ has to be replaced by the corresponding area $v(W \cap b(x_i, r))$ of a circle of radius $r$ around point $i$ which overlaps $W$.

2. The term $\lambda(r) \lambda(r)$ has to be replaced by the integral $\int_0^r t \bar{\gamma}_W(t)dt$ to yield the expected area of a disk with radius $r$ within $W$:

$$
\hat{\lambda}_S(r) = \frac{\sum_i^n v_d(W \cap \partial b(x_i, r))}{2\pi \bar{\gamma}_W(r)}
$$

$$
\hat{\lambda}_V(r) = \frac{\sum_i^n v(W \cap b(x_i, r))}{2\pi \int_0^r t \bar{\gamma}_W(t)dt} = \hat{\lambda}_W \left[ \frac{1/n}{2\pi} \sum_i^n v(W \cap b(x_i, r)) \int_0^r t \bar{\gamma}_W(t)dt/A \right]
$$

(3.35)

**FIGURE 3.14**
Estimators of the $K$-function. (a) Estimators of $\lambda^2 K(r)$ for a random pattern of 1000 points within a $100 \times 100$ m$^2$ area. (b) Estimators of $\lambda^2 K(r)$ for the clustered pattern in Figure 3.10b, consisting of 626 points within a $100 \times 100$ m$^2$ area. (c) Estimates of the $L$-function for a random pattern as in panel (a). (d) Estimates for the $L$-function of the clustered pattern in Figure 3.10b. Panels (a) and (b) show values for $\lambda \lambda_i$ and $\lambda \lambda_n$ in addition to $K$-function estimates.
When factoring out $\hat{\lambda}_n = n/A$ in Equation 3.35, we find that $\hat{\lambda}_V(r)$, the estimator of lambda, is basically given by the naive estimator of the intensity multiplied by a correction factor (i.e., the expression within parentheses in the numerator of Equation 3.35). The correction term adjusting the value of the naive estimator is composed of a numerator, which is the average area of disks $b(x_i, r)$ overlapping $W$ calculated from the points of the pattern, and denominator, which is the expected area overlapping $W$ of disks centered in $W$. This is analogous to the case of the intensity estimator adapted to the pair-correlation function (Equation 3.23).

Continuing the example shown in Figure 3.14, we calculated the squared adapted intensity estimate $\hat{\kappa}_V^2(r)$, which is shown as dashed dark gray lines in Figure 3.14a and b. These figures confirm that the correction factor indeed captures some of the small bias in $\hat{\kappa}(r)/\pi r^2$ as also shown for the product density (Figure 3.10). The $\hat{\lambda}_V(r)$ will be smaller/larger than $\hat{\lambda}_n = n/A$ if many/few points are accidently located close to the border.

Wiegand and Moloney (2004) proposed an estimator of $\hat{\lambda}K(r)$ based on a grid approximation to facilitate calculation of the $K$-function for irregularly shaped study areas. The quantity $\hat{\lambda}K(r)$ yields the expected number of points of the pattern within distance $r$ of the typical point of the pattern. The quantity $\hat{\lambda}K(r)$ is thus a conditional summary statistic, similar to $O(r) = \lambda g(r)$; that is, it yields the expected number of points within a disk, given that there is a point at the center of the disk. With this definition, the WM estimator first counts, for each point $i$ of the pattern, the number of points $j$ located within distance $r$ of $i$ that lie inside $W$ (i.e., $\sum_{j=1}^{n_{i,r}} 1(\|x_i - x_j\|, r)$) and then averages the number of points meeting this criterion for all of the points $i$ of the pattern: $(1/n)\sum_{i=1}^{n} (\sum_{j=1}^{n_{i,r}} 1(\|x_i - x_j\|, r))$. The estimate of $\hat{\lambda}K(r)$ is then obtained by dividing the average number of points in disks with radius $r$ by the average area of the disks that lie inside the observation window $W$ (i.e., $(1/n)\sum_{i=1}^{n} v(W \cap b(x_i, r))$:}

$$\hat{\lambda}K_{WM}^2(r) = \pi r^2 \frac{(1/n)\sum_{i=1}^{n} (\sum_{j=1}^{n_{i,r}} 1(\|x_i - x_j\|, r))}{(1/n)\sum_{i=1}^{n} v(W \cap b(x_i, r))} = \frac{\hat{\kappa}_{WM}^2(r)}{\hat{\lambda}_V(r)} \quad (3.36)$$

It can be easily verified that dividing $\hat{\kappa}_{WM}^2(r)$ of Equation 3.34 by the adapted intensity estimator $\hat{\lambda}_V(r)$ of Equation 3.35 yields the estimator of $\hat{\lambda}K(r)$ given in Equation 3.36.

Following the suggestion by Stoyan (2006a) and Illian et al. (2008), we obtain three alternative estimators for the $K$-function, based on the different edge correction methods (i.e., the Ripley, Stoyan, and Ohser weights shown in Equation 3.34). These divide $\hat{\kappa}(r)$ by the square of the adapted intensity estimator (Equation 3.35):
\[
\hat{K}^{WM}(r) = \frac{\hat{\kappa}^{WM}(r)}{\hat{\lambda}(r)\hat{\lambda}_m} \quad \text{WM}
\]
\[
\hat{K}^{O}(r) = \frac{\hat{\kappa}^{O}(r)}{\hat{\lambda}(r)^2} = \frac{\hat{\kappa}^{WM}(r)}{\hat{\lambda}(r)^2} \quad \text{Ohser}
\]
\[
\hat{K}^{S}(r) = \frac{\hat{\kappa}^{S}(r)}{\hat{\lambda}(r)^2} \quad \text{Stoyan}
\]
\[
\hat{K}^{R}(r) = \frac{\hat{\kappa}^{R}(r)}{\hat{\lambda}(r)^2} \quad \text{Ripley}
\]

For completeness, we also show the estimator \(\hat{K}^{WM}(r)\) presented in Wiegand and Moloney (2004), which divides the estimator of \(\lambda K(r)\) by the natural estimator of the intensity (i.e., \(n/A\)).

Continuing our example shown in Figure 3.14, we calculated the four \(K\)-functions for the two patterns and show the corresponding \(L\)-functions, where \(L_1(r) = \sqrt{\pi^{-1}K(r)} - r\). The resulting differences among the estimators for the \(L\)-function are relatively small. In general, the \(L\)-function estimated with the Ohser and the Stoyan estimators (which are almost identical) falls somewhere in-between the estimates obtained with the Ripley and the Wiegand–Moloney methods (Figure 3.14c,d). For clustered patterns, with a large cluster in one corner of the observation window (Figure 3.10b), the Ripley method overestimates the clustering (Figure 3.14d) and the Wiegand–Moloney method tends to indicate somewhat reduced clustering.

The approach to developing estimators for the inhomogeneous \(K\)-function is fully analogous to the approach used for the pair-correlation function. The corresponding BMW estimator of the inhomogeneous pair-correlation function corresponding to Equation 3.25, therefore, yields

\[
\hat{K}^{BMW}(r, \lambda(x)) = \frac{1}{A} \sum_{i=1}^{n} \sum_{j=1}^{n, j \neq i} \mathbf{1}(\|x_i - x_j\|, r) \frac{w_{ij}(r)}{\lambda(x_i)\lambda(x_j)}.
\]

For each individual point pair \(i-j\) lying closer than distance \(r\) apart, the estimator applies the common weighting term \(w_{ij}(r)/[\lambda(x_i)\lambda(x_j)]\) for edge correction and adjustment for the heterogeneity. The Ripley and Stoyan edge corrections methods are therefore applied, since they weight point point pairs individually (Equation 3.9).

The alternative WM method for estimation of the inhomogeneous \(K\)-function is based on the ratio of the mean number of points \(\bar{N}^{KW}_{WM}(r)\) within distance \(r\) of the points of the pattern, relative to the corresponding expected number of points \(\bar{N}^{K}_{WM}(r)\):
\[ \hat{K}_{\text{WM}}(r) = \pi r^2 \frac{\hat{Z}_{\text{WM}}(r)}{N_{\text{WM}}(r)} \]  

(3.39)

The mean number of points within distance \( r \) of the points of the pattern yields, as before,  
\[ \hat{Z}_{\text{WM}}(r) = \frac{1}{n} (\sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{1}(\|x_i - x_j\|, r)) \], cf. Equation 3.36. It then follows that the expectation under the heterogeneous Poisson process yields an equation analogous to Equation 3.28 for the nonadapted case:

\[ \hat{N}_{\text{O}}^K = \frac{1}{n} \int_{W} \lambda(x) \left[ \int_{W} \lambda(a) \mathbf{1}(\|x - a\|, r) da \right] dx \]  

(3.40)

The corresponding estimator of the \( K \)-function becomes

\[ K_{\text{inhom}}^O(r, \lambda(x)) = \pi r^2 \frac{\frac{1}{n} \sum_{i=1}^{n} \left( \sum_{j=1}^{n} \mathbf{1}(\|x_i - x_j\|, r) \right)}{\frac{1}{n} \int_{W} \lambda(x) \left[ \int_{W} \lambda(a) \mathbf{1}(\|x - a\|, r) da \right] dx} \]  

(3.41)

The analogous adapted estimator that modifies the estimator \( \hat{N}_n(r) \) does not average over all focal points \( x \) within \( W \), but only over the points \( x_i \) of the pattern:

\[ \hat{K}_{\text{inhom}}^W(r, \lambda(x)) = \pi r^2 \frac{\sum_{i=1}^{n} \left( \sum_{j=1}^{n} \mathbf{1}(\|x_i - x_j\|, r) \right)}{\sum_{i=1}^{n} \int_{W} \lambda(a) \mathbf{1}(\|x_i - a\|, r) da} \]  

contains intensity function \( \lambda(x) \)

\[ = \pi r^2 \frac{\sum_{i=1}^{n} \left( \sum_{j=1}^{n} \mathbf{1}(\|x_i - x_j\|, r) \right)}{\sum_{i=1}^{n} \left( \sum_{j=1}^{n} \mathbf{1}(\|x_i - a\|, r) \right)} \]  

approximates \( \lambda(x) \) by points \( a_i \) of an auxiliary pattern

(3.42)
3.1.3 Nearest-Neighbor Summary Statistics for Univariate Patterns (Data Type 1)

3.1.3.1 Distribution Functions for Nearest-Neighbor Distances

Summary statistics based on distances to the kth neighbor provide an alternative way of characterizing the small-scale correlation structure of point patterns and complement the use of the second-order summary statistics. For homogeneous patterns, the function \( D_k(r) \) represents the (cumulative) distribution of the distances \( r \) to the kth nearest neighbor as measured from the typical point of the pattern. \( D(r) \) is used to refer to the distribution function of distances to the nearest neighbor [i.e., \( k = 1 \), \( D(r) = D^1(r) \)]. Nearest-neighbor statistics are “short-sighted” and sense only the immediate neighborhood of the typical point. They are, therefore, particularly sensitive to local cluster structures and can differentiate among dissimilar cluster processes that have identical second-order summary statistics (e.g., Wiegand et al. 2007c, 2009).

In estimating \( D_k(r) \), we first need to determine the distance \( d_{ik} \) to the kth neighbor for each point \( i \) in \( W \). The naive estimator of \( D_k(r) \) without edge correction is given by

\[
\hat{D}_k(r) = \frac{1}{n} \sum_{i=1}^{n} 1(0 < d_{ik} \leq r)
\]

(3.43)

where \( 1(0 < d_{ik} \leq r) \) is an indicator function, which yields a value of 1, if the kth neighbor is within distance \( r \) (i.e., \( 0 < d_{ik} < r \)), and zero otherwise. However, the “real” \( d_{ik} \) for some of the points may be less than the estimated \( d_{ik} \), since only information on points within \( W \) is available. This error occurs when an unobserved point lying outside \( W \) is the true kth neighbor for a focal point, that is, \( d_{ik} \) is greater than the distance \( \beta_i \) from point \( i \) to the nearest border of \( W \) (Figure 3.15a). This edge effect produces a small bias in the estimation of \( D_k(r) \).

The buffer method, which was introduced by Ripley (1977), is able to avoid this bias. It uses only focal points \( i \) located within an “inner window” surrounded by a buffer of width \( b \) in calculating \( D_k(r) \) (Figure 3.15b). If we restrict the calculation of \( D_k(r) \) to distances \( r \leq b \), \( \hat{D}_k(r) \) can be correctly estimated based on the points of the pattern within \( W \). The estimator that uses the buffer method is similar to the estimator without edge correction (Equation 3.43), but only includes focal points \( i \) that are located within the inner window:

\[
\hat{D}^{k,b}(r) = \frac{\sum_{i=1}^{n} 1(0 < d_{ik} \leq r) \times 1(b < \beta_i)}{\sum_{i=1}^{n} 1(b < \beta_i)}, \quad r \leq b
\]

(3.44)

The indicator function \( 1(b < \beta_i) \) yields a value of 1 if \( b < \beta_i \) and zero otherwise.

Although Equation 3.44 removes the bias in the estimate of \( D_k(r) \), it has the disadvantage of only producing reliable results for a small range of distances...
For larger values of the buffer \( b \), the inner window becomes small compared to \( W \) and almost no focal points remain to be analyzed. An additional disadvantage of this estimator is that it disregards valuable information for points in the buffer for which the distance to the \( k \)th neighbor can be correctly evaluated. This can be seen in Figure 3.15b where the nearest neighbor for point \( c \) lies within \( W \), that is, point \( e \) is closer to \( c \) than \( c \) is to the border.

Some of the issues related to the use of the buffer method can be illustrated by exploring some examples. Figure 3.16a,b show two patterns together with a buffer area of 100 m. The points lying outside the inner rectangle are not used as foci for estimation of \( D^k(r) \), if we apply the buffer method. As a consequence, the buffer method excludes the information for many points that are located within the buffer area, but for which the \( k \)th nearest neighbor can be, in fact, correctly estimated. For example, the open circles are those points for which the border is closer than the first nearest neighbor, which means that the nearest neighbor (i.e., \( k = 1 \)) cannot be accurately determined. It can be seen that there are only a few points for which this is the case, so most points can actually be included in the analysis without producing a bias. However, as the rank \( k \) of the neighbor increases, fewer focal points can be accurately assessed. Even so, the buffer method excludes a number of points that can be included in the analysis, since the \( k \)th nearest neighbor in \( W \) lies closer to the focal point than the distance between the focal point and the border. For example, the closed black disks in Figure 3.16a,b represent points that can be included in an analysis for \( k = 25 \). In essence, we may lose a substantial amount of information if we do not consider the points within the buffer area for which the distance to the \( k \)th neighbor can be correctly determined.
Hanisch (1984) proposed an estimator of $D_k(r)$ that uses more information than allowed by Equation 3.44 and overcomes the disadvantages of the buffer estimator. It is similar to the buffer estimator, but instead of the buffer being fixed for all points, it is individually adjusted for each focal point $i$. In this way, the Hanisch estimator applies edge correction by using only focal points $i$ for which the $k$th neighbor is definitely located within $W$. The $k$th neighbor of point $i$ can be assessed if the distance $d_{ik}$ to its $k$th neighbor is shorter than the distance $\beta_i$ to the nearest border of $W$. For points not meeting this condition, the $k$th neighbor may potentially be located outside $W$ and the point is therefore not considered in the analysis. Given these considerations, the Hanisch estimator can be defined as follows:

$$
\hat{D}^{k,H}(r) = \frac{\sum_{i=1}^{n} I(0 < d_{ik} \leq r) \times I(d_{ik} < \beta_i) \times w(d_{ik})}{\sum_{i=1}^{n} I(d_{ik} < \beta_i) \times w(d_{ik})}
$$

(3.45)

FIGURE 3.16
Comparison of estimators for the nearest-neighbor distribution function $D_k(r)$. (a) Random pattern with 1296 points within a $1000 \times 500$ m$^2$ observation window. (b) Clustered pattern with 626 points within a $500 \times 500$ m$^2$ observation window. (c) Nearest-neighbor distribution functions $D_k(r)$ for the pattern in panel (a). (d) Nearest-neighbor distribution functions $D_k(r)$ for the pattern in panel (b). The rectangles in panels (a) and (b) enclose the area of focal points used for the buffer method; open points are those excluded by the Hanisch estimator for $k = 1$, and gray points were those additionally excluded for $k = 25$. In panels (c) and (d), Hanisch estimators are given (symbols), and estimators without edge correction are shown as solid gray lines which often overlaps with the estimates of the Hanisch estimator.
The indicator function \( 1(d_{ik} < \beta_i) \) tests for the condition that the distance \( d_{ik} \) to the \( k \)th nearest neighbor is shorter than the distance \( \beta_i \) to the nearest border and yields a value of 1 if \( d_{ik} < \beta_i \) and zero otherwise. A weighting factor \( w(d_{ik}) \) is included in Equation 3.45 to compensate for excluded points. This can be defined by noting that the proportion of points in \( W \) located further from the border than distance \( d_{ik} \) is given for homogeneous patterns by \( (l_x - d_{ik})(l_y - d_{ik})/l_x l_y \) for \( W \) defined by a rectangle with sides of length \( l_x \) and \( l_y \). Thus, for \( r < \min(l_x/2, l_y/2) \), an appropriate weighting factor \( w(d_{ik}) \) that compensates for the points for which \( d_{ik} > \beta_i \) is given by

\[
w(d_{ik}) = \frac{1}{(l_x - d_{ik})(l_y - d_{ik})} \quad (3.46)
\]

However, for large values of \( r \) the Hanisch estimator becomes unreliable because few points will exist that are farther than distance \( r \) from the nearest border. As an upper limit, \( r \) is certainly constrained to be less than the maximal distance inside \( W \) from the nearest border. For a rectangle, this is half of the length of its smaller side.

Equation 3.47 outlines the similarity between the buffer method and the Hanisch method by presenting a slightly rewritten version of the buffer method (Equation 3.44):

\[
\hat{D}^{k,h}(r) = \frac{\sum_{i=1}^{n} 1(0 < d_{ik} \leq r) \times 1(d_{ik} < \beta_i) \times w(d_{ik})}{\sum_{i=1}^{n} 1(d_{ik} < \beta_i) \times w(d_{ik})} 
\]

\[
\hat{D}^{k,b}(r) = \frac{\sum_{i=1}^{n} 1(0 < d_{ik} \leq r) \times 1(b < \beta_i) \times w(b)}{\sum_{i=1}^{n} 1(b < \beta_i) \times w(b)}, \quad (3.47)
\]

Basically, the distance \( d_{ik} \) to the \( k \)th neighbor has to be replaced by the constant buffer width \( b \). Because the weight \( w(b) \) of the buffer method is the same for all focal points \( i \) it cancels out in Equation 3.47, thus yielding Equation 3.44. In contrast, the Hanisch method determines for each individual point \( i \) whether or not it is included in the calculations. Note that the estimators given in Equation 3.47 are adapted estimators that use a rationale similar to that used for the adapted estimators of the second-order statistics presented in the previous sections. The numerator is an estimator of \( \lambda D^k(r) \) and the denominator is an adapted estimator of the intensity \( \lambda \) (Stoyan 2006b).

To obtain an intuitive appreciation of the Hanisch method of edge correction, we show the points of the pattern used by the Hanisch method as focal points in estimating \( D^k(r) \)’s for a few chosen values of \( k \) and \( r \) (Figure 3.16a,b). For the nearest neighbor (i.e., \( k = 1 \)), only a few points very close to
the border are excluded from the calculation (open points in Figure 3.16a,b). In this case, the estimates using the Hanisch method versus not applying edge correction should be very close. However, for higher neighbor ranks \(k = 25\); Figure 3.16a,b) substantially more points are excluded as focal points by the Hanisch method in estimating \(D^k(r)\) (gray points). All of the excluded points are suspected to have their 25th nearest neighbor lying outside of \(W\). It is also evident from Figure 3.16 that the buffer method excludes substantially more points as compared to the Hanisch method.

Figure 3.16 also provides a comparison of the estimator with no edge correction and that of the Hanisch method. The estimates of the buffer and the Hanisch method coincide for the random pattern (not shown). For small neighborhood ranks of say \(k \leq 8\), the estimates without edge correction are also very close to those with edge correction (Figure 3.16c). However, for larger neighborhood ranks we find differences between the method without edge correction (gray lines) and the two methods that use edge correction (Figure 3.16c). While the estimates for \(D^k(r)\) still agree for smaller values of \(D^k(r)\), the farthest distances to the \(k\)th neighbor [i.e., the approach of \(D^k(r)\) to a value of 1] are overestimated when no edge correction is applied. This means that the proportion of points having a nearby \(k\)th neighbor (i.e., clustered points) is usually well estimated by the estimator that does not use edge correction, but the proportion of points having their \(k\)th neighbor farther away (i.e., more isolated points) shows a bias. This is expected because the estimator without edge correction may in many cases overestimate the distances to the \(k\)th neighbor for points close to the border. This effect will be more pronounced for more isolated points (i.e., points with larger distances to the \(k\)th neighbor) and for larger neighborhood ranks \(k\). Keeping these results in mind is important for correctly interpreting results of \(D^k(r)\) when edge correction is not applied.

Figure 3.16d shows the results of the two methods of edge correction applied to nearest-neighbor analysis for a clustered pattern (Figure 3.16b). Interestingly, differences among the three estimators are smaller than for the random pattern. For all distances \(r\), the estimates obtained agree in close approximation. Because the pattern is highly clustered, the distribution functions to even the 25th nearest neighbor agree in all three estimators. In this case, the summary statistics \(D^k(r)\) mostly measure the properties of the cluster of points surrounding the focal points or the points of the nearest cluster and problems with isolated points observed for the random pattern do not arise. Summarizing the two examples, a good approximation of \(D^k(r)\) will arise even if edge correction in the estimation of the distribution function to the \(k\)th neighbor is not employed, unless the pattern comprises a substantial proportion of isolated points. In the latter case, bias is introduced if the neighborhood rank \(k\) is high, but only as \(D^k(r)\) approaches a value of 1. However, it is also clear that we may expect a bias, especially for patterns with few points. More technical details and comparisons among estimators of the nearest neighbor distance distribution function can be found in Stoyan (2006b).
3.1.3.2 Mean Distance to kth Neighbor

The mean distance to the nearest-neighbor \( m_D \) is a measure that has been frequently used to characterize univariate point patterns, particularly in earlier studies. It is closely related to the classical Clark–Evans index CE (Clark and Evans 1954) that divides \( m_D \) by the corresponding expectation of \( m_D \) under CSR and has been used to determine whether a pattern is hyperdispersed, random, or clustered. The \( m_D \) metric can be generalized to characterize a pattern by means of the mean distances \( nn(k) \) to the \( k \)th neighbor. This summary statistic has the convenient property of approximating a power law (Hubbell et al. 2008) and thus may have the potential to reveal scale independent features of a pattern.

For a random pattern, \( nn(k) \) can be estimated analytically and approximated by a power law (Table 2.2):

\[
nn(k) = \frac{1}{\sqrt{\lambda}} \frac{k(2k)!}{(2k)!} \approx \frac{1}{\sqrt{\pi \lambda}} k^{1/2}
\]  

(3.48)

Figure 3.17 shows the estimates for \( nn(k) \) for three different patterns, employing the Hanisch edge correction method (disks) and estimates without edge correction (lines in Figure 3.17d). The estimates for the random pattern agree almost perfectly with those of the power law (Figure 3.17d). The slope fit to the observed values for the Hanisch estimator yield a value of 0.507 instead of the expected value of 0.5 and the observed and expected mean neighborhood distances are 8.5 and 8.9, respectively, which is in good agreement. Figure 3.17e shows that \( nn(k) \) for clustered patterns also follows the power law very well, but the slope differs from 0.5. Clearly, the value at \( k = 1 \), the mean nearest-neighbor distance, depend on the intensity of the process and the local cluster structure. If the intensity and/or the clustering are greater, the mean nearest-neighbor distance is shorter and \( nn(k = 1) \) is smaller. The slope depends on the clustering of the pattern; the more clustered in the pattern, the larger the distance to the \( k \)th neighbor becomes for large neighborhood ranks \( k \). Additionally, the fewer the number of individuals there are, the greater the slope (Hubbell et al. 2008). Figure 3.17d shows the estimates of \( nn(k) \) based on the Hanisch edge correction method (disks) and the estimates without edge correction as lines. For the first 64 neighborhood ranks the mean distance to the \( k \)th neighbor is approximated well, but at ranks greater than 64 the difference between the methods becomes visible.

3.1.3.3 Spherical Contact Distribution

Second-order and nearest-neighbor statistics describe the properties of a pattern from the viewpoint of the points of the pattern and, therefore, mostly characterize neighborhood properties of the typical point. However, they
are not well suited to provide a statistical description of the properties of the “holes” of the pattern. This gap is filled with the spherical contact distribution $H_s(r)$ that characterizes the pattern from the perspective of “test points” that are randomly or regularly distributed in the observation window. Because of this property, the $H_s(r)$ provides additional information for characterizing patterns that can be used, for example, when testing the fit of parametric point-process models.

The spherical contact distribution can be viewed as a bivariate nearest-neighbor distribution function in which the focal pattern is a “test pattern” and the second pattern is the observed pattern. Therefore, the spherical contact distribution $H_s(r)$ is the distribution of the distances $r$ from the typical test location to the nearest point of the pattern. A random pattern can be used to generate the test locations or they can be based on the nodes of a grid. In general, the test locations should “sufficiently” cover the observation window to produce good resolution in the estimate of $H_s(r)$.

**FIGURE 3.17**
Summary statistic for the mean distance $nn(k)$ to the kth neighbor for three patterns (a) Random pattern with 1000 points in a 500 × 500 m² study area. (b) Clustered pattern with 626 points in a 500 × 500 m² study area. (c) Inhomogeneous pattern with 1296 points in a 1000 × 500 m² study area. (d) Summary statistic $nn(k)$ estimated for the three patterns (see legend in panel (e)) with continuous lines representing estimates without edge correction and dots estimates using the Hanisch method. (e) Same as panel (d), but lines show a fit with a power law to the estimates of mean nearest-neighbor distances using the Hanisch method.
The spherical contact distribution can be estimated as a bivariate distribution function to the nearest neighbor using Hanisch edge correction:

\[
\hat{H}_s(r) = \frac{\sum_{i=1}^{n} 1(0 < d_i \leq r) \times 1(d_i < \beta_i) \times w(d_i)}{\sum_{i=1}^{n} 1(d_i < \beta_i) \times w(d_i)}
\] (3.49)

where \(d_i\) is the distance from test location \(t\) to the nearest point of the pattern and \(\beta_i\) is the distance from the test location to the nearest border of \(W\). The weighting factor \(w(d_i)\) is the same as in Equation 3.46.

An alternative way of estimating the spherical contact distribution is to take advantage of its interpretation as a morphological function (Mecke and Stoyan 2005). In this case, each point of the pattern is enlarged to be a disk of radius \(r\). As a result, a pattern of potentially overlapping disks is obtained. The topology or morphology of this pattern is then analyzed for different values of \(r\). On first view, the spherical contact distribution is, somewhat surprisingly, the fraction of the observation window covered by the union of all disks of radius \(r\) centered on the points of the pattern (Illian et al. 2008, p. 200). This quantity, referred to as \(A_2(r)\), yields the probability that the first neighbor to an arbitrary “test point” in \(W\) is within distance \(r\) (Figure 2.14). Illian et al. (2008, p. 204) recommend the use of minus sampling edge correction with the estimator

\[
\hat{H}_s(r) = \frac{v(X_r \cap W_{\Theta r})}{v(W_{\Theta r})}
\] (3.50)

where \(X_r\) is the union of all (possibly overlapping) disks of radius \(r\) centered on the points of a pattern (the gray shaded area in Figure 2.14), the symbol \(v(X)\) denotes the area of \(X\), and \(W_{\Theta r}\) is the inner rectangle with area \(v(W_{\Theta r}) = (a - r)(b - r)\) of the rectangle \(W\) with sides of length \(a\) and \(b\). The numerator estimates the area of \(X_r\) within the reduced window \(W_{\Theta r}\). It is often estimated based on an underlying grid. Brodatzki and Mecke (2002) provide an algorithm that is able to estimate (3.50) without approximation.

3.1.4 Summary Statistics for Bivariate Patterns (Data Type 2)

3.1.4.1 Partial Product Densities

The estimation of the second-order statistics for bivariate patterns is completely analogous to the approach used for univariate patterns. We, therefore, do not present the estimators for bivariate patterns in much detail. The general form of the estimator of the product density given in Equation 3.19...
for univariate patterns is maintained. However, a new indicator function \( C_{lm}(x_i, x_j) \) is required that returns a value of 1 for point pairs when the first point \( x_i \) is of type \( l \) and the second point \( x_j \) is of type \( m \), and a value of zero otherwise. The general form of the estimator of the “partial” product densities for a pattern of \( n \) points, which may be of different types (e.g., species), therefore yields

\[
\hat{\rho}(r) = \frac{1}{2\pi r} \frac{1}{A} \sum_{i=1}^{n} \sum_{j=1}^{n, \neq} k(\|x_i - x_j\| - r) w_{i,j} \quad \text{univariate}
\]

\[
\hat{\rho}_{lm}(r) = \frac{1}{2\pi r} \frac{1}{A} \sum_{i=1}^{n} \sum_{j=1}^{n, \neq} C_{lm}(x_i, x_j) k(\|x_i - x_j\| - r) w_{i,j} \quad \text{bivariate (3.51)}
\]

where \( k() \) is the kernel function that selects point pairs that are located approximately at distance \( r \) apart, \( C_{lm}(x_i, x_j) \) selects point pairs where the first point is of type \( l \) and the second of type \( m \), and \( w_{ij} \) is the edge correction weight for the point pair \( i-j \). All four edge correction weights (summarized in Equation 3.21) apply without change. Note that the estimator of the univariate “partial” pair-correlation function for type \( m \) points yields

\[
\hat{\rho}_{mm}(r) = \frac{1}{2\pi r} \frac{1}{A} \sum_{i=1}^{n} \sum_{j=1}^{n, \neq} C_{mm}(x_i, x_j) k(\|x_i - x_j\| - r) w_{i,j} \\
= \frac{1}{2\pi r} \frac{1}{A} \sum_{i=1}^{n_m} \sum_{j=1}^{n_m, \neq} k(\|x_i^m - x_j^m\| - r) w_{i,j} \quad (3.52)
\]

where \( n_m \) is the number of type \( m \) points, and the superscript \( m \) indicates that the points \( x_i^m \) and \( x_j^m \) of the partial pattern consist of type \( m \) points.

### 3.1.4.2 Partial Pair-Correlation Functions

For the bivariate case, the (partial) pair-correlation functions based on the product densities given in Equation 3.51 require two adapted estimators of the intensity functions, one for the partial pattern of type \( l \) points, and one for the partial pattern of type \( m \) points. For points of type \( m \)

\[
\hat{\lambda}_m^m(r) = \frac{\sum_{i}^{n} C_m(x_i) \times v_{d-1}(W \cap \partial B(x_i, r))}{2\pi r \gamma_W(r)} \quad (3.53)
\]
Equation 3.53 includes the indicator function \( C_m(x) \) that selects only points of type \( m \) from the pattern. Thus, the estimators for the partial pair-correlation functions yield

\[
\hat{g}_{lm}^{R}(r) = \frac{\hat{\rho}_{lm}^{R}(r)}{\lambda_{l}^{m}(r)\lambda_{m}^{l}(r)} \quad \text{Ripley}
\]

\[
\hat{g}_{lm}^{S}(r) = \frac{\hat{\rho}_{lm}^{S}(r)}{\lambda_{l}^{m}(r)\lambda_{m}^{l}(r)} \quad \text{Stoyan}
\]

\[
\hat{g}_{lm}^{O}(r) = \frac{\hat{\rho}_{lm}^{O}(r)}{\lambda_{l}^{m}(r)\lambda_{m}^{l}(r)} \quad \text{Ohser}
\]

\[
\hat{g}_{lm}^{WM}(r) = \frac{\hat{\rho}_{lm}^{WM}(r)}{\lambda_{l}^{m}(r)\lambda_{m}^{l}(r)} = \frac{\hat{\rho}_{lm}^{O}(r)}{\lambda_{l}^{m}(r)\lambda_{m}^{l}(r)} \quad \text{WM}
\]

Note that the WM estimator uses the adapted intensity estimator for the second pattern (i.e., type \( m \) points), but the natural estimator for the focal pattern \( l \). The reason for this is that the estimator of the bivariate product density in Equation 3.51 can be rewritten as \((1/2\pi)(n/l)(1/n)\sum_{i=1}^{n} \sum_{j=1}^{n} C_{lm}(x_i, x_j)k(||x_i - x_j|| - r)\omega_{i,j} \), which includes the natural estimator of the intensity of pattern \( l \).

Figure 3.18 compares the different estimators provided in Equation 3.54 for a bivariate pattern comprising a clustered focal pattern and an independent random pattern (Figure 3.18a). Because the two patterns are independent by construction, the expectation for this pattern yields \( g_{12}(r) = 1 \). We find that the WM, Ohser and Stoyan estimators yield almost identical estimates, whereas the Ripley estimator shows a slight positive bias (Figure 3.18b). This is because the focal pattern is a clustered pattern with several clusters close to the border, which yields weights that are too large, analogous to the univariate case for the focal pattern (Figure 3.14d). For the “real life” pattern shown in Figure 3.18d, in which there is no strong clustering in the focal pattern close to the border, we find only small differences among the various estimators (Figure 3.18e,f).

To put the small departures among estimators into perspective, we also show in Figure 3.18b the simulation envelopes of a null model, based on the WM estimator. In the simulations, the clustered focal pattern is unchanged, but pattern 2 is randomized based on CSR. The width of the simulation envelopes for the pair-correlation function, at any distance \( r \), is greater than the differences among estimators. However, the Ripley estimator shows departure from the other estimators, approaching a magnitude of half of the width of the simulation envelopes, at largest distances \( r \) (Figure 3.18b). These examples outline that the differences among the three estimators for the bivariate pair-correlation function (WM, Ohser, and Stoyan) are negligible in practical applications, but that the Ripley estimator may in some cases produce biased results.
3.1.4.3 Partial Inhomogeneous Pair-Correlation Functions

The estimators of the inhomogeneous (cross) pair-correlation functions require two separate estimates of the intensity functions \( \lambda_l(x) \) and \( \lambda_m(x) \) for the type \( l \) and \( m \) patterns, respectively. Analogous to Equation 3.25 for the univariate, pair-correlation function, the BMW estimator of the inhomogeneous partial pair-correlation function yields

**FIGURE 3.18**
Estimators for bivariate second-order statistics. (a) Simulated bivariate pattern composed of a clustered focal pattern (closed disks) and a second, random pattern (open disks). (b) Estimators of the bivariate pair-correlation functions for the pattern shown in panel (a) together with simulation envelopes of the CSR null model based on the WM estimator. (c) Same as (b) but for the bivariate \( L \)-function. (d) Bivariate pattern of two tropical tree species. (e) Estimators of the bivariate pair-correlation functions for the pattern shown in panel (d). (f) Estimators of the bivariate \( L \)-function for the pattern shown in panel (d).
\[ \hat{g}_{lm}^{\text{BMW}}(r, \lambda_l(x), \lambda_m(x)) = \frac{1}{2\pi r A} \sum_{i=1}^{n_l} \left( \sum_{j=1}^{n_m} C_{lm}(x_i, x_j) \times k(\|x_i - x_j\| - r) \right) \frac{\hat{w}_i(r)}{\lambda_l(x_i) \lambda_m(x_j)} \]  

(3.55)

And, analogous to Equation 3.29, the estimator of the inhomogeneous partial pair-correlation function based on a generalization of the Ohser edge correction yields

\[ \hat{g}_{lm}^{\text{O}}(r, \lambda_l(x) \lambda_m(x)) = \frac{1}{\lambda_l \lambda_m} \frac{1}{2\pi r A} \sum_{i=1}^{n_l} \left( \sum_{j=1}^{n_m} C_{lm}(x_i, x_j) k(\|x_i - x_j\| - r) \right) \frac{A}{\hat{\gamma}_W(r, \lambda_l(x) \lambda_m(x_j))} \]  

(3.56)

where the \( \hat{\lambda}_l \) and \( \hat{\lambda}_m \) are the natural estimators of the intensity of type \( l \) and type \( m \) points, respectively. In this case, the generalized, isotropic-set covariance yields

\[ \hat{\gamma}_W(r, \lambda_l(x) \lambda_m(x)) = \frac{1}{\lambda_l \lambda_m} \frac{1}{2\pi r} \int W \left[ \int W \lambda_l(a) k(\|x - a\| - r) da \right] dx \]  

(3.57)

Analogous to Equation 3.31, the estimator of the inhomogeneous, partial pair-correlation function based on a generalization of the WM edge correction yields

\[ \hat{g}_{lm}^{\text{WM}}(r, \lambda_l(x), \lambda_m(x)) = \frac{1}{n_l} \sum_{i=1}^{n_l} \left( \sum_{j=1}^{n_m} C_{lm}(x_i, x_j) k(\|x_i - x_j\| - r) \right) \] \[ \frac{1}{n_l} \sum_{i=1}^{n_l} C_l(x_i) \int_W \lambda_m(a) k(\|x_i - a\| - r) da \] \[ \approx \frac{\sum_{i=1}^{n_l} \left( \sum_{j=1}^{n_m} C_{lm}(x_i, x_j) k(\|x_i - x_j\| - r) \right)}{(n_m - 1)/s \sum_{i=1}^{n_l} \left( \sum_{j=1}^{n_m} C_{lm}(x_i, a_j) k(\|x_i - a_j\| - r) \right)} \]  

(3.58)

As can be seen in the lower expression, the integral in the denominator can be approximated by distributing \( s \) type \( m \) points (i.e., \( a_j \)) within the observation window, using a heterogeneous Poisson process with intensity function \( \lambda_m(x) \). The \( n_l \) are the number of points of type \( l \). Note that the intensity function of the focal pattern \( l \) does not enter into this estimator, because the type...
190 Handbook of Spatial Point-Pattern Analysis in Ecology

l points are used as focal points and are assumed to already be the result of thinning a homogeneous pattern with the intensity function \( \lambda(x) \).

### 3.1.4.4 Partial K-Functions

The quantity \( \lambda_l \lambda_m K_{lm}(r) \) is used for the estimation of the partial K-functions and is analogous to Equation 3.32:

\[
\hat{K}_{lm}(r) = \frac{1}{A} \sum_{i=1}^{n} \sum_{j=1}^{n_i} C_{lm}(x_i, x_j) \times 1(\|x_i - x_j\|, r) \times w_{ij}
\]

(3.59)

The edge correction weights for the univariate case given in Equation 3.34 can be used. The adapted estimator of the intensity of type \( l \) points yields

\[
\hat{\lambda}_l(r) = \frac{\sum_i^n C_l(x_i) \times v(W \cap b(x_i, r))}{2\pi \int_0^r t\bar{f}_l(t)dt}
\]

(3.60)

where the indicator function \( C_l(x_i) \) selects all points \( i \) of the focal type \( l \). Analogous to Equation 3.37, the various estimators of the K-function are

- \( \hat{K}_{lm}^{WM}(r) = \frac{\hat{K}_{lm}^{WM}(r)}{\hat{\lambda}_l(r)\hat{\lambda}_m(r)} \) WM
- \( \hat{K}_{lm}^{O}(r) = \frac{\hat{K}_{lm}^{O}(r)}{\hat{\lambda}_l(r)\hat{\lambda}_m(r)} \) Ohser
- \( \hat{K}_{lm}^{S}(r) = \frac{\hat{K}_{lm}^{S}(r)}{\hat{\lambda}_l(r)\hat{\lambda}_m(r)} \) Stoyan
- \( \hat{K}_{lm}^{R}(r) = \frac{\hat{K}_{lm}^{R}(r)}{\hat{\lambda}_l(r)\hat{\lambda}_m(r)} \) Ripley

(3.61)

where the type of point used in estimating lambda is indicated as a superscript.

The estimators of the bivariate L-function derived from the K-function (Equation 3.61) mirror the results for the bivariate, pair-correlation function (Figure 3.18). The estimates based on the WM, Ohser and Stoyan estimators are practically identical, whereas the estimate of the L-function based on the Ripley edge correction shows a notable difference, with a positive bias in the example pattern (Figure 3.18c). When contrasting the L-function and the corresponding simulation envelopes for the WM estimator (Figure 3.19b) with

© 2014 Taylor & Francis Group, LLC
Estimators and Toolbox

those of the Ripley estimator (Figure 3.19c), we find that the Ripley estimator shows a considerable bias in the expectation (gray line) not shown by the WM estimate (and the Ohser and Stoyan estimator; not shown). This is because the Ripley estimator weights the large number of points located in the clusters close to the border of $W$ with too high a value, resulting in an overestimation based on the number of nearby points in the second pattern. This creates increasingly positive values of the $L$-function with increasing distance $r$ (Figure 3.19c). However, if the inference is based on simulation envelopes, this is not a problem because the estimates of the realizations of the null model are subject to the same bias and Figure 3.19c still indicates independence. However, the bias in the Ripley estimator may become a problem if a parametric fit is used.

Again, we see a situation paralleling the inhomogeneous case for the univariate estimators in developing estimators for the bivariate case. So briefly, the BMW estimator of the inhomogeneous cross $K$-function yields (cf. Equation 3.38)

$$\hat{K}_{lm}^{\text{BMW}}(r, \lambda_l(x), \lambda_m(x)) = \frac{1}{A} \sum_{i=1}^{n_l} \sum_{j=1}^{n_m} C_{lm}(x_i, x_j) \times 1(\|x_i - x_j\|, r) \frac{w_{ij}(r)}{\lambda_l(x_i)\lambda_m(x_j)}, \quad (3.62)$$
and the Ohser edge correction yields (cf. Equation 3.41)

\[
\hat{K}^{\text{Oh}}_{lm}(r, \lambda_l(x), \lambda_m(x)) = \pi r^2 \sum_{i=1}^{n} \left( \sum_{j=1}^{n} C_{lm}(x_i, x_j) \times \mathbf{1}(\|x_i - x_j\|, r) \right) \int_{W} \lambda_l(x) \int_{W} \lambda_m(a) \mathbf{1}(\|x - a\|, r) da \ dx
\]  

(3.63)

and finally the WM edge correction yields (cf. Equation 3.42)

\[
\hat{K}^{\text{WM, inhom}}_{lm}(r, \lambda_l(x), \lambda_m(x)) = \pi r^2 \sum_{i=1}^{n} \left( \sum_{j=1}^{n} C_{lm}(x_i, x_j) \times \mathbf{1}(\|x_i - x_j\|, r) \right) \sum_{i=1}^{n} C_i(x_i) \int_{W} \lambda_m(a) \mathbf{1}(\|x - a\|, r) da \]

(3.64)

### 3.1.4.5 Partial Nearest-Neighbor Summary Statistics

The estimators of the nearest-neighbor distribution functions for bivariate patterns can be derived in a natural way from those for univariate patterns. If we have a homogeneous and bivariate pattern composed of type \(l\) and type \(m\) points, the summary statistic \(D^k_{lm}(r)\) is the distribution function of the distances from a typical point of type \(l\) to the \(k\)th nearest point of type \(m\). In this case, we can generalize the estimators of the univariate case by introducing an indicator function \(C_{lm}(x_i, x_j)\) that selects from among all point pairs \(x_i\) and \(x_j\) only those where the first point is of type \(l\) (i.e., point \(x_i\)) and the second is of type \(m\) (i.e., point \(x_j\)). If, as before, we define \(d_{ik}\) as the distance from point \(x_i\) to the \(k\)th neighbor of type \(m\) (i.e., points \(x_j\)), the naive estimator of \(D^k_{lm}(r)\) without edge correction is then analogous to Equation 3.43:

\[
\hat{D}^k_{lm}(r) = \sum_{i=1}^{n} \frac{\mathbf{1}(0 < d_{ik} \leq r) \times C_{lm}(x_i, x_j)}{\sum_{i=1}^{n} C_i(x_i)}
\]  

(3.65)

where the indicator function \(C_{lm}(x_i, x_j)\) selects pairs of points in which the first point \(i\) is of type \(l\) and the second point \(j\) is of type \(m\). \(C_i(x)\) selects all points \(i\) of the focal type \(l\) and the indicator function \(\mathbf{1}(0 < d_{ik} \leq r)\) yields the value 1, if \(d_{ik}\) is smaller than or equal to \(r\), and zero otherwise. Thus, the summation counts, for a given distance \(r\), all focal points \(i\) which are of type \(l\) and have their \(k\)-nearest-neighbor type \(m\) within distance \(r\). The denominator then normalizes by counting the number of focal points \(i\) which are of type \(l\).
The estimator using the buffer method is analogous to the univariate estimator, but now we need to add the condition that the distance $\beta_i$ of the focal points $x_i$ to the border is greater than the buffer distance $b$. This condition is implemented as the indicator function $1(b < \beta_i)$ which yields a value of 1 if distance $\beta_i$ is smaller than $b$ and zero otherwise:

$$\hat{D}_{lm}^{b,h}(r) = \frac{\sum_{i=1}^{n} [1(0 < d_{ik} \leq r) \times [C_{lm}(x_i, x_i) \times 1(b < \beta_i)]]}{\sum_{i=1}^{n} [C_l(x_i) \times 1(b < \beta_i)]}, \quad r \leq b \quad (3.66)$$

The Hanisch estimator selects the focal points $i$ not based on the buffer, but decides for each point $i$ of type $l$, individually, if it is used as a focal point. To this end, focal points are selected only if the $k$-nearest-neighbor of type $m$ can be correctly determined. The condition for this is that the distance $d_{ik}$ to the $k$-nearest-neighbor of type $m$ is smaller than the distance $\beta_i$ to the border. This condition is implemented by using the function $1(d_{ik} < \beta_i)$, which yields a value of 1 if the $d_{ik}$ is smaller than $\beta_i$ and zero otherwise:

$$\hat{D}_{lm}^{h,H}(r) = \frac{\sum_{i=1}^{n} [1(0 < d_{ik} \leq r) \times [C_{lm}(x_i, x_i) \times 1(d_{ik} < \beta_i) \times w(d_{ik})]]}{\sum_{i=1}^{n} [C_{lm}(x_i, x_i) \times 1(d_{ik} < \beta_i) \times w(d_{ik})]} \quad (3.67)$$

The weighing factor $w(d_{ik})$ compensates for the points of type $l$ which are not used as focal points and, for a rectangular observation window with side length $l_x$ and $l_y$, yields

$$w(d_{ik}) = \frac{1}{(l_x - d_{ik})(l_y - d_{ik})} \quad (3.68)$$

Comparing Equations 3.66 and 3.67 shows that the indicator function $1(b < \beta_i)$, which defines the buffer in Equation 3.66, is replaced by the weighted indicator function $1(d_{ik} < \beta_i) \times w(d_{ik})$, which defines the focal points of the Hanisch estimator, in Equation 3.67.

While the estimators for bivariate patterns are completely analogous to those for univariate patterns, there is one important difference with respect to sensitivity of the estimators to departures from homogeneity. To illustrate this potential problem, we use an extreme case of a heterogeneous bivariate pattern (Figure 3.20a) in which type 1 points (indicated as closed disks) are randomly distributed in the western half ($x < 250$) of a $500 \times 500$ m$^2$ observation window and type 2 points (open disks) are randomly distributed in the eastern half ($x > 250$). In this case, we observe a large difference between the naive estimator without edge correction and the Hanisch estimator (Figure 3.20c). Somewhat
surprisingly, the Hanisch estimator of $D_{12}(r)$ yields, on first examination, a value of 1 for distances greater than 125 m, whereas the naive estimator yields a value of approximately 1 only at distances greater than 250 m. The latter was expected because points with an $x$-coordinate close to zero will have their nearest neighbor of type 2 approximately at a distance of 250 m, and all points with an $x$-coordinate larger than zero will have their nearest type 2 neighbor at distances below 250 m. This suggests that, the naive estimator approximates the true $D_{12}(r)$ better than the Hanisch estimator, at least in this example employing an extremely heterogeneous pattern.

The difference between the Hanisch and naive estimators can be explained by the characteristics of the two in relation to the pattern. The Hanisch

![Figure 3.20](image-url)
Estimators and Toolbox

195

The estimator uses only focal points \(i\) for which the distance \(d_{ii}\) to the nearest neighbor is smaller than its distance to the nearest border of the observation window. In our particular example, those focal points are located inside the gray shaded trapezoid as shown in Figure 3.20a. In the univariate case, there is no reason to assume that the (true) nearest-neighbor distances of points closer to the border (i.e., outside the trapezoid) would differ principally from those used for estimation (i.e., inside). However, this is not the case in our example. The distances to the nearest type 2 neighbor inside the trapezoid are always smaller than 125 m, but outside they may reach distances up to 250 m (for points with \(x\) coordinates close to zero). In particular, the nearest-neighbor distance is greater than 125 m for all points with \(x\) coordinates smaller than 125 m. For this reason, the Hanisch estimator yields \(D_{12}(r = 125\,\text{m}) = 1\). However, this means that the focal points selected by the Hanisch (and the buffer) estimator are not “representative” of the type 1 points (with respect to their type 2 neighbors) and this particular selection of focal points results in the bias observed in Figure 3.20c.

This extreme example illustrates that the buffer and the Hanisch estimator for bivariate patterns may not be robust against heterogeneities of a specific type, in which the selection of the focal points is not “representative” with respect to the nearest-neighbor distances. However, if type 2 points do overlap sufficiently well with the area occupied by type 1 points, this problem does not occur. This is illustrated in Figure 3.20e, which provides the results of a bivariate pattern analysis, with the focal type 1 points located only in the western part of the plot (as in Figure 3.20a) and the type 2 points randomly distributed within the entire plot (as for the pattern in Figure 3.20b, but with the open disks being pattern 1 and the closed points being pattern 2). In this case, the type 2 points overlap the area covered by type 1 points and the Hanisch and naive estimators agree in their results (Figure 3.20e). However, if we switch the role of type 1 and 2 (Figure 3.20b), the problem reappears as shown in Figure 3.20d. In this case, the type 1 focal points are randomly distributed within the entire observation window, but type 2 points overlap only half of this area. The resulting expectation of \(D_{12}(r)\) is a straight line between points \((0, 0.5)\) and \((250, 1)\) (Figure 3.20d), which is captured by the naive estimator. The straight line actually arises because half of the focal points have their type 2 nearest neighbor quite close by (those in the western part of the plot), whereas those in the eastern part of the window have their nearest type 2 neighbor located at an \(x\)-coordinate of approximately 250.

The above examples suggest that it may be safer to use the naive estimator without edge correction for bivariate patterns that are suspected of heterogeneity, due to the focal points not being “representative” of the nearest-neighbor distances found within the entire observation window. The smaller bias introduced by not always having the true nearest neighbor inside the observation window may be smaller in this case than the bias arising through edge correction.
3.1.5 Summary Statistics for Multivariate Patterns (Data Type 3)

Although the modification of summary statistics to accommodate bivariate patterns is, in general, straightforward, generalization to truly multivariate summary statistics is not. The issues for accomplishing this are, in fact, not yet entirely resolved. As a result, truly multivariate summary statistics that would be able to summarize the spatial structure of multivariate point-pattern data, in a simple and ecologically meaningful way, are lacking. The few attempts to develop summary statistics for multivariate patterns represent logical extensions of the summary statistics for uni- or bivariate patterns. They can, in fact, be expressed as the sum of bivariate summary statistics. Each one generally characterizes only one specific aspect of the potentially complex structure of multivariate patterns and often represents a spatial analog to classical nonspatial diversity indices.

3.1.5.1 Spatially Explicit Simpson Index

One approach to analyzing multivariate patterns, which has been applied to multispecies communities, such as tropical forests, considers summary statistics that average the various possible pairwise, partial, summary statistics into a community average. For example, the spatially explicit Simpson index is basically the conditional probability $\beta(r)$ that two points separated by distance $r$, belong to different types (Plotkin et al. 2000; Shimatani 2001; Chave and Leigh 2002). The formula for this is given by

$$
\beta(r) = 1 - \sum_{m=1}^{S} \frac{\lambda_m^2 g_{mm}(r)}{\lambda^2 g(r)}
$$

(3.69)

where $S$ is the number of species in the community, $g(r)$ is the pair-correlation function for the entire community (i.e., all individuals of all species are treated as one pattern), $g_{mm}(r)$ is the partial pair-correlation function of species $m$, $\lambda_m$ is the intensity of type $m$ points, and $\lambda$ is the intensity of all points of the multivariate pattern. If all patterns are random patterns, the expectation of $\beta(r)$ is a constant and yields $\beta^* = 1 - \sum (\lambda_m / \lambda)^2$. This is the classical Simpson index (Simpson 1949) and yields the probability that two arbitrarily chosen individuals (without replacement) are different species (Shimatani and Kubota 2004). This index quantifies the degree of evenness in species composition. $\beta^*$ is small if one (or a few) species dominate the community, while $\beta^*$ reaches its maximum, $\beta^* = 1 - 1/S$, if the abundances $n_m$ of all species are equal, that is, $n_m = n/S$, where $n$ is the total number of individuals in the community and $S$ the total number of species.

If the patterns of the individual species tend to be clumped, the spatially explicit Simpson index $\beta(r)$ will be below the expectation $\beta^*$ for small distances $r$ and, if the individual species tend toward hyperdispersed patterns, $\beta(r)$ will be above the expectation $\beta^*$ for small distances $r$. In the case of
conspecific clustering and heterospecific segregation, the index $\beta(r)$ would yield substantially lower values than $\beta^*$ at small distances, because in this case most neighbors would be conspecifics.

Figure 3.21 provides an example of the spatially explicit Simpson index $\beta(r)$ calculated for recruits from the 2005 census at the BCI forest plot in Panama (Box 2.6). In this case, recruits are defined to be trees that were smaller than 1 cm diameter at breast height in 2000 and were larger than that in 2005. There were, in total, 23,290 recruits of 246 species (Figure 3.21a). Thus, the classical Simpson index $\beta^*$ for an evenly distributed community would yield $\beta^* = 0.996$, however, we find $\beta^* = 0.968$. Thus, the recruits are clearly not uniformly distributed. Figure 3.21b shows the spatially explicit Simpson index $\beta(r)$. It indicates that there is a strong, spatial structure at shorter distances of $r$ (<10 m), at which the recruits of different species

![Spatially explicit Simpson index $\beta(r)$ for $\beta$ diversity.](image-url)

**FIGURE 3.21**
Spatially explicit Simpson index $\beta(r)$ for $\beta$ diversity. (a) Spatial pattern of all recruits (i.e., individuals that grew to a dbh $> 1$ cm during a five-year period) for a 500 x 500 m² area of the BCI plot. (b) Simpson index $\beta(r)$ for the pattern of recruits (closed disks) and for all reproductive individuals (open disks), as well as the corresponding cumulative indices $\alpha(r)$ (triangles). The horizontal lines show the asymptotic expectation $\beta^*$ for $\beta(r)$ at large scales. (c) Same as panel (b), but for the normalized indices $\beta(r)/\beta^*$ and $\alpha(r)/\beta^*$. (d) Pair-correlation function of all recruits (closed disks) and all reproductive individuals (open disks).
tend to be clustered. Two recruits located at distance \( r = 1 \) m are, with an 85% probability, not of the same species, but for the nonspatially explicit Simpson index \( \beta^* = 0.968 \) the expectation is that 97% of all randomly selected pairs of individuals will be heterospecific. Two recruits located at distance \( r = 10 \) m still have a 94.7% probability of being heterospecific. The pair-correlation function of all recruits shows strong clustering up to distances of say \( r = 30 \) m (Figure 3.21d), with recruits having a 4 times higher neighborhood density at 0.5 m than expected by their over all density [i.e., \( g(r = 0.5) = 4 \)]. This indicates that the community of recruits exhibits clustering (see also Figure 3.21a), with some intraspecific clustering at smaller scales. However, the relatively high values of \( \beta(r) \) at short distances [e.g., \( \beta(r = 1 \) m) = 0.85] indicate that the clusters are composed of a mixture of species.

Figure 3.21b also shows an analysis for all reproductive individuals (i.e., individuals with a diameter above a species-specific threshold; open symbols). We find in total 71,597 reproductive individuals from 272 species. Thus, the classical Simpson index \( \beta^* \) for an evenly distributed community would yield \( \beta^* = 0.996 \); however, we find \( \beta^* = 0.831 \). Clearly, the abundances of reproductive individuals are not uniformly distributed and are even more unevenly distributed than recruits. Figure 3.21b shows the spatially explicit Simpson index \( \beta(r) \). We find a strong spatial structure at distances \( r \) of less than approximately <30 m, at which scales the different species tend to be clustered. Two reproductive trees located at distance \( r = 1 \) m are, with a 69% probability, heterospecifics, and two reproductive individuals at distance \( r = 10 \) m still have a 77% probability of being heterospecifics. The pair-correlation function of all reproductive individuals shows clustering up to distances of say \( r = 30 \) m (Figure 3.21d). This pattern, however, is much weaker than for recruits.

If the multivariate pattern does not show strong heterogeneity, the spatially explicit Simpson index \( \beta(r) \) approaches the nonspatial Simpson index for large distances \( r \). Because of this, it is somewhat complicated to compare the pure spatial effects in the decline of similarity in local species composition (i.e., Figure 3.21b). To compare the spatial structure between different communities, we may therefore normalize the spatially explicit Simpson index \( \beta(r) \) with its nonspatial expectation \( \beta^* \) (Figure 3.21c). This analysis shows that reproductive individuals have a stronger spatial structure than recruits, where the species richness is less evenly distributed locally than that of the recruit community.

Note that the numerator and denominator of Equation 3.69 are product densities and that the partial product density of species \( m \), divided by the full product density, yields a mark connection function, which is discussed in more detail in the next section. If we insert the product densities from Equation 3.51 into Equation 3.69, we obtain an estimator of the spatially explicit Simpson index \( \beta(r) \):
The index can be simplified by using the indicator function \( 1(m_i = m_j) \), which yields one if the species identifier of point \( i (= m_i) \) is the same as the species identifier of point \( j (= m_j) \) and zero otherwise. If the Ohser weights are used, the edge correction factors \( w_{ij} \) of the numerator and denominator cancel out, because they are independent of the concrete point pair \( i-j \). This suggests that this estimator, which is the ratio of two product densities, does not need edge correction. Note that the index \( \beta(r) \) is conditional by nature, since dividing by \( g(r) \) factors out the overall pattern of the entire community.

As we will see in Section 3.1.7.6, the spatially explicit Simpson index can be generalized to consider different levels of similarity between species. While Equation 3.70 treats all heterospecific individuals equally [i.e., \( 1(m_i = m_j) = 1 \) for \( i \neq j \)], a more nuanced view of the spatial structures in a community can be gained by considering alternative axes of biodiversity, such as phylogenetic and functional diversity. For example, closely related (or functionally similar) species might be found in similar environments, but due to competitive exclusion the local neighborhoods around species may be characterized by more distantly related species (e.g., Weiher and Keddy 1995; Cavender-Bares et al. 2006; Swenson et al. 2006). Thus, consideration of species similarity in spatially explicit, summary statistics of multivariate patterns may allow us to better understand the mechanisms underlying observed biodiversity patterns. In Section 3.1.7.6 we will extend the spatially explicit Simpson index by including a continuous measure of species dissimilarity into Equation 3.70, instead of the indicator function \( 1(m_i = m_j) \). Since this index works with quantitative marks provided by the continuous dissimilarity index, the resulting summary statistics are mark-correlation functions and are treated therefore in detail in Section 3.1.7.

The cumulative spatially explicit Simpson index \( \alpha(r) \) is based on \( K \)-functions and is given by

\[
\alpha(r) = 1 - \sum_{m=1}^{S} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n,x} \lambda^2_{mm} K_{mm}(r)}{\lambda^2 K(r)}
\]  

(3.71)

Equation 3.71 expresses the probability that two points with a separation distance of no more than \( r \) are of different types (Shimatani 2001). The index
Handbook of Spatial Point-Pattern Analysis in Ecology

\( \alpha(r) \) is based on an estimation of the quantity \( \lambda^2 K(r) \). Here, the WM edge correction weight (Equation 3.34), which approximates the Ohser weight, also cancels out. Figure 3.21b shows that this index behaves in a manner similar to \( \beta(r) \), but because it is a cumulative index it shows scale effects less clearly.

### 3.1.5.2 Individual Species–Area Relationship ISAR

Another summary statistic that has been proposed for multivariate point patterns is the *spatial species richness* \( d(r) \) (Shimatani and Kubota 2004). It is defined as the expected number of different species present within distance \( r \) from an arbitrarily chosen point in an observation window \( W \). Thus, this summary statistic is the point-pattern analog to the species–area relationship (SAR), if distance \( r \) is converted to area \( \pi r^2 \). Usually, the SAR is estimated by dividing the observation window into quadrats, counting the number of species in every quadrat, and calculating the average (Shimatani and Kubota 2004). However, the spatial species richness \( d(r) \) can be calculated by decomposing the index into the sum of the spherical contact distributions of all the univariate patterns:

\[
d(r) = \sum_{m=1}^{s} H_{s}^m(r)
\]

(3.72)

where \( H_{s}^m(r) \) is the partial spherical contact distribution for species \( m \). Decomposing the spatial species richness \( d(r) \) into the “detectability” \( H_{s}^m(r) \) of each species allows the role of each species in the resulting overall species diversity to be examined (Shimatani and Kubota 2004).

Figure 3.22 shows the spatial species richness \( d(r) \) for the recruits and reproductive individuals of the 2005 BCI census (Box 2.6). The species at the BCI forest are not well mixed. At a distance of 10 m, only 4.4% of all recruit species can be found on average at a random location within the plot (Figure 3.22b). At sample plots with a 40 m radius we find on average 27.2% of all species and, at sample plots with a radius of 88 m, we find 50% of all species (Figure 3.22b). This is, in part, a consequence of the spatial structure, as revealed by the spatially explicit Simpson index, but is also due to many species with low abundances. Only 163 species (= 66% of the total of 246 species represented by recruits) have more than 10 recruits. A similar result was found for the reproductive community, in which 175 species had more than 10 individuals. A more uneven distribution of abundances of reproductive individuals (compared with recruits), together with the spatial structure revealed by the spatially explicit Simpson index, yields a \( d(r) \) index below that of the recruits, even if total species richness is greater (272 vs. 246 species) (Figure 3.22a).

Another summary statistic closely related to \( d(r) \) is the *individual species–area relationship* ISAR,\( _x(r) \) introduced by Wiegand et al. (2007b). This summary statistic is defined to be the expected number of species present within
distance \( r \) of the typical individual of a focal species \( f \). It determines the average neighborhood species richness around individuals of the focal species and is therefore the point-centered analog to \( d(\cdot) \). The ISAR can be calculated by decomposing the index into the sum of the cross, nearest neighbor, distribution functions between the focal species \( f \) and all other species \( m \):

\[
\text{ISAR}_f(r) = \sum_{m=1}^{S} D_{fm}(r)
\]  

(3.73)

where \( D_{fm}(r) \) is the probability that an individual of the focal species \( f \) has its nearest neighbor of species \( m \) within distance \( r \). Given the potential problems with the Hanisch estimator (Figure 3.20), it is safer to estimate the ISAR function without edge correction or use the buffer method (as in Wiegand et al. 2007b). To complete the analogy with the other summary statistics presented in this section, the ISAR can also be averaged over all species:

\[
\overline{\text{ISAR}}(r) = \frac{1}{n} \sum_{f=1}^{S} n_f \text{ISAR}_f(r) = \frac{1}{n} \sum_{f=1}^{S} n_f \sum_{m=1}^{S} D_{fm}(r)
\]  

(3.74)

where \( n_f \) is the number of individuals of the focal species and \( n \) is the total number of individuals for the entire community. This summary statistic yields the average number of species within distance \( r \) of the typical individual of the community and is the point-centered analog to the location related, spatial species richness \( d(\cdot) \).

The special characteristic of the ISAR is that it views the structures in local species richness from the “plant’s-eye” viewpoint of individuals of a
given focal species (Wiegand et al. 2007b). Figure 3.23a shows an example of
the community of all large trees (dbh > 10 cm) at the BCI plot (1995 census;
Box 2.6) and sampling areas with a 10 m radius around the individuals of
the focal species *Simarouba amara*. The ISAR function determines the num-
ber of species in each of these sampling areas and then their average value.
If we vary the neighborhood radius \( r \) we obtain the full ISAR\((r)\) function
that quantifies scale-dependent spatial structures in local species richness
around a given focal species \( f \) (Figure 3.23c).

The ISAR function can be used in different ways to explore how the indi-
viduals of a focal species are located within the “landscape” of local species
richness as shown in Figure 3.23a. For example, it allows us to quantify if
a given species is predominately located in areas of below or above aver-
age species richness across a range of spatial scales. This can be tested by
contrast the observed ISAR curve to that of a null model that repeatedly

![Figure 3.23](image-url)

(See color insert.) ISAR for large trees (dbh > 10 cm) at the BCI plot, with *Simarouba amara* used
as an example. (a) “Landscape of local species richness” for neighborhoods of 10 m, estimated
as the number of species within 10 m of nodes on a regular grid. Open disks are the loca-
tions of the individuals of the species *S. amara*. (b) Observed values of ISAR\((r)\) adjusted by the
expectation for the null model of spatial species richness for *S. amara*. (c) Observed values of
ISAR\((r)\) for *S. amara*. In panels (b) and (c), closed disks represent the observed values, whereas
solid lines represent the simulation envelopes and expectation. The null model (i.e., CSR) was
implemented by randomly relocating *S. amara* within the observation window \( W \) and calculat-
ing the resulting ISAR\((r)\) values. *S. amara* is located, on average, in areas of higher local spe-
cies richness, as indicated by the positive departure from the null model in panel (b). Because
we subtracted the expectation of the null model (i.e., ISAR\(_{exp}\)) in (b) from the ISAR index, the
expectation under the null model is zero at all scales.
relocates the individuals of the focal species to random locations within the entire plot (Wiegand et al. 2007b). This null model, which approximates the spatial species richness \( d(r) \) (SAR; Shimantani and Kubota 2004), is the point-pattern analog of the common SAR. It estimates the mean number of species in neighborhoods of test points and links the ISAR concept with the SAR concept. The key question here is to find out whether the ISAR, \( r \) function of species \( f \) is significantly different from the point-pattern SAR, and identify at what spatial scales this occurs. Figure 3.23b shows that the species \( S. amara \) is predominantly located in areas of higher local species richness.

### 3.1.5.3 Phylogenetic Extension of the ISAR

In Section 3.1.5.1, we argued that measures of species diversity, when used alone, are relatively information poor, because they assign each species pair the same level of dissimilarity. Accordingly, we proposed a way of including continuous measures \( d(f, m) \) of species dissimilarity (for species \( f \) and \( m \)) into summary statistics of multivariate patterns. The phylogenetic Simpson index, which generalizes the spatially explicit Simpson index, is based on the sum of bivariate (partial) pair-correlation functions (see Section 3.1.7.6). A similar approach is also possible for multivariate summary statistics derived from the sum of bivariate (partial) nearest-neighbor distribution functions. Thus, we can generalize the ISAR to yield a multivariate summary statistic that includes a continuous measure of species dissimilarity.

Recall that the individual species–area relationship ISAR, \( r \) for species \( f \) can be estimated as the sum of the probabilities \( D_{fn} (r) \) over all species \( m \) that an individual of the focal species \( f \) has its nearest species \( m \) neighbor within distance \( r \). We generalize this expression by weighting the different species \( m \) according to their phylogenetic (or functional) distance \( d(f, m) \) from the focal species \( f \):

\[
\text{ISAR}_f (r) = \sum_{m=1}^{S, m \neq f} D_{fm} (r)
\]

\[
\text{PISAR}_f (r) = \sum_{m=1}^{S} d(f, m)D_{fm} (r)
\]

where \( S \) is the total number of species in the observation window. The PISAR function estimates the average phylogenetic dissimilarity \( d(f, s) \) (or phylogenetic distance) between all species \( m \) located within a neighborhood of radius \( r \) around the typical individual of the focal species \( f \). Or, in other words, it is the mean phylogenetic distance (MPD) of the typical focal individual to all species that are located in its neighborhood.

The PISAR function given in Equation 3.75 contains a signal of both the multivariate spatial structure of the community (captured by the ISAR function) and a signal of phylogenetic spatial structure. If there is no phylogenetic spatial structure present, or if all heterospecific species show the same
dissimilarity (i.e., a star shaped phylogeny), the phylogenetic PISAR function will basically collapse to the ISAR function. Thus, the difference between PISAR and ISAR is due to the phylogenetic spatial structure embedded in the multivariate pattern. To investigate the “pure” spatial phylogenetic signal, independent of species richness within the neighborhood, we normalize the PISAR$_f(r)$ with the ISAR$_f(r)$ to yield the rISAR function:

$$r\text{ISAR}_f(r) = \frac{\text{PISAR}_f(r)}{\text{ISAR}_f(r)} = \frac{\sum_{m=1}^{S} d(f,m)D_{fm}(r)}{\sum_{m=1}^{S_{m \neq f}} D_{fm}(r)}$$  \hspace{1cm} (3.76)

The rISAR function has a straightforward interpretation; it is the expected phylogenetic distance between the focal species and an arbitrarily chosen species from a neighborhood with radius $r$. We can calculate the expectation $\Delta_f^p$ of rISAR$_f(r)$ for very large neighborhoods $r$. If the neighborhoods approach the size of the plot, the $D_{fm}(r)$ has an approximate value of 1 (i.e., species $m$ is present in the plot) and we find

$$\Delta_f^p = \frac{\sum_{m=1}^{S} d(f,m)}{S - 1}$$  \hspace{1cm} (3.77)

Thus, rISAR$_f(r)$ will asymptotically approach the mean phylogenetic distance $\Delta_f^p$ of the focal species to all other species in the observation window. Note that the index $\Delta_f^p$ is analogous to the index $\Delta^p$ in Hardy and Senterre (2007) that measures phylogenetic distinctness based on species incidence within a given community. However, we restrict $\Delta_f^p$ to comparisons of the focal species $f$ with all other species $m$ present in the plot (i.e., we do not average over species $f$), similar to the manner in which we handled the ISAR. If the rISAR function is larger than $\Delta_f^p$ in small neighborhoods $r$, the focal species is locally surrounded by a subset of phylogenetically more dissimilar species (compared to all species present in the plot), and if the rISAR is less than $\Delta_f^p$ the focal species is surrounded by a subset of phylogenetically more similar species.

Analogous to the approach used with ISAR, we can define an underlying “landscape of phylogenetic neighborhood dissimilarity” for the rISAR function, which measures, at each location $x$ of the observation window, the mean pairwise phylogenetic distance $\Delta_f^p(x,r)$ between the focal species $f$ and all species located within a given distance $r$. Figure 3.24a shows an example of such a landscape for the species *Gustavia superba* in the BCI forest dynamics plot (Box 2.6). For some areas of the plot, we find that the phylogenetic neighborhoods are more dissimilar than expected (yellow to red areas), given the over all phylogenetic dissimilarity $\Delta_f^p$ of the focal species to all species.
However, there are also areas where the phylogenetic neighborhoods are more similar than expected (green to blue areas in Figure 3.24a). Figure 3.24b shows the same landscape, but now derived for neighborhoods with radius $r = 10$ m. The overall pattern of neighborhood dissimilarity is maintained, but additional small-scale structures appear. Note that the landscape of phylogenetic neighborhood dissimilarity will be different for each focal species.

The rISAR function can be used in different ways to explore whether the location of an individual species is correlated with the landscape $\Delta^p_I(x, r)$ of phylogenetic neighborhood dissimilarity; that is, revealing spatial structures in the phylogenetic relatedness of focal species to the other species in the community. For example, we can use the homogeneous Poisson null model, which relocates the individuals of the focal species to random positions present in the plot.
within the plot, to reveal whether a focal species is surrounded locally by species that are phylogenetically more similar or dissimilar, compared with what is found on average within the entire plot. An even better null model might additionally maintain the observed spatial autocorrelation structure by using pattern reconstruction (see Section 4.3.3).

Plotting the locations of *G. superba* individuals within the landscape of phylogenetic neighborhood dissimilarities [i.e., the $\Delta_P^f(x,r)$] indicates that this species may show significant association with areas surrounded by phylogenetically more similar species (Figure 3.24b). Application of the null model shows that the observed $r\text{ISAR}(r)$ is below the expectations of the null model at all spatial scales, which indicates a highly significant effect. This suggests that the distribution pattern of the species *G. superba* could be influenced by habitat filtering of phylogenetically similar species (Figure 3.25). This species dominates ca. 2 ha of forest (4% of the total area) located at the north-western part of the plot [centered on coordinate (700, 450)]. This area was cleared during the nineteenth century and is occupied by species that are phylogenetically similar to *G. superba*. *G. superba* is also surrounded by more similar local species assemblages in other parts of the plot (Figure 3.24b).

### 3.1.6 Summary Statistics for Qualitatively Marked Point Patterns

Qualitatively marked patterns carry a qualitative (discrete or categorical) mark that is descriptive of their state, such as surviving versus dead trees of a given species. Importantly, the qualitative mark is produced by a process acting *a posteriori* over a given univariate (unmarked) pattern. Examples of
qualitatively marked patterns are infected vs. noninfected plants or surviving vs. dead trees. This represents a fundamental difference to bivariate patterns in which the qualitative mark distinguishes among points where the marks were produced \textit{a priori}, a good example of which is given by points whose marks represent different species (Goreaud and Pélissier 2004).

The basic interest in the analysis of qualitatively marked patterns is to determine whether the process that distributed the marks among points acted in a spatially uncorrelated way. To this end, several test statistics of qualitatively marked patterns are compared with those of the null model of \textit{random labeling}, where the qualitative mark is randomly shuffled over the points of the univariate (unmarked) pattern.

In the following, we present summary statistics for two data types of qualitatively marked patterns that are relevant in ecological applications. The first data type is the standard situation, where a univariate pattern is qualitatively marked. The second data type also comprises a qualitatively marked pattern, but here the focus is on exploring the influence of an additional focal pattern on the marking. We have thus a bivariate pattern where the second pattern carries an additional qualitative mark. In this case, random labeling of the qualitatively marked subpattern provides the appropriate null model. Since this data type carries three types of points, it is referred to as \textit{trivariate random labeling}. For these types of marked point pattern, we present \textit{mark connection functions} as the appropriate adapted summary statistics.

3.1.6.1 Random Labeling (Data Type 4)

The simplest data structure with qualitative marks is a univariate pattern that carries a qualitative mark, for example, trees of a given species that are surviving versus dead. Because this data type consists of two types of points, similar in essence to a bivariate pattern, bivariate summary statistics, such as the pair-correlation function or Ripley’s $K$-function, can be used as test statistics to characterize the spatial correlation structure of the marks. However, the bivariate summary statistics still contain the signal of the fixed spatial structure of the underlying univariate pattern (e.g., both dead and surviving trees), which is often not of primary interest. Mark connection functions $p_{lm}(r)$ are designed with this in mind, as they factor out the signature of the underlying univariate pattern.

Intuitively, the idea behind the mark connection function $p_{lm}(r)$ is that it represents the conditional probability that a pair of points, picked at random and separated by distance $r$, consists of a focal point of type $l$ and a second point of type $m$. Thus, $p_{lm}$ is symmetric and $p_{lm}(r) = p_{ml}(r)$. We can also define $p_{mm}(r)$ as the conditional probability that both points are of type $m$. It follows that the sum of all mark connection functions yields one. In the case of the two types $l$ and $m$, we have $p_{ll}(r) + p_{lm}(r) + p_{ml}(r) + p_{mm}(r) = 1$. 
A mark connection function can be expressed in terms of product densities or pair-correlation functions (Illian et al. 2008, p. 331):

\[
p_{lm}(r) = \frac{\rho_m(r)}{\rho(r)} = \frac{\lambda_l \lambda_m g_{lm}(r)}{\lambda^2 g(r)} = p_{lm} g_{lm}(r) / g(r). \tag{3.78}
\]

for \( r > 0 \) and \( \rho(r) > 0 \), with \( p_l = \lambda_l / \lambda \) being the proportion of points having mark \( l \). The denominators \( \rho(r) \) and \( g(r) \) are, respectively, the univariate product density and pair-correlation function applied to the over all pattern (i.e., points of both mark types). The numerators are the partial functions for points with mark \( l \) and \( m \).

On the basis of Equation 3.78, the mark connection functions are best estimated using the ratio of the product densities (Equation 3.51):

\[
\hat{p}_{lm}(r) = \frac{\hat{\rho}_{lm}(r)}{\hat{\rho}(r)} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n, \neq} C_{lm}(x_i, x_j) \times k(\|x_i - x_j\| - r) \times w_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{n, \neq} k(\|x_i - x_j\| - r) \times w_{ij}} \tag{3.79}
\]

The indicator function \( C_{lm}(x_i, x_j) \) yields one if the focal point \( i \) is of type \( l \) and the second point \( j \) of type \( m \), and zero otherwise, and the \( w_{ij} \) are the edge correction weights for the different estimators derived in Section 3.1.2.1 and shown in Equation 3.21. Because the Ohser edge correction weights do not depend on the individual points \( i \) and \( j \), they cancel out and the corresponding estimator

\[
\hat{p}_{lm}(r) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n, \neq} C_{lm}(x_i, x_j) \times k(\|x_i - x_j\| - r)}{\sum_{i=1}^{n} \sum_{j=1}^{n, \neq} k(\|x_i - x_j\| - r)} \tag{3.80}
\]

can be interpreted as the mean value of the “test function” \( C_{lm}(x, x) \) over all pairs of points \( i \) and \( j \), which are separated approximately by distance \( r \) [the latter is captured by the kernel function \( k() \)]. The test function \( C_{lm}(x, x) \) yields a value of 1 if point \( i \) is of type \( l \) and point \( j \) of type \( m \) and zero otherwise. Thus, the estimator of \( p_{lm}(r) \) shown in Equation 3.80 gives the proportion of point pairs separated by distance \( r \) where the first point is of type \( l \) and the second point is of type \( m \). Equation 3.80 also shows that the estimator of the spatially explicit Simpson index (Equation 3.70) can be interpreted as a multivariate mark connection function, with a test function \( 1(m_i = m_j) \) that yields one if the species identifier of point \( i \) (\( = m_i \)) is the same as the species identifier of point \( j \) (\( = m_j \)) and zero otherwise.

Figure 3.26 provides an example for the application of mark connection functions. We used for this purpose the pattern of small surviving and dead
individuals (dbh < 10cm) of the species *Shorea congestiflora* at the Sinharaja World Heritage Site (Sri Lanka) (Figure 3.26a). Although the pattern of small trees is strongly clustered, the empirical mark connection functions are in good agreement with the expectations of the random mortality hypothesis (solid horizontal lines in Figure 3.26b). This result suggests that mortality of small trees of *S. congestiflora* is not subject to density dependent effects as for example shown in Figure 2.15.

### 3.1.6.2 Trivariate Random Labeling (Data Type 5)

Trivariate random labeling explores the effect of an antecedent focal pattern $f$ on the process that distributes a qualitative mark (type $l$ and type $m$) over a second pattern. Based on this data structure we can explore whether the qualitative mark of the second pattern depends on the distance from a point of the antecedent pattern $f$. For example, fire-induced mortality of a shrub species 2 may depend on proximity to individuals of a shrub species 1 that are readily killed by fire (Biganzoli et al. 2009). Killing its neighbor could be an evolved strategy of (focal) species 1 to escape from competition with nearby shrubs of species 2.

We thus have points of a focal pattern (subscript $f$), and type $l$ and type $m$ points of a qualitatively marked pattern. The appropriate summary statistic for this data structure selects pairs of points separated by distance $r$, where the first point is from the focal pattern (i.e., type $f$) and the second point is...
from the second qualitatively marked pattern (i.e., type $l$ or $m$), and calculates the probability that the second point is of type $l$:

$$
 p_{fl}(r) = \frac{\lambda_l}{(\lambda_l + \lambda_m)} g_{fl}(r) \frac{\lambda_f \lambda_l}{\lambda_f (\lambda_l + \lambda_m)} g_{f,l+m}(r)
$$

(3.81)

Here, $(\lambda_l + \lambda_m)$ and $\lambda_l$ are the intensities of the points of the qualitatively marked pattern and its type $l$ points, respectively, and $g_{f,l+m}(r)$ and $g_{fl}(r)$ are the corresponding bivariate, pair-correlation functions. Note that the intensity $\lambda_l$ of the focal pattern cancels out; however, if we leave $\lambda_l$ in Equation 3.81, we see that the summary statistic is analogous to a mark connection function:

$$
 p_{fl}(r) = \frac{\rho_{fl}(r)}{\rho_{f,l+m}(r)}
$$

(3.82)

which can be calculated as the ratio of two bivariate product densities.

An estimator of the summary statistic for trivariate random labeling is given by

$$
 \hat{p}_{fl}(r) = \frac{\hat{\rho}_{fl}(r)}{\hat{\rho}_{f,l+m}(r)} = \frac{\sum_{i=1}^{n_f} \sum_{j=1}^{n_{i+m}} C_i(x_j) \times k(||x_i - x_j|| - r) \times w_{ij}}{\sum_{i=1}^{n_f} \sum_{j=1}^{n_{i+m}} k(||x_i - x_j|| - r) \times w_{ij}}
$$

(3.83)

where $n_f$ is the number of points of the focal pattern and $n_{i+m}$ the number of points of the qualitatively marked pattern. The indicator function $C_i(x_j)$ yields a value of 1 if the second point $j$ is type $l$ from the qualitatively marked pattern. Given these definitions, we can see that the numerator counts all type $f-l$ point pairs at distance $r$ and the denominator counts all $f-l$ and $f-m$ point pairs at distance $r$. Thus, this summary statistic yields the proportion of type $l$ points at distance $r$ from type $f$ points. In our example above, it therefore produces the proportion of burned shrubs of species 2 at distance $r$ from shrubs of species 1.

Figure 3.27 provides an example of trivariate random labeling. We used a clustered pattern with 500 individuals and generated 200 dead individuals (pattern 1) and 300 surviving individuals (pattern 2) in such a way that the probability of mortality was dependent on the number of points of a third (focal) pattern within 40 m (shown in Figure 3.27a as open squares). Analysis with the mark connection functions shows some indication of nonrandom mortality with small-scale clustering of dead individuals, and some repulsion between surviving and dead individuals.
However, the effects are difficult to interpret. Clearly, this is because mortality was not density dependent, but dependent on points of a focal pattern, which are not considered in this analysis (open squares in Figure 3.27a). The dependence in the previous example can be revealed by employing trivariate random labeling (Figure 3.27c). We used the summary statistic $p_f(r)$, which gives the probability of mortality $p_f(r)$ for an individual, which is dependent on the distance $r$ to points of the focal pattern. We then compare this summary statistic to simulation envelopes arising from random mortality (i.e., random permutation of labels 1 and 2). As expected, the probability of mortality is substantially higher than expected at random, if a focal point is located at a distance less than 40 m from the type 1 or 2 individuals. The aggregation of dead individuals at small distances can be explained by collective mortality of a cluster, if one or two focal points were close by. Proximity or nonproximity of a focal point, therefore, produced dead or surviving clusters, respectively, and imprinted a signal of segregation between surviving and dead individuals on the overall pattern.

(Figure 3.27b).
3.1.7 Summary Statistics for Quantitatively Marked Point Patterns

The marks carried by a univariate pattern can be quantitative, as well as qualitative. For example, the trees of a given species may be characterized by their diameter at breast height (dbh) or height. What is generally of interest in analyzing quantitative marks is to find out whether the marks show some sort of spatial correlation, conditional on the locations of the corresponding unmarked pattern. The summary statistics adapted to quantitative marks are “mark-correlation functions,” a concept introduced by Stoyan (1984). Mark-correlation functions are only rarely used in the ecological literature and then mostly in their simplest form, where one quantitative mark is attached to a univariate pattern. However, a plethora of data types are possible for quantitative marks (see Section 2.2.3), since one or more quantitative marks may be attached to bivariate and qualitatively marked patterns (Figure 2.3, data types 6–9).

Here, we present mark-correlation functions for univariate, quantitatively marked patterns, as developed by Illian et al. (2008, their Section 5.3.3), and analogous functions for patterns with two marks. The latter includes cases where two quantitative marks are attached to a univariate pattern (data type 7 in Figure 2.3), cases in which the pattern carries one qualitative and one quantitative mark (data type 8 in Figure 2.3), and cases where a bivariate pattern carries one quantitative mark (data type 9 in Figure 2.3). An interesting new perspective in the application of mark-correlation functions is provided by analyses of spatial structures in traits or phylogenetic spatial structures in fully mapped communities. For example, a plant trait such as leaf area, wood density, or maximal height can be defined as a quantitative mark, and mark-correlation functions can then be used to reveal spatial structures in the trait distribution within communities. Similarly, the phylogenetic distance between two individuals in a community can be used as a test function and phylogenetic, mark-correlation functions can be used to calculate the mean phylogenetic distance between two individuals of the community that are separated by distance $r$ (Section 3.1.7.6).

3.1.7.1 Univariate Quantitatively Marked Pattern (Data Type 6)

This is the simplest case of a quantitatively marked pattern, where one quantitative mark (e.g., size) is attached to a univariate pattern (e.g., trees of a given species). Analogous to the use of mark connection functions for qualitative marks, mark-correlation functions are used to analyze the spatial relationships among points containing quantitative marks. Point $i$ at location $x_i$ has mark $m_i$, where $m_i$ is usually a positive, real number. The basic goal is to determine whether the joint properties of the marks of two points depend on the distance $r$ separating them. Univariate mark-correlation functions are the summary statistics adapted to this data structure, and the empirical
summary statistics are compared to those arising from independent marking, which represents the fundamental division related to the absence of any spatial structure in the marks.

The basic idea of mark-correlation functions is analogous to that of the interpretation of the mark connection function in Equation 3.79 as described above: they are the mean value of a test function of the marks of all point pairs \( i \) and \( j \) that are separated approximately by distance \( r \) (Illian et al. 2008, p. 341). Thus, the first step in developing mark-correlation functions is to identify appropriate test functions that characterize some relationship between the marks \( m_i \) and \( m_j \) of two points \( i \) and \( j \). One example of a test function is simply to use the value of the mark of the second point \( \bar{m} = m_j \). The associated mark-correlation function thus yields the mean mark of a point, given that this point has a neighbor at distance \( r \). Note that this conditional mean may be quite different from the (unconditional) mean mark.

Intuitively, an estimator of the mark-correlation functions visits all pairs of points separated by distance \( r \) and calculates the mean value of the test function over these pairs. This is then repeated for different distances \( r \). Thus, analogous to the estimator of the mark connection function (Equation 3.79), we find

\[
\hat{c}_i(r) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} t(m_i, m_j) \times k(\|x_i - x_j\| - r) \times w_{i,j}}{\sum_{i=1}^{n} \sum_{j=1}^{n} k(\|x_i - x_j\| - r) \times w_{i,j}}
\]

(3.84)

where \( t(m_i, m_j) \) is the test function that uses the mark \( m_i \) of point \( i \) and the mark \( m_j \) of point \( j \), \( k() \) is the kernel function that defines which points are located approximately at distance \( r \) and \( w_{i,j} \) is the edge correction. This estimator is based on the analogous second-order product density for quantitatively marked patterns (Illian et al. 2008, p. 354 their Equation 5.3.5.3):

\[
\hat{\rho}_i(r) = \frac{1}{2\pi r A} \sum_{i=1}^{n} \sum_{j=1}^{n} t(m_i, m_j) \times k(\|x_i - x_j\| - r) \times w_{i,j}
\]

(3.85)

where the inner sum \( \sum_{j=1}^{n} t(m_i, m_j) \times k(\|x_i - x_j\| - r) \times w_{i,j} \) sums up all points \( j \) located at distance \( r \) from the focal point \( i \), and weights them with the test function \( t(m_i, m_j) \). Thus, the second-order product density \( \rho_i(r) \) for a pattern with quantitative marks contains information on both the spatial structure of the pattern and the correlation structure of the marks. As a parallel to the development of the mark connection functions (Equation 3.79), we factor out the conditional spatial structure of the points contained in the product density and define the (nonnormalized) mark-correlation function \( c_i(r) \) based on test function \( t \) as
\[ \hat{c}_i(r) = \frac{\hat{\rho}_i(r)}{\hat{\rho}(r)}, \quad (3.86) \]

where \( \hat{\rho}_i(r) \) is the second-order product density for a pattern with quantitative marks and test function \( t \) and \( \hat{\rho}(r) \) is the second-order product density for the corresponding univariate pattern. It can be easily verified that this definition yields the estimator given in Equation 3.84.

If there is no spatial correlation among the marks, the nonnormalized, mark-correlation function \( c_i(r) \) yields the average \( c_i \) of the test function \( t(m_i, m_j) \) over all pairs of points \( i \) and \( j \) of the pattern, that is, the nonspatial mean of the test function. It is, therefore, useful to normalize the \( c_i(r) \) with its asymptotic value \( c_i \) to make mark-correlation functions \( k_i(r) \) independent of the distribution and values of the marks:

\[ \hat{k}_i(r) = \frac{1}{c_i} \frac{\hat{\rho}_i(r)}{\hat{\rho}(r)}. \quad (3.87) \]

The subscript \( t \) in Equation 3.87 refers to the particular test function \( t \) used. Depending on the test function \( t \), different mark-correlation functions \( k_i(r) \) arise. Important examples of test functions used to construct (univariate) mark-correlation function, as pointed out in Illian et al. (2008, p. 343), are

- \( k_{mn}(r): t_1(m_i, m_j) = m_i m_j \)
- \( k_{m}(r): t_2(m_i, m_j) = m_i \)
- \( k_{m}(r): t_3(m_i, m_j) = m_j \)
- \( \gamma_m(r): t_4(m_i, m_j) = (m_i - m_j)^2/2 \)
- \( I_m(r): t_5(m_i, m_j) = (m_i - \mu)(m_j - \mu) \)
- \( I_m(r): t_6(r, m_i, m_j) = [m_i - \mu(r)](m_j - \mu(r)] \quad (3.88) \)

where \( m_i \) and \( m_j \) are the marks of the two points \( i \) and \( j \), \( \mu \) is the mean mark over all points in the pattern, and \( \mu(r) \) is the mean mark of points pairs separated by distance \( r \). Note that test functions \( t_1(m_i, m_j) \) to \( t_5(m_i, m_j) \) are defined for any two points \( i \) and \( j \), whereas, for reasons to be explained later, the test function \( t_6(r, m_i, m_j) \) is only defined for points that are separated by distance \( r \). The normalizing factors of the different test functions given in Equation 3.88 yield

\[ c_1 = \mu^2 \]
\[ c_2 = \mu \]
Estimators and Toolbox

\[ c_3 = \mu \]  
\[ c_4 = \sigma_{\mu}^2 \]  
\[ c_5 = \sigma_{\mu}^2 \]  
\[ c_6 = \sigma_{\mu}^2 \]

where \( \mu \) is the mean mark over all points in the pattern and \( \sigma_{\mu}^2 \) is the mark variance.

The simplest test functions, \( t_2 \) and \( t_3 \), result in mark-correlation functions that yield the mean mark of a point that has another point at distance \( r \). These summary statistics are called “r-mark-correlation functions” \( k_m(r) \) and \( k_m(r) \), respectively. Note that the point in the subscripts indicate that the marks of the second point (= \( m_i \)) and the first point (= \( m_i \)) not used for estimation of the r-mark-correlation functions \( k_m(r) \) and \( k_m(r) \), respectively (see Equations 3.84 and 3.88). They describe a basic statistical property of the spatial correlation structure between the marks of two points separated by distance \( r \), that is, their (conditional) mean value. The basic interests here are to find out how the conditional mean mark depends on distance \( r \) and if it departs significantly from the mean mark \( \mu \).

Figure 3.28b provides examples for the r-mark-correlation function \( k_m(r) \). The constructed marks of the pattern shown in Figure 3.28a are the inverse of the number of neighbors within distance 10 m (including the focal point), thus isolated points have larger marks and points occurring in clusters have smaller marks. The marks in Figure 3.29a are the size of the trees of the species of Shorea congestiflora, a dominant species within a 25-ha plot in a rain forest at Sinharaja World Heritage Site (Sri Lanka). The r-mark-correlation function for both patterns shows a significantly negative departure from the null model at short distances \( r \), indicating that nearby trees are smaller on average than trees selected at random. This was expected for the pattern shown in Figure 3.28a due to the way it was constructed. For the pattern of S. congestiflora, the significant effect was caused by small trees occurring together in clumps. However, Figure 3.29a shows that larger trees are often close to clusters of smaller trees. For this reason, the response of the r-mark-correlation function (Figure 3.29b) is weaker than for the constructed marks shown in Figure 3.28b.

Schlather et al. (2004) proposed the test function \( t_6 \) and its corresponding summary statistic to investigate how the marks of two points separated by distance \( r \) differ from their conditional mean value \( \mu(r) \). This results in a Moran’s I type summary statistic \( I_m(r) \), which is a spatial variant of the classical Pearson correlation coefficient (Shimatani 2002). It characterizes the covariance, a second basic property of the spatial correlation structure between the marks of two points separated by distance \( r \). The basic interest here is to determine whether the marks show a spatial correlation, for example, individuals may be smaller than average if they are located close to a neighbor.
and larger than average if they are located farther away from a neighbor. Note that the test function $t_5$, which is similar to $t_6$, adjusts for $\mu$, the population mean, not the conditional mean $\mu(r)$, as required for a correlation coefficient. Figure 3.28c shows that the two variants of $I_m(r)$, based on $t_5$ and $t_6$, differ a great deal for the constructed marks. The reason for this is the strong effect of the $r$-mark-correlation function. Test function $t_5$ compares the marks $m_1$ and $m_2$ to the over all mean mark $\mu$, whereas test function $t_6$ compares to the conditional-mean mark $\mu(r)$. Or, in other words, the mark-correlation function resulting from the Schlather test function $t_6$ looks only at pairs of points separated by distance $r$ and compares the marks $m_1$ and $m_2$ for each point pair with the actual mean mark $\mu(r)$ resulting from all pairs of points separated by distance $r$. Clearly, test functions $t_5$ and $t_6$ will yield similar results, as shown in Figure 3.29e, if the conditional mean $\mu(r)$ agrees with the over all mean $\mu$ (this is tested by the $r$-mark-correlation function; Figure 3.29b). However, if $\mu(r)$ differs strongly at smaller distances $r$ from the mean $\mu$ mark, as shown in Figure 3.28b, the resulting $I_m(r)$ summary statistics differ as well at these distances (Figure 3.28c). Thus, the Schlather test function $t_6$ must be selected if we want to obtain a mark-correlation function with an interpretation as a Moran’s $I_m(r)$ summary statistic, whereas the test function $t_5$ may indicate spurious correlation caused by departures from the mean $\mu(r)$.
Estimators and Toolbox

that are not due to correlations in the marks of points separated by distance $r$ (i.e., Figure 3.28c).

Test function $t_4$ characterizes the difference between the marks of point pairs as a function of distance $r$ and yields what is called the mark variogram $\gamma_m(r)$. Small values in the mark variogram indicate similarity in magnitude between points separated by distance $r$. Figure 3.29d shows that the sizes of two nearby $S. congestiflora$ trees tend to be more similar than expected compared to two trees taken at random. Finally, test function $t_5$ yields the mean product of the two marks of points separated by distance $r$ and the corresponding summary statistic is called the mark-correlation function $k_{mm}(r)$. It is interesting to note that the mark-correlation function $k_{mm}(r)$ does not show
a significant effect for the species \textit{S. congestiflora} (Figure 3.29c). Because the \textit{r}-mark-correlation function shows a significant effect, this means that we may often have a large tree close to a small tree, which “neutralizes” the mark product. Thus, the mark-correlation function may miss out on important effects shown by the \textit{r}-mark-correlation function and can only be properly interpreted in conjunction with the \textit{r}-mark-correlation function.

The second-order product density for quantitatively marked patterns given in Equation 3.85 can also be interpreted in a way similar to the pair-correlation function. To this end the quantity

\[ \hat{g}_t(r) = \frac{1}{\lambda^2 c_t} \hat{\rho}(r) \]  

is considered, where the subscript \( t \) refers to the test function \( t \) used. Note that normalization by the mean of the test function \( c_t \) is necessary to yield a value of 1, when a spatial pattern in the points and the marks is absent.

The test function \( t_1(m_i, m_j) = m_i m_j \) is of particular interest, as it yields the multiplicatively weighted, pair-correlation function \( g_{mm}(r) \):

\[ \hat{g}_{mm}(r) = \frac{1}{\lambda^2 c_t} \frac{1}{2\pi r A} \sum_{i=1}^{n} \sum_{j=1}^{n} k(\| x_i - x_j \|- r) \times (m_i m_j) \times w_{i,j} \]  

This summary statistic characterizes the spatial distribution of the “mark mass” rather than the point distribution (Illian et al. 2008, p. 348). The quantity \( g_{mm}(r) \) basically acts the same as the pair-correlation function \( g(r) \), but weights each point pair additionally by its mark (i.e., the factor \( m_i m_j \) in Equation 3.91).

If the marks show no spatial correlation, the multiplicatively weighted, pair-correlation function \( g_{mm}(r) \) collapses to the pair-correlation function, and if the univariate pattern is a random pattern \( g_{mm}(r) \) collapses to the mark-correlation function. The multiplicatively weighted, pair-correlation function \( g_{mm}(r) \), therefore, contains signals from potential correlations in the marks and from potential correlations in the spatial distribution of the points. Note that \( g_{mm}(r) \) is related to the pair-correlation function derived for objects with finite size and real shape (see Section 3.18).

An interesting example for the application of the multiplicatively weighted pair-correlation function \( g_{mm}(r) \) was given in Law et al. (2009). They analyzed the distribution of biomass over space in a 1 ha plot of temperate forest at Rothwald, Austria, to assess the extent to which biomass is decoupled from the spatial pattern of tree locations. The pair-correlation functions revealed aggregation of the tree locations in the forest (Law et al. 2009; Figure 3). They investigated the relationship between the spatial pattern of biomass and trees using the multiplicatively weighted pair-correlation function. For this analysis, they used allometric relationships to convert tree height into biomass,
which was then used as a quantitative mark. Comparison of the results of the pair-correlation function \( g(r) \) of all trees, the multiplicatively weighted pair-correlation function \( g_{mw}(r) \), and the mark-correlation function \( k_{mm}(r) \) provided interesting insights into the spatial structure of these forests. While \( g(r) \) indicated aggregation of trees up to 20 m, \( g_{mw}(r) \) yielded values close to 1, indicating a near random spatial distribution of biomass. In contrast, \( k_{mm}(r) \) showed that trees situated close to one another were characterized by below average biomasses (Law et al. 2009; Figure 4). Thus, the spatial pattern of biomass was substantially different from the aggregated pattern of tree locations. A more uniform distribution of biomass (compared to the clustered tree distribution) was reached by closely spaced trees having below average biomass. This compensatory effect was visible up to about 20 m.

### 3.1.7.2 Univariate Marked K-Functions (Data Type 6)

Cumulative mark-correlation functions can be derived for any test function \( t \) as a natural generalization of Ripley’s \( K \)-functions (Illian et al. 2008, pp. 350–352). Recall that the \( K \)-function can be estimated as

\[
\hat{K}(r) = \frac{1}{\lambda^2} \frac{1}{A} \sum_{i=1}^{n} \sum_{j=1}^{n, j} \mathbf{1}([x_i - x_j]_r) \times w_{ij}
\]

where the count function \( \mathbf{1}(d, r) \) has a value of 1, if point \( j \) is located within distance \( r \) of point \( i \), and zero otherwise. As usual, \( w_{ij} \) is the edge correction term. The cumulative mark product function \( K_t(r) \) (Shimatani and Kubota 2004), based on a test function \( t \), can be estimated as

\[
\hat{K}_t(r) = \frac{1}{\lambda^2} \frac{1}{A} \frac{1}{\tilde{c}_i} \sum_{i=1}^{n} \sum_{j=1}^{n, j} t(m_i, m_j) \times \mathbf{1}([x_i - x_j]_r) \times w_{ij}.
\]

This approach basically consists of counting all point pairs \((i, j)\) separated by a distance less than \( r \), weighted by the value of the test function \( t(m_i, m_j) \) of the respective marks \( m_i \) and \( m_j \). Thus, mark-weighted \( K \)-functions are the normalized means of the sum of values of the test function \( t \) formed by the mark of the typical point and all points in the disk of radius \( r \) centered at the typical point. Because of the accumulative nature of the \( K \)-functions we find that

\[
K'_t(r) = 2\pi rk_t(r)g(r)
\]

where \( K'_t(r) \) is the derivative of \( K_t(r) \) with respect to \( r \) (Equation 5.3.40 in Illian et al. 2008). Thus, the mark weighted \( K \)-functions depend on both the spatial structure of the unmarked pattern [represented by the pair-correlation function \( g(r) \) in Equation 3.94] and the correlation structure of the marks [represented by the mark-correlation function \( k_t(r) \)].
Finally, it might be desirable to remove the signal of the univariate pattern from the cumulative mark-product function $K(r)$ to obtain a cumulative mark-correlation function. This can be done by estimating the quantity

$$
\hat{\kappa}_i^{\text{cum}}(r) = \frac{1}{\hat{c}_i} \frac{\hat{\kappa}_i^{(2)}(r)}{\hat{K}_i^{(2)}(r)} = \frac{1}{\hat{c}_i} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} l(x_i, m_i) \times 1(||x_i - x_j||, r) \times w_{i,j}}{\sum_{i=1}^{n} \sum_{j=1}^{n} 1(||x_i - x_j||, r) \times w_{i,j}}.
$$

which is analogous to the approach taken in Equations 3.84 and 3.87. The quantity $\hat{\kappa}_i^{\text{cum}}(r)$ has a simple interpretation: it is the mean value of the test function for points pairs $i$ and $j$ that are located within distance $r$ of each other.

Figure 3.30 compares the cumulative mark-correlation functions with their noncumulative counterparts for the pattern shown in Figure 3.29a. Clearly, the cumulative nature of the $k_i^{\text{cum}}(r)$ glosses over the details of the spatial patterns shown in their noncumulative counterpart $k_i(r)$, and the scales of significant departures from random are often different. The cumulative functions $k_i^{\text{cum}}(r)$ are also more likely to show departures from the null model, while the noncumulative $k_i(r)$ shows, in general, scale-dependent effects much more clearly. This is because the cumulative functions are based on a higher number of point pairs, which reduces the effects of stochastic noise. The question as to whether or not to use the noncumulative or the cumulative functions ultimately depends on the biological question: is the contribution of all individuals within the neighborhood $r$ important or is the goal to reveal specific scale effects among individuals separated by a certain distance $r$?

### 3.1.7.3 Two Quantitative Marks Attached to a Univariate Pattern (Date Type 7)

The first data structure that yields bivariate mark-correlation functions is a univariate pattern with two quantitative marks $m_1$ and $m_2$ (data type 7 in Figure 2.3). An example for this data type is a pattern of trees that hosts two species (or groups of species) of orchids, with the marks representing the number of orchids of each species (or group) on a given tree (Raventós et al. 2011). Stoyan (1987) used another example in which the points were the locations of pine saplings and the marks were their heights and age. The basic interest in analyzing these types of data is to find out whether the two marks show some spatial correlation that depends on the distance $r$ between points. This can be tested by comparing the empirical summary statistics to simulation envelopes arising from a null model that represents the absence of any spatial structure in the two marks.

Depending on the ecological question, different null models need to be used. For example, if the marks are the number of orchids of two species, we may ask whether they are independently distributed over the host trees. In this case, we can condition on the number of orchids of the first species and...
shuffle the second mark (i.e., number of individuals of the second orchid species) randomly over the trees of the univariate pattern. However, if we analyze the spatial correlation in marks representing height and dbh of trees, we cannot separate height and dbh, since these are most likely correlated with one another. In this case, we shuffle the vector of marks of the individual trees $i$, given by $(m_{i1}, m_{i2})$, randomly over the trees of the univariate pattern.

An application of this null model that studied spatial structures in epiphytic orchids caused by hurricane damage is given in Wiegand et al. (2013b). The number of affected orchids of a given species that were located at a host tree

---

**FIGURE 3.30**
Cumulative mark-correlation functions vs. noncumulative mark-correlation functions based on the example pattern of the species *Shorea congestiflora* a shown in Figure 3.29a. Left column: cumulative mark-correlation functions. Right column: corresponding noncumulative mark-correlation functions.
was mark $m_1$, and the corresponding number of non-affected orchids was mark $m_2$. They used the bivariate mark variogram to test if the number of affected orchids at the focal tree and the number of non-affected orchids at nearby trees tended to be relatively similar or dissimilar compared to those of pairs of trees taken at random.

Summary statistics for the univariate case can be easily modified to perform an analysis of data structures in which two quantitative marks are attached to a univariate pattern. In this case, the mark $m_{1i}$ will refer to the mark of a focal point $i$ and the mark $m_{2j}$ to a second mark for point $j$, which is located at a distance $r$ from the focal point $i$. The estimator of the nonnormalized mark-correlation functions therefore yields

$$
\hat{c}_i(r) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n,x} t(m_{1i}, m_{2j}) \times k(\|x_i - x_j\| - r) \times w_{i,j}}{\sum_{i=1}^{n} \sum_{j=1}^{n,x} k(\|x_i - x_j\| - r) \times w_{i,j}}
$$

(3.96)

The test function $t_1$ yields a bivariate mark-correlation function $k_{m1m2}(r)$, which returns the mean product of the marks $m_{1i}$ and $m_{2j}$. The normalization factor for this test statistic is given by $c_1 = \mu_1 \mu_2$, where $\mu_1$ and $\mu_2$ are the mean of the first and second mark, respectively. The test functions $t_2$ and $t_3$ are not of interest here because they yield the univariate, $r$-mark-correlation functions for the first and second mark $m_1$ and $m_2$, respectively. The test function $t_4$ yields a bivariate mark variogram $\gamma_{m1m2}(r)$ that returns the mean of the squared differences of the two marks separated by distance $r$. However, this test function only makes sense if the values of the two marks are normalized to the same mean, otherwise it will basically depict differences in the numerical values of the two marks. If the marks are normalized, the bivariate mark variogram $\gamma_{m1m2}(r)$ tests whether the two marks of nearby points tend to be relatively similar or dissimilar. The test function $t_4$ returns a bivariate, distance-dependent correlation coefficient of the two marks $m_{1i}$ and $m_{2j}$ separated by a distance $r$. Note that in this case Equation 3.88 must be modified to $t_4(m_{1i}, m_{2j}) = [m_{1i} - \mu_1(r)][m_{2j} - \mu_2(r)]$, where $\mu_i(r)$ and $\mu_j(r)$ are the mean of the first and second mark, respectively, given that points $i$ and $j$ are separated by distance $r$. This Moran’s $I$ type summary statistic $I_{m1m2}(r)$ determines, for example, if the dbh and the height of nearby trees are correlated. The normalization constant is given by $c_{16} = \sigma_{12}$, where $\sigma_{12}$ is the covariance of the two marks.

Figure 3.31 provides an example of an analysis of a univariate pattern that is augmented with two quantitative marks. It is the pattern of the species *S. congestiflora*, as shown in Figure 3.29a, together with the two marks of size ($m_1$) and a “constructed mark” ($m_2$) representing the number of neighbors within 20 m of an individual. The first $r$-mark-correlation function $k_{m1}(r)$ is identical with the univariate mark-correlation function of the pattern for the mark size (cf. Figures 3.29b and 3.31b), and indicates that trees separated by distances up to 40 m are smaller than expected. The second $r$-mark-correlation
FIGURE 3.31
Bivariate mark-correlation function. (a) Pattern of the species Shorea congestiflora a shown in Figure 3.29 together with marks for dbh \((m_1);\) open disks and number of neighbors within 20 m \((m_2);\) gray disks. Note that the second mark has small values in most cases and is barely visible. (b)–(f) Different bivariate mark-correlation functions analyzing spatial structures in the relationship between the size of a tree and the number of neighbors. (g) Relationship between number of neighbors within 20 m and the size of a tree. In panel (d), we include additional information concerning the product of the two \(r\)-mark-correlation functions (gray disks) and, in panel (f), we compare Schlather’s version of Moran’s \(I\) mark-correlation function (black) with the common version (gray). The null model in this analysis shuffled the first mark \(m_1\) (i.e., size) randomly over all trees, whereas the constructed mark \(m_2\) (i.e., the number of neighbors within 20 m) remained unchanged.
The function $k_{m_2}(r)$ exhibits a strong spatial structure in the number of neighbors within 10 m: for distances up to 40 m we find that two trees separated by distance $r$ have more neighbors than expected (this is due to the strong clustering of the pattern), whereas for distances larger than 40 m two trees separated by distance $r$ have fewer neighbors than expected (Figure 3.31c). The latter happens because trees separated by more than 40 m are often “isolated” trees away from clusters. The bivariate mark-correlation function does not reveal much new information. It basically provides the product of the two $r$-mark-correlation functions (cf. closed black and closed gray disks in Figure 3.31d). To produce a meaningful variogram, we normalized the marks with their mean value, thus replacing the mark $m_1$ by $m_1/\mu_1$ and the mark $m_2$ by $m_2/\mu_2$. When doing this we do not find any significant relationship in the mark differences associated with the distance between trees (Figure 3.31e).

However, the Moran’s $I$ type summary statistic $I_{m_1m_2}(r)$ shows a negative correlation between tree sizes and number of neighbors, which indicates that small trees tend to have many neighbors within 20 m.

### 3.1.7.4 One Qualitative and One Quantitative Mark (Data Type 8)

The second data structure that yields bivariate mark-correlation functions is a univariate pattern with one quantitative mark and one qualitative mark (data type 8 in Figure 2.3). An example for this data structure is a pattern of surviving and dead trees of a given species (qualitative mark), where the size of the trees is also known (the quantitative mark). The basic interest in analyzing these types of data is to find out whether the quantitative mark (e.g., size) shows a spatial correlation with the qualitatively marked points (i.e., surviving and dead). For example, we may expect that dead trees located near surviving trees would be smaller than expected given the overall sizes of surviving and dead trees. This question requires a null model where the qualitative marks (e.g., surviving and dead) are fixed, but the size of the quantitative mark is randomly shuffled over all points (e.g., surviving and dead trees). Alternatively, we may randomly reshuffle the qualitative marks (e.g., surviving vs. dead) over all locations and retain the quantitative mark (e.g., size) as fixed.

Essentially, the same test statistics are used for this data structure as with the univariate mark-correlation functions. The qualitative mark distinguishes between type $l$ and type $m$ points, where the focal type $l$ point $i$ carries the mark $m_{il}$ (e.g., size of surviving tree $i$) and a type $m$ point $j$ carries mark $m_{jm}$ (e.g., size of dead trees $j$). The estimator of the nonnormalized mark-correlation functions for this data structure therefore yields

$$
\hat{c}_{lm,l}(r) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n,z} t(m_{il}, m_{jm}) \times C_{lm}(x_i, x_j) \times k(\|x_i - x_j\| - r) \times w_{i,j}}{\sum_{i=1}^{n} \sum_{j=1}^{n,z} C_{lm}(x_i, x_j) \times k(\|x_i - x_j\| - r) \times w_{i,j}}
$$

(3.97)
The indicator function $C_{lm}(x_i, x_j)$ evaluates to a value of 1, if point $i$ is a type $l$ point and point $j$ is a type $m$ point. It is zero otherwise. This summary statistic is thus a mixture between a univariate, mark-correlation function (Equation 3.84) and a mark-connection function (Equation 3.79). However, note that the denominator includes the indicator function $C_{lm}(x_i, x_j)$, which selects only point pairs $i, j$ of type $l$ and $m$, respectively. Thus, the test functions are calculated for pairs of points that fulfill two conditions: the focal point must be of type $l$ and the second point of type $m$ and the points must be separated by distance $r$.

The test function $t(m_l, m_m)$ in the numerator of Equation 3.97 can be taken from the set of functions listed in Equation 3.88 and is applied as follows: The test function $t_1$ yields a bivariate mark-correlation function $k_{m1m2}(r)$, which returns the mean product of the marks of all pairs of type $l$ and $m$ points that are a distance $r$ apart. Because the randomization of this data structure shuffles the marks over the entire unmarked pattern (or randomly shuffles the mark surviving and dead) the normalization constant yields $c_{l1} = \mu_{lm}^2$ where $\mu_{lm}$ is the mean of the marks taken over all points of the underlying univariate pattern (i.e., type $l$ and type $m$ points). The test function $t_2$ yields the mean mark of type $l$ points (e.g., surviving), which are distance $r$ away from a type $m$ point (e.g., dead), whereas $t_3$ yields the mean mark of a type $m$ point (e.g., dead) located distance $r$ from a type $l$ point (surviving). In both of the latter two cases, the normalization constant is equal to $\mu_{lm}$. Note that there is a subtle difference in the univariate and bivariate $r$-mark correlation function $k_{m1}(r)$ when it is based on test function $t_2$. In the univariate case, the condition is that there is a surviving tree at distance $r$ from a surviving tree, whereas in the bivariate case the condition is that there is a dead tree at distance $r$ from a surviving tree.

The test function $t_4$ yields the mean squared difference of the marks of all pairs of type $l$ and type $m$ points, which are a distance $r$ apart. In the tree example given above, incorporating the quantitative mark for size and the qualitative marks of dead or alive, the bivariate-mark variogram measures whether the sizes of neighboring dead and surviving trees are more similar (low values) or dissimilar (high values) than the sizes of a pair of trees taken at random. The test functions $t_5$ and $t_6$ yield a distance-dependent, cross-correlation coefficient of the mark of a pair of points of type $l$ and type $m$. They must be modified such that

$$t_5(m_l, m_m) = [m_l - \mu_l][m_m - \mu_m] \quad \text{(3.98a)}$$

$$t_6(m_l, m_m) = [m_l - \mu_l(r)][m_m - \mu_m(r)] \quad \text{(3.98b)}$$

where $\mu_l$ and $\mu_m$ are the mean marks of points of type $l$ and type $m$, respectively. $\mu_l(r)$ is the mean mark of a type $l$ point (e.g., surviving) that is a distance $r$ away from a type $m$ point (e.g., dead), and $\mu_m(r)$ is the mean mark of a type $m$ point (e.g., dead) that is distance $r$ away from a type $l$ point (e.g., surviving). Thus, $\mu_l(r)$ and $\mu_m(r)$ are the two nonnormalized $r$-mark-correlation...
functions associated with the test function $t_2$ and $t_3$, respectively that arise for this data type. In test function $t_5$ these conditional means are replaced by the over all means $\mu_l$ and $\mu_m$. The normalization constant yields the variance $\sigma_{lm}^2$ of the joint pattern of type $l$ and type $m$ points.

Figure 3.32 provides an example of an analysis of a spatial pattern with one qualitative mark (surviving vs. dead) and one quantitative mark (size) using the species *S. congestiflora*. This example illustrates the exploration of distance-dependent spatial patterns between pairs of surviving and dead trees. The $r$-mark-correlation function $k_{m1}(r)$, which estimates the mean size of surviving trees (that have a dead tree at distance $r$), indicates that the size of surviving trees located at distance $r$ from dead trees is significantly smaller than expected (Figure 3.32b; simulation envelopes are produced by reshuffling the mark size over surviving and dead trees). This, however, is mostly a consequence of the univariate structure of surviving trees, as shown

**FIGURE 3.32**
Mark-correlation function for one qualitative and one quantitative mark. (a) Pattern of surviving (open circles) and dead (closed circles) individuals of the species *Shorea congestiflora*, with a quantitative mark for size ($m_1$), defined as by dbh, being shown by the size of the circle. (b)–(e) Different bivariate mark-correlation functions analyzing spatial structures with respect to size for pairs of surviving and dead trees located at distance $r$. 

© 2014 Taylor & Francis Group, LLC
in previous analyses (e.g., Figure 3.29b). However, note that the univariate $r$-mark correlation function of surviving trees (Figure 3.29b) is different from the bivariate $r$-mark correlation function (Figure 3.32b), because in the first case the condition is that the surviving focal tree has a surviving tree at distance $r$, whereas in the second case the condition is that the surviving focal tree has a dead tree at distance $r$. The $r$-mark-correlation function $k_m(v_2)(r)$ does not show a significant impact of the proximity of surviving trees on the size of dead trees (Figure 3.32c). The mark variogram shows that nearby (<10 m) surviving and dead trees tend to have similar sizes and more distant pairs (>30 m) tend to have more dissimilar sizes (Figure 3.32d). The same result is depicted by the correlation coefficients that show a tendency to more similar sizes at shorter distances and a tendency to negative correlation in size at greater distances (Figure 3.32e).

### 3.1.7.5 Bivariate Pattern with One Quantitative Mark (Data Type 9)

Although the case of a bivariate pattern with one quantitative mark seems to be very similar to the case of one qualitative and one quantitative mark, a fundamental difference arises in the randomization of the data structure for the fundamental division. In the case of one qualitative and one quantitative mark, the null model randomizes the marks over the unmarked pattern (either the qualitative or the quantitative mark), whereas for a bivariate pattern with one quantitative mark, the marks must only be randomized within each pattern (i.e., a point of pattern 1 should not receive a mark of a point of pattern 2 and vice versa). Because the two marks in data type 8 are a posteriori marks (i.e., they characterize an existing univariate pattern), a point of the univariate pattern could theoretically receive each qualitative mark and one value from any of the quantitative marks. However, if we have a bivariate pattern that carries a quantitative mark, the two univariate component patterns are a priori different and therefore we can study potential correlation in their marks only by randomizing the quantitative mark within each component pattern.

We can characterize the bivariate pattern by defining two types of points: type $v$ and type $w$. A point $i$ of type $v$ carries the mark $m_{iv}$ (e.g., size of a tree of species $v$) and a point $j$ of type $w$ carries the mark $m_{iw}$ (e.g., size of tree of species $w$). The estimator of the nonnormalized, mark-correlation functions of this data structure is, therefore, the same as for data type 8 (i.e., Equation 3.97), except that $l$ now stands for type $v$ and $m$ stands for type $w$. However, the normalization constants of the different test functions differ from those of data structure 8, because the randomization of the data structure shuffles the mark only within data types.

For test function $t_1$, the normalization constant is given by $c_{t_1} = \mu_v \mu_w$, where $\mu_v$ and $\mu_w$ are the mean marks for points of type $v$ and $w$, respectively. The test function $t_2$ yields the mean mark of type $v$ points, taken over all pairs that are a distance $r$ apart, where the first point is a type $v$ point and the second a type $w$ point. In this case, the normalization constant is $c_{t_2} = \mu_v$. Note that
the analysis with $t_2$ in this situation is similar to trivariate random labeling (see Section 3.1.6.2), which allowed us to determine how the proximity of a point of the second pattern (i.e., type $w$) influenced the mark of the focal point (i.e., type $v$). Similarly, test function $t_3$ allows us to determine how the proximity of a focal point of type $v$ influences the marks of nearby points of type $w$. Here, the normalization constant is $c_{t_3} = \mu_w^2$. The test function $t_4$ only makes sense if the values of the two marks are normalized to the same mean, otherwise it basically assesses differences in the numerical values of the two marks. The test functions $t_5$ and $t_6$ yield a distance-dependent, cross-correlation coefficient of the marks of a pair of points of type $v$ and type $w$ and is the same as for data type 8 (Equation 3.98). However, in contrast to data type 8, the normalization constant is given by the covariance $\sigma_{vw}^2$ of the marks of the two patterns.

To illustrate the analysis of this data structure we analyzed data from the large trees (dbh $> 10$ cm) of the BCI plot (Box 2.6). The bivariate pattern consists of the focal species *Trichilia pallida* (a canopy species) and the individuals of all the other species, which comprise the second pattern. In all cases, we consider individuals with a dbh greater than 10 cm. The quantitative mark in this case is the size of the trees. The bivariate mark-correlation functions, therefore, relate the size of the focal species to the size of individuals of all other species that are located at distance $r$ away from the focal individual. In this example, we are interested in finding out whether there is a correlation between the sizes of the individuals of the focal species *T. pallida* and those of nearby heterospecific trees. We, therefore, randomize the mark size within the focal species, but keep the marks of all heterospecific trees unchanged. This null model assumes that the sizes of *T. pallida* trees are independent of the sizes of nearby heterospecific trees. Note that the univariate summary statistics use the condition that the focal *T. pallida* individual has another *T. pallida* individual at distance $r$, whereas the bivariate summary statistics use the condition that the focal *T. pallida* individual has a heterospecific tree at distance $r$.

First, we analyze the spatial correlation structure of the size of the large individuals of the focal species *T. pallida* (Figure 3.33a). The univariate $r$-mark-correlation function reveals that individuals of this species have a tendency to be smaller than expected, if another *T. pallida* is in the neighborhood (Figure 3.33b). The sizes of individuals located between 5 and 35 m apart are more similar (smaller) than expected by the null model (Figure 3.33c). They are also positively correlated (Figure 3.33d). Thus, individuals located in clumps tend to be smaller than more isolated individuals.

The bivariate $r$-mark-correlation function, corresponding to the test function $t_2$, explores whether the size of *T. pallida* individuals depends on the proximity of nearby heterospecific trees. Indeed, we find a significant small-scale effect. *T. pallida* individuals that have a heterospecific neighbor within 10 m are smaller than expected (Figure 3.33e). Interestingly, the $r$-mark-correlation function corresponding to the test function $t_3$ (which returns the mean size of heterospecific trees at distance $r$ of individuals of *T. pallida*) reveals that nearby heterospecific trees tend to be larger than expected (Figure 3.33f). At
Figure 3.33
Mark-correlation function for a quantitatively marked, bivariate pattern derived from the pattern of all trees larger than 10 cm dbh in the BCI plot. The bivariate pattern is given by the individuals of the species *Trichilia pallida* (pattern 1) and the individuals of all other species (pattern 2). The quantitative mark is the dbh of the trees. (a) Map of *T. pallida*, with the area of the disks proportional to basal area. (b–d) Univariate mark-correlation analysis using the independent marking null model. (e–g) Bivariate analysis exploring spatial correlation between dbh of *T. pallida* relative to the dbh of all heterospecific individuals located distance *r* away. In the null model, dbh of the focal *T. pallida* trees was randomized among locations following independent marking, whereas the dbh of all other (heterospecific) individuals was held constant.
distances of up to 2 m they are approximately 10% larger. Thus, it looks as if individuals of *T. pallida* have a tendency to be located close to larger heterospecific trees. This hypothesis is confirmed by the results of Schlather's correlation function $I_{m_{1}m_{2}}(r)$ (i.e., test function $t_{6}$), which show that the sizes of the focal *T. pallida* trees and those of heterospecific individuals are strongly and negatively correlated up to distances of 7 m (Figure 3.33g).

### 3.1.7.6 Phylogenetic Mark-Correlation Functions

Traditional measures of alpha or beta diversity treat all species as equivalent. For example, the spatially explicit Simpson index (Section 3.1.5.1, Equation 3.70) can be expressed in simplified form as

$$\hat{\beta}(r) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n, i \neq j} I(m_{i} \neq m_{j})k(\|x_{i} - x_{j}\| - r)}{\sum_{i=1}^{n} \sum_{j=1}^{n, i \neq j} k(\|x_{i} - x_{j}\| - r)} = \sum_{l,m} p_{lm}(r) \quad (3.99)$$

where the “test function” $I(m_{i} \neq m_{j})$ yields 1, if individuals $i$ and $j$ are heterospecific, and 0 if they are the same species. Thus, in this formulation the spatially explicit Simpson index can be viewed as a mark-correlation function that yields the proportion of heterospecific pairs of individuals $i-j$ at distance $r$. However, it is also the sum of all heterospecific mark connection functions $p_{lm}(r)$.

In Section 3.1.5.3 we generalized the ISAR to a multivariate summary statistic that considers a continuous measure of phylogenetic or functional dissimilarity between species. We can do the same with the Simpson index in a straightforward way, by replacing the function $I(m_{i} \neq m_{j})$, in Equation 3.99, by a distance matrix $d(a, b)$ that represents the phylogenetic or functional distance between species $a$ and $b$. The phylogenetic Simpson index $\beta_{phy}(r)$ is then defined as the conditional mean of the test function $d(s_{i}, s_{j})$ taken over all pairs of points $i$ and $j$ that are distance $r$ apart:

$$\beta_{phy}(r) = \frac{\sum_{i,j} d(s_{i}, s_{j})k(\|x_{i} - x_{j}\| - r)}{\sum_{i,j} k(\|x_{i} - x_{j}\| - r)} = \sum_{l,m} d(l, m) p_{lm}(r) \quad (3.100)$$

The kernel function $k(\|x_{i} - x_{j}\| - r)$ yields 1, if the distance between points $i$ and $j$ is within the range $(r - dr/2, r + dr/2)$ for a bandwidth of $dr/2$, and 0 otherwise. The phylogenetic distance between species $i$ and $j$ is given by $d(s_{i}, s_{j})$, where $s_{i}$ and $s_{j}$ are the species identifiers of species $i$ and $j$, respectively.

The nonspatial expectation of the phylogenetic Simpson index $\beta_{phy}(r)$ yields the mean pairwise phylogenetic distance over all pairs of individuals present in the observation window. Comparison with Equation 3.99 shows that the only difference, when compared to the spatially explicit Simpson index, is that the indicator function $I(m_{i} \neq m_{j})$ is replaced by the distance matrix $d(s_{i}, s_{j})$. 

© 2014 Taylor & Francis Group, LLC
The phylogenetic Simpson index $\beta_{\text{phy}}(r)$, given in Equation 3.100, contains a signal of both the multivariate spatial structure of the community (captured by the spatially explicit Simpson index $\beta(r)$ shown in Equation 3.99) and a signal of the phylogenetic spatial structure. In fact, if there is no phylogenetic spatial structure, or if all heterospecific species have the same phylogenetic distance (i.e., a star type phylogeny), the phylogenetic Simpson index $\beta_{\text{phy}}(r)$ will yield the nonnormalized Simpson index. Thus, the difference between $\beta(r)/\beta^*$ and $\beta_{\text{phy}}(r)/\beta_{\text{phy}}^*$ is due to the phylogenetic spatial structure embedded in the pattern. To illustrate this, Figure 3.34a compares the normalized spatially explicit Simpson index $\beta(r)/\beta^*$ with the normalized phylogenetic Simpson index $\beta_{\text{phy}}(r)/\beta_{\text{phy}}^*$ for the spatial pattern of all recruits from the BCI plot (see Figure 3.21a; Box 2.6). We see that the phylogenetic spatial structure is largely driven by the underlying multivariate pattern captured by the Simpson index. However, at distances up to about 20 m, the normalized phylogenetic Simpson index (black dots in Figure 3.34a) is less than the normalized spatially explicit Simpson index (gray line in Figure 3.34a). This indicates that there is phylogenetic spatial structure among recruits, with neighboring recruits tending to be phylogenetically more similar. This is shown more clearly in Figure 3.34b where we divided the normalized spatially explicit Simpson index by the null model expectation.

![Figure 3.34](image)

**Figure 3.34**
Phylogenetic mark-correlation function. (a) Comparison of the normalized spatially explicit Simpson index $\beta(r)$ and the phylogenetic Simpson index $\beta_{\text{phy}}(r)$ for the spatial pattern of all recruits from the BCI plot (see Figure 3.21). (b) Phylogenetic mark correlation function $k_d(r)$ (c) Same as panel (a), but also including the simulation envelopes resulting from the species shuffling null model. Note that the spatially explicit Simpson index is now the expectation under the null model (gray line) because it corresponds to the case of no spatial phylogenetic structure. (d) Same as panel (b), but now including simulation envelopes. Phylogenies of the tree community in the BCI plot were constructed as in Figure 3.24.
phylogenetic Simpson index $\beta_{\text{phy}}(r)/\beta^*_{\text{phy}}$ by its expectation under no spatial phylogenetic structure, that is, the normalized spatially explicit Simpson index $\beta(r)/\beta^*$.

To remove the confounding effect of the multivariate spatial structure of the community [captured by the spatially explicit Simpson index $\beta(r)$], we follow the approach already taken in the generalization of the phylogenetic species area relationship PISAR (Section 3.1.5.3; Equation 3.76). This is done by normalizing the phylogenetic Simpson index $\beta_{\text{phy}}(r)$ with the spatially explicit Simpson index $\beta(r)$ to yield the phylogenetic mark-correlation function $k_d(r)$:

$$k_d(r) = \frac{\beta_{\text{phy}}(r)}{\beta^*_{\text{phy}}} \left( \frac{\beta(r)}{\beta^*} \right) = \frac{1}{c_d} \frac{\beta_{\text{phy}}(r)}{\beta(r)}$$  (3.101)

with $c_d = \beta^*_{\text{phy}}/\beta^*$. Shen et al. (in press). It can be easily shown that the phylogenetic mark-correlation function $k_d(r)$ can be estimated as

$$\hat{k}_d(r) = \frac{1}{c_d} \sum_{i,j} d(s_{p_i}, s_{p_j}) 1(s_{p_i} \neq s_{p_j}) k(\|x_i - x_j\| - r)$$  (3.102)

Thus, the phylogenetic mark-correlation function $k_d(r)$ uses basically the same estimator as the phylogenetic Simpson index $\beta_{\text{phy}}(r)$ (Equation 3.100), the difference being that conspecific pairs are excluded. This is accomplished by the indicator function $1(s_{p_i} \neq s_{p_j})$, which selects only heterospecific pairs and results in a value of 1, if the individuals $i$ and $j$ belong to different species, and zero otherwise.

The advantage of the phylogenetic mark-correlation function over the phylogenetic Simpson index $\beta_{\text{phy}}(r)$ is that it has the straightforward expectation of $k_d(r) = 1$, when there is no spatial correlation in the phylogenetic structure. However, if the neighbors of individuals tend to be more phylogenetically similar to the focal individual than expected, we have $k_d(r) < 1$ (i.e., phylogenetic clustering). In contrast, we expect $k_d(r) > 1$ (i.e., phylogenetic overdispersion), if the neighbors tend to be more phylogenetically distant species than would be the case, if there were no spatial correlation in the phylogenetic structure.

Figure 3.34c shows the results of an analysis to determine whether the recruits of tree species in the 2005 census of the BCI plot exhibited spatial structure in their phylogenetic relationships. To this end, we contrasted the observed phylogenetic Simpson index to the relationships arising from 199 simulations, where the vector of species names was randomly reshuffled. This null model constrains the entire multivariate pattern and only randomizes the phylogenetic distance matrix $d(s_{p_i}, s_{p_j})$, representing the case of no
Estimators and Toolbox

The expectation under the null model differs from a value of 1 because the recruits show a strong tendency to conspecific clustering (Figure 3.34c). The analysis shows that there is a tendency for phylogenetic clustering at small distances (<30 m), but the GoF test shows that this tendency was not significant over the range of distances less than 30 m ($P = 0.06$). Figure 3.34c also shows that the spatially explicit Simpson index (bold gray line) is the same as the expectation of the phylogenetic Simpson index under the null model of no phylogenetic spatial structure. Figure 3.34d shows the results of the same analysis, but now we used the phylogenetic mark-correlation function Equation 3.101, which excludes conspecifics. The expectation of the phylogenetic mark-correlation function under no spatial phylogenetic structure is one and, therefore, departures from the expectation are easier to visualize.

To put the phylogenetic mark-correlation function into perspective, we outline the hierarchical structure of plant communities. For example, several plots of tropical forest, arranged for example along environmental gradients over larger distances, may differ in their overall phylogenetic community structure. The latter can be determined by collecting data on the presence/absence (or abundances) of species in each of the plots and using the phylogeny of the regional species pool. Analyses such as these, presented in Kraft et al. (2007) using the program PHYLOCOM (Webb et al. 2008), then allow for quantification of the overall phylogenetic community structure, taking into account the phylogeny of the regional pool. For example, if a plot contains more phylogenetically related species than expected, given the regional pool, the local community is characterized by phylogenetic clustering.

In principle, the individuals present in a given plot could be arranged in quite different ways in space. For example, individuals of phylogenetically similar species may be located close to each other or individuals of phylogenetically dissimilar species could be located close to each other, while the relative abundances of the species in the plot (and therefore the overall phylogenetic community structure) could be exactly the same. Thus, there is an additional level of complexity hidden in the small-scale placement of individuals with respect to their phylogenetic similarity that may be manifested for pairs of individuals as correlation between their phylogenetic distance and spatial separation distance $r$. Interestingly, this correlation structure in the small-scale placement of individuals at the plant neighborhood scale is completely independent of the overall phylogenetic community structure of the local community at the plot scale. If the plot is fully mapped, we can use the phylogenetic mark-correlation function to measure how individuals of a local (fully mapped) community are arranged in space with respect to their phylogenetic similarity. This provides a means for hierarchical analyses of phylogenetic structure by decoupling analysis of the overall phylogenetic community structure of a plot (e.g., Kraft et al. 2007) and the smaller-scale spatial phylogenetic structures based on the phylogenetic mark-correlation function.
3.1.7.7 Summary for Mark-Correlation Functions

This chapter has presented the test functions and estimators for mark-correlation functions that correspond to different data types. Mark-correlation functions are the summary statistics for patterns that carry a quantitative mark, such as the size of a tree, or the phylogenetic (or trait) distance between objects, such as two species. Basically, a mark-correlation function looks at pairs of points that are separated by distance $r$ and calculates the mean value of a test function involving the marks. For example, the $r$-mark-correlation function can estimate the conditional mean size of objects that are separated from another object by distance $r$. Depending on the structure of the data, the pairs of points are subject to different conditions and only certain combinations of points are selected from the entire marked pattern. While the literature deals mostly with the simplest case of univariate, quantitatively marked patterns (e.g., the pattern of trees of one species and their size), the framework of mark-correlation functions is very flexible. Mark-correlation functions can be used with a variety of test functions, marks, and data structures, which opens up interesting applications in ecology. The framework is especially promising because it also allows for the straightforward spatial analysis of phylogenetic structures and species traits.

The univariate mark-correlation functions $k_t(r)$ allow us to determine whether there is spatial correlation among marks that depends on the distance between points. It is evident that this permits a deep investigation of the spatial-correlation structure of marks, such as the size of an individual or the phylogenetic distance between two species separated by distance $r$. In this regard, Law et al. (2009) have demonstrated that the joint use of the pair-correlation function $g(r)$ for univariate patterns, the mark-correlation function $k_{mm}(r)$, and the related, multiplicatively weighted pair-correlation function $g_{mm}(r)$ can be used to characterize the relationship between the spatial pattern of marks (tree biomass in their case) and the spatial pattern of the individual objects containing the marks. In fact, the multiplicatively weighted pair-correlation function $g_{mm}(r)$ incorporates the influence of both the spatial pattern of the points and the product of the marks of the two points separated by distance $r$. If the univariate pattern is random, it yields the mark-correlation function, and if the marks are spatially uncorrelated, it yields the pair-correlation function. Analyzing all three summary statistics together allowed Law et al. (2009) to explore whether the spatial distribution of biomass (the mark) was different from the pattern of locations. Indeed, while the locations were clustered, the marks showed inhibition, indicating, in their case, compensatory effects of nearby individuals having lower than average biomass.

The first data structure we presented for bivariate mark-correlation functions involved a univariate pattern with each point carrying two quantitative marks. This data structure allows an exploration of the spatial correlation between two properties of the points, such as size and height. It is an especially powerful approach, if a constructed quantitative mark is used in the analysis. A
constructed mark is a property of the focal point that is constructed based on neighboring points or marks. For example, in Figure 3.31 we used a mark that was constructed from the number of neighbors within 20 m of a focal individual to explore the spatial relationships between the size of the focal tree and the number of neighbors. Indeed, size and number of neighbors were negatively correlated, indicating a density-dependent effect. Other constructed marks could be the total basal area of heterospecific trees located within a given distance (e.g., distance 20 m) from a focal individual or distance to the nearest neighbor. In this case, the cross-correlation function $I_{m1m2}(r)$ is a powerful tool to reveal correlations between a quantitative mark and a constructed quantitative mark. However, the Schlather test function $t_6$ must be used if we want to yield a summary statistic that can be interpreted as a cross-correlation coefficient.

The second data structure presented with two marks carried one qualitative and one quantitative mark. Prominent examples of this type of data structure are surviving versus dead trees (the qualitative mark) and a quantitative mark that reflects some aspect of fitness, such as size (Figure 3.32). However, this data structure also has the potential for producing interesting applications when the quantitative mark is a constructed mark. For example, we could explore if dead individuals of a focal species are likely to be associated with competitive interactions by constructing marks from the total basal area (or the total number) of conspecific trees within a certain distance $r$ around the focal individual.

The final data structure we examined involved a bivariate pattern and one quantitative mark. This data structure can be used to explore if, for example, the size of a focal species exhibits a correlation with the sizes of neighboring heterospecific individuals. Indeed, in the example we used, we found that the individuals of the species Trichilia pallida had sizes that were negatively correlated with the sizes of nearby trees and that $T.$ pallida was mostly located close to larger, heterospecific trees.

In summary, mark-correlation functions provide a very flexible tool for analyzing the spatial-correlation structure of quantitative marks and we expect that they will open up new avenues for interesting analyses in ecology. This is especially true when combined with constructed marks or phylogenetic data. Over all, the methods presented in this chapter enable tests of very specific hypotheses regarding the potential drivers of spatial structure in the distribution of marks, when used creatively.

### 3.1.8 Objects of Finite Size (Data Type 10)

One of the general assumptions of point-pattern analysis is that the objects studied (e.g., plants) are dimensionless points. The point approximation is valid when the size of the object is small in comparison with the spatial scales being investigated. However, when the size of the object is roughly the same order of magnitude as the scale of interest, the point approximation
may obscure the real spatial relationships (e.g., Simberloff 1979; Prentice and Werger 1985). For ecologists this can be a real issue as the relationships they are often interested in, for example, the interactions among plants, occur at distances not much greater than the areal extent of the individual objects being investigated (Purves and Law 2002). In these circumstances, the conventional form of point-pattern analysis may suffer from a bias introduced by reducing the scale of the objects to being dimensionless points. For example, when we are interested in exploring the relationships among shrubs to determine whether facilitation or competition is a more important process (e.g., Wiegand et al. 2006), we may be interested in determining if their canopies touch more or less than expected to randomly placed, nonoverlapping shrubs. In such cases, we may need to analyze the spatial pattern of objects of finite size and real shape. Although this case falls a bit outside the scope of traditional point-pattern analysis, it is covered by the more general field of stochastic geometry (e.g., Stoyan, Kendall, and Mecke 1995). In fact, some of the tools of point-pattern analysis can be appropriately used to interpret relationships within this additional data category.

Wiegand et al. (2006) introduced an approach for dealing with objects of finite size and irregular shape by extending a grid-based estimator of second-order summary statistics that was introduced in Wiegand and Moloney (2004). The basic idea was to represent objects in a study area by means of a categorical raster map with a cell size smaller than the size of the object (Figure 3.35). One or several adjacent grid cells represent an object, depending on size and shape. These are then included in a map representing several categories such as bare ground, cover of species 1, cover of species 2, and so on (Figure 3.35a). The categorical map is then used to formally construct a point pattern in which the center of the cell is used as a point location and the categories define the types of points (Figure 3.35b). This allows conventional methods of point-pattern analysis to be used. However, consideration of the finite size and irregular shape of objects requires specification of more detail than standard point-pattern analysis. Specifically, rules and options are needed to construct objects from a categorical map, to decide on the rules of overlap among objects, and to randomize the position of the objects.

The aim of an analysis of objects with finite size and irregular shape is to test simple baseline null hypotheses on the system. Although this approach corresponds to conventional point-pattern analysis, the explicit consideration of real-world structures (i.e., finite size and irregular shape) prevents an analytical treatment. Analysis of objects with finite size and irregular shape is, therefore, a simulation-based approach for testing specific hypothesis about the spatial dependencies of objects of a system under consideration.

### 3.1.8.1 Grid-Based Estimators of Second-Order Statistics

The estimation of the univariate $K$-function in standard point-pattern analysis is based on counting the average number of points located within distance
Estimators and Toolbox

237

Similarly, the estimation of the $O$-ring statistics is based on counting the number of points at distance $r$ from the points of the pattern. Thus, the focal point is not counted. However, if we extend the analysis to objects of finite size, which may occupy several cells within a grid, a decision has to be made as to how we should generalize this rule. If we are primarily interested in the spatial structure of the map, the $O$-ring estimator $O_{lm}(r)$ has the interpretation of being the probability of finding a cell of category $m$ at distance $r$ away from a cell of category $l$ (Wiegand et al. 1999). In the univariate case, $O_{ll}(r)$ has the interpretation of being the probability of finding a cell of the same category at distance $r$ away from a cell of category $l$. In this case, the estimators of the conventional point-pattern analysis are applied to the point pattern that is formally constructed from the categorical map. The approach presented in Wiegand et al. (2006) does not count the focal cell, but allows the other cells of the focal object to be counted in the analysis. A potential disadvantage of this approach is that it conserves a strong signal from the shape of the objects, since point pairs belonging to the same focal object are considered (Nuske et al. 2009).

If we are primarily interested in the univariate spatial structure of the objects, that is, the probability that a cell of a given category is located at distance $r$ from another cell of the same category, but of a different object, additional modifications are required. Two approaches can be employed.

**FIGURE 3.35**
Typical examples of a bivariate “point” pattern with objects of real size and irregular shape. (a) A $13 \times 20$ m$^2$ plot of semiarid shrubland at Patagonia, Argentina, analyzed in Wiegand et al. (2006), showing the area occupied by three shrub species. The light gray objects are the areas occupied by the dominant shrub species *Mulinum spinosum* and the black objects are the two other dominant species *Adesmia volckmannii* and *Senecio filaginoides*. (b) Approximation of the original raster map as a point pattern for the area within the smaller rectangle of panel (a).
One approach, which uses only the nearest distance between two objects, was proposed by Nuske et al. (2009). In the second approach, distances between all cells within a focal object are excluded from the analysis. The latter approach is analogous to the standard univariate point-pattern analysis, where the focal point is not counted when estimating the $K$-function or the pair-correlation function.

When the second approach to analyzing objects of finite size described above is applied, the grid-based estimator for second-order statistics is analogous to the WM estimator presented in Equations 3.13, 3.17, and 3.36. The WM O-ring estimator divides the mean number $\bar{n}^R(r)$ of points located within rings of radius $r$ and width $dr$, centered on the points of the pattern, by the mean area $\bar{\sigma}(r)$ of the rings; that is, $\hat{O}_{WM}(r) = \bar{n}^R(r)/\bar{\sigma}(r)$. For the grid-based O-ring estimator, rings $R_i^{dr}(r)$ of radius $r$ and width $dr$ are approximated using the underlying grid structure in the calculations. The numerator $\sum_{i=1}^{n_l} \text{Points}_m[R_i^{dr}(r)]$ of the estimator is calculated from the number of type $m$ points, lying within the study region, that occur inside rings centered on the focal points $i$ of the pattern (gray cells in Figure 3.6, with a ring width of one cell). The denominator of the estimator is obtained by calculating the area of the rings $\sum_{i=1}^{n_l} \text{Area}[R_i^{dr}(r)]$ that fall within the observation window, yielding

$$\hat{O}_{\text{WM}}^{dr}(r) = \frac{\sum_{i=1}^{n_l} \text{Points}_m[R_i^{dr}(r)]}{\sum_{i=1}^{n_l} \text{Area}[R_i^{dr}(r)]} \quad (3.103)$$

Estimator 3.103 basically counts the number of cells of type $m$ in rings with radius $r$ around cells of type $l$ and divides by the corresponding area of the rings. The estimator of the $K$-function can be developed in an analogous fashion, yielding

$$\lambda_l K_{\text{WM}}(r) = \pi r^2 \frac{\sum_{i=1}^{n_l} \text{Points}_m[C_i(r)]}{\sum_{i=1}^{n_l} \text{Area}[C_i(r)]} \quad (3.104)$$

where the rings $R_i^{dr}(r)$ are replaced by circles $C_i(r)$ with radius $r$ centered on the cells $i$ of the focal pattern $l$. Here, $\lambda_l$ represents the intensity of the cells of the focal pattern $l$ (= number of cells of type $l$ divided by number of cells within the observation window). As mentioned above, the estimators given in Equations 3.103 and 3.104 can be evaluated in the univariate case (i.e., $l = m$) in two different ways. First, we may also count pairs of cells that belong to the same object, and second, we may exclude cell pairs belonging to the same object.

For interpretation of the univariate summary statistics, it is worthwhile to note that, if the distance $r$ among objects is relatively large compared to
the size of the objects, the resulting pair-correlation and $K$-functions can be approximated by multiplicatively weighted pair-correlation and $K$-functions (see Section 3.1.7.1, Equation 3.91). This results from the fact that, if the focal object contains $s_i$ cells and the second object contains $s_j$ cells, the estimator of the pair-correlation function counts the distance between two objects $s_i \times s_j$ times and all $s_i \times s_j$ distances between cells of the two objects are approximately the same. The resulting calculation is essentially the same as the value produced by the multiplicatively weighted pair correlation where the marks are the size of the objects.

### 3.1.8.2 Additional Features Required for Analysis of Objects of Finite Size

An important difference between the analysis of points versus objects is that objects may or may not overlap whereas points, as one-dimensional objects, cannot overlap if they are approximated by their coordinates (unless they occupy exactly the same location). In particular, randomization of the position of objects of finite size and real shape requires rules that determine whether or not the objects are allowed to overlap. For example, in the grid-based framework as described above, a category (and not a number of points) is assigned to each cell, and overlap of two objects of the same pattern is not allowed. This has several implications. First, the $O$-ring statistic, the most important summary statistic for studying objects of finite size and real shape, has a slightly different interpretation than in conventional point-pattern analysis. Additional rules are also required for estimation of the $O$-ring statistic and the $K$-function. Second, the cell size (i.e., the grain of the map) needs to be adapted to the biological question at hand. For example, if the question requires that small objects be explicitly considered, the cell size must be made small enough to allow individual objects to be represented with at least one cell. Finally, depending on the null hypothesis, it may or may not be allowed for objects of two different types to overlap. This rule will affect data collection and mapping, as well as the application of null models.

### 3.1.8.3 Algorithm to Define Objects for Randomization

Null models for objects of finite size and irregular shape require randomization of locations of the objects (Wiegand et al. 2006). Therefore, an algorithm must be employed that determines which neighboring cells of a categorical map belong to the same object. One approach uses a “fire spread” (or “flood- ing”) algorithm for this purpose. This algorithm repeats three operations. It starts with selection of random cells. If it encounters a cell $i$ that is of a given category $c$, the first operation is to place this cell into a queue. The next operation is to select the first element of the queue (in the first step of the algorithm this is cell $i$), assign it to the object and delete it from the queue. However, the third operation visits the eight (or four) adjacent cells in turn to determine whether they also fall into category $c$. The neighbors of cell $i$, which are in
category $c$, are marked and put into the queue. Now the second step of the algorithm starts. Again, the first element of the queue is selected, assigned to the object, deleted from the queue, and its neighbors are examined. If the neighbors are in category $c$ and do not yet belong to the object, they are marked and put into the queue. The third step of the algorithm again selects the first element in the queue and so on. This algorithm is repeated until no cells are left in the queue. Thus, the algorithm starts with a random cell $i$, which is of category $c$, and determines all adjacent cells of category $c$ that are connected with cell $i$ via direct neighbors of category $c$. The algorithm stops if no cell of the object has a direct neighbor of category $c$, which is not yet assigned to the object. As a result, all adjacent cells of category $c$ are marked and form one object. This algorithm is repeated until all cells of category $c$ are assigned to an object.

One potential problem with this approach is that two objects that touch will be assigned to a single object. This can be avoided by using additional information on the identity of objects that must be collected during the field survey. If it is known which adjacent cells belong to a single object we can represent category $c$ (e.g., one species of shrubs) by several subcategories, say $c_1, c_2,$ and $c_3,$ and assign these “dummy” categories in such a way that two objects, which touch, will still belong to different subcategories. In this way, the software can separate objects of the same category when they touch. This requires slightly more effort in developing the grid-based maps for analysis.

3.1.8.4 Randomization of the Position of Objects

Randomization of points involves assigning random coordinates, whereas randomization of objects of finite size needs to preserve the shape of the object, which may occupy several cells. This can be achieved by rotating and mirroring the object by $0^\circ, 90^\circ, 180^\circ,$ or $270^\circ$ (each of the eight variants being equally probable) and then randomly shifting its location in space, as a whole, but not allowing overlap with objects of the same type (Figure 3.36a,b).

Repeated trials are performed for each individual object: if, after being randomly mirrored, rotated, and shifted, an object overlaps with an already distributed object of the same type (or another type if appropriate), or falls partly outside of an irregularly shaped study area, or does not conform with the selected method of edge correction (see below), the trial is rejected. The procedure is repeated until a location is found for all objects.

3.1.8.5 Masking (Space Limitation)

In the analysis of ecological objects, competition for space is often an important ingredient to consider in developing null models for individuals of finite size. We may encounter situations where individuals of a focal species cannot inhabit some areas, for example, those already occupied by other species. In this case, we may ask if the species of interest is randomly distributed,
conditioned on the locations of plants of a second species. The categorical map approach facilitates inclusion of such space restrictions in an elegant way. All nonaccessible cells are assigned to a category called “mask” and all cells belonging to the mask are excluded from the study area. In practice, this means that, during the repeated randomization trials of the given species, a trial is rejected if the tentatively placed individual overlaps a cell included in the mask category.

### 3.1.8.6 Edge Correction

The finite size of objects requires edge correction, since randomly displaced objects may fall partly outside the observation window. In the null model, this would reduce the proportion, $\lambda$, of occupied cells and produce a slight (positive) bias toward aggregation. There are three methods that can be used to mitigate this effect (Wiegand et al. 2006). First, randomized objects are not allowed to fall partly outside the observational window. This produces a negative bias toward hyperdispersion since fewer objects of the null model will be distributed close to the border. The second method avoids this problem by treating the observational window encompassing the study region as a torus, that is, the part of an object lying outside the window reappears at the opposite border. However, breaking relatively large objects into two smaller objects produces a slight positive bias. A third method uses the torus correction, but calculates the summary statistics only inside an inner rectangle excluding cells close to the border (guard area). For a guard area wider than the diameter of the largest plants, the biases of the first and the second method disappear, but this may reduce the size of the study rectangle and thus the number of objects considered in the analysis. Therefore, the guard

![FIGURE 3.36](image-url)

Randomization of the position of objects using a null model. (a) Pattern of objects with real shape and size. (b) Randomization of objects shown in panel (a), highlighting the change in position for three examples with arrows.
area selected needs to be wider than the diameter of most objects, but still small enough to yield a large enough sample.

### 3.1.8.7 Univariate Analysis of Objects of Finite Size

In univariate analyses, the fundamental division is provided by the *completely random distribution of objects*, as opposed to aggregation or hyperdispersion. Objects of finite size and irregular shape are, by analogy to CSR, distributed randomly as described above. Depending on the biological question, the available area may be the entire observation window, an irregularly shaped subarea of the observation window, or the area remaining after masking regions of the grid that are hypothesized to constrain the locations of the focal objects, for example, through competitive effects.

Figure 3.37 shows an example of a univariate analysis for a pattern based on the distribution of the shrub species *Mulinum spinosum* (Figure 3.35a). Simulation envelopes were constructed by producing 199 simulations of the null model, which randomized the locations of the plants. We show results of the analysis of two cases, one in which pairs of cells of the focal object were counted (as in Wiegand et al. 2006) and one in which we considered only pairs of cells of different objects. Analysis by the univariate O-ring statistic $O(r)$ gives the probability that a cell located at a distance $r$ away from a cell occupied by a shrub is also occupied by a shrub. Clearly, in the first case, where pairs of cells of the focal object are included in the analysis, $O(r)$ essentially depicts the spatial structure (i.e., size distribution) of the individual shrubs at small scales, which is not of particular interest. This can be seen in Figure 3.37a, which shows that the probability of having a neighbor of the same category (here a *M. spinosum* shrub) declines rapidly with increasing distance from the focal cell, but reaches the expectation given by the over all shrub cover at distances greater than 1 m. The solid gray line in Figure 3.37a gives the expectation of $O(r)$ under the null model and basically captures the geometry of the objects. However, potential departures from the null model are difficult to see because of the steep decline in the value of $O(r)$ at scales up to about 1 m.

Including only distances between different objects in the analysis (open disks in Figure 3.37a) completely removes the signal of the shape of individual objects and places the focus on the spatial relationships among objects. In our example, this reveals a negative departure from the null model at distances less than 30 cm (Figure 3.37b), indicating an additional, subtle effect of regularity at small scales. This information was obscured in Figure 3.37a by including pairs of cells of the focal object in the analysis. However, the effect between shrubs indicated by the analysis may actually be an artifact produced by the method used in mapping the pattern, since shrubs were defined in our example as consisting of adjacent cells of the same type, in effect prohibiting two shrubs from having cells that touched.

The spatial pattern of the focal shrub species may be constrained by the placement of the other shrubs in the system. To test for this potential effect
we considered the pattern of the other two species *Adesmia volckmanni* and *Senecio filaginoides* (Figure 3.35a) as a mask, which prohibited relocated shrubs of the focal species from overlapping the areas occupied by the two other shrub species. We obtain basically the same results, with and without masking (cf. Figure 3.37b and c), which indicates that competition for space exerted by the other two dominant shrub species is not an overriding process shaping the spatial pattern of the focal shrub species.
It is instructive to compare the results of the analyses of objects with finite size and irregular shape to those of the point approximation and mark-correlation analysis. Figure 3.37d shows the results of a simple point-pattern analysis, where the objects were approximated by their center of gravity. There is repulsion at very small scales but, in contrast to the results shown in Figure 3.37b, the regularity is caused by the physical size of the objects (i.e., a soft core effect). Next, we marked the point pattern with the size of the corresponding basal areas. The $r$-mark-correlation function, which returns the (normalized) average size of basal areas at distance $r$ from the focal object, shows that neighboring individuals are smaller, if they are closer than 50 cm to the focal object (Figure 3.37e). The mark-correlation function, which calculated the mean normalized product of the sizes of two basal areas separated by distance $r$, depicts basically the same significant effect (Figure 3.37f). Thus, either explicitly considering the shape of the shrubs in the analysis or using the size information in a mark-correlation function yields the same result, that is, there is small-scale repulsion among the shrubs.

### 3.1.8.8 Bivariate Analysis of Objects of Finite Size

In bivariate analyses the fundamental division for objects of finite size is given by independently distributed patterns. This provides the separation line between attraction and repulsion (segregation). The null model testing for independence of two-point patterns assumes that independent stochastic processes created the component patterns. This leads to the idea of using a toroidal shift to produce a null model testing for the independence of the point patterns (see Section 3.1.1). The same approach applies for bivariate patterns of plants of finite size: pattern 1 remains fixed, whereas pattern 2 is randomly shifted as a whole across the study area, using a torus (toroidal shift) to reposition the parts of the pattern that cross an edge.

One aspect should be noted when using the toroidal shift as a null model: it requires that objects of different patterns be allowed to overlap. This may introduce a bias toward repulsion at small scales if the objects are known not to overlap. As an alternative, a univariate null model could be introduced to describe the second-order structure of pattern 2 and, if appropriate, an overlap of plants could be prohibited. In the simplest case, this would be the null model where the plants of pattern 1 remain unchanged, but plants of pattern 2 are distributed at random, with the condition that they are not allowed to overlap plants of pattern 1 (i.e., a completely random distribution of objects of pattern 2 constrained by the objects of pattern 1). Note that one may also use a null model that conditions on the larger scale pattern of the plants of pattern 2 by means of a heterogeneous Poisson process (Sections 2.6.1.2, 2.6.3.3, and Box 2.3). Another option would be to extend the techniques of pattern reconstruction (Section 3.4.3) to objects and create, for the independence null model, stochastic replicate patterns that also conserve intermediate scale correlations in the placement of objects.
Figure 3.38a shows a bivariate pattern composed of the shrub *A. volckmanni* (pattern 1, black) and the pattern of two other dominant shrub species *M. spinosum* and *S. filaginoides* (pattern 2, gray). The bivariate O-ring statistic is zero at very small distances, indicating no overlap between the two patterns, and increases up to a distance of 50 cm, before reaching saturation at the expectation.
of independence under the null model (i.e., the probability that a cell is occupied by the second pattern; i.e., *M. spinosum* or *S. filaginoides* shrubs) (Figure 3.38b). As expected, application of the toroidal shift null model exhibits repulsion at small scales (caused by the nonoverlap of the two patterns). At greater distances no significant departures from the null model were observed.

An approach that is more rewarding than using the toroidal shift as a null model is the application of a null model where the *A. campestris* pattern remains fixed and the basal areas of all other shrubs are randomly distributed without overlap. We find that the two groups of shrubs show repulsion at small scales up to 30 cm (Figure 3.38b). This means that the shrubs of the second pattern are located somewhat farther away from *A. campestris* shrubs than would be expected through random placement without overlap. This may be an effect of competition between the two groups of plants.

Using the point approximation and testing for independence with the toroidal shift null model indicates independence between the two patterns. The signal of repulsion between the two groups of shrubs seen in the corresponding analysis with the full shape can only be seen in the point approximation as a nonsignificant tendency (cf. Figure 3.38b and d). This indicates that retaining the size and shape information may be critical in this analysis. The importance of the size information is also outlined by application of the \( r \)-mark-correlation function, which yields the (normalized) mean size of the shrubs of the second pattern at distance \( r \) away from shrubs of the focal pattern (Figure 3.38e). We used a null model where the sizes of the shrubs of the second pattern were randomly shuffled among shrubs of this pattern. This analysis revealed that the size of individuals of the second shrub pattern was smaller, if a shrub of the focal pattern was nearby (i.e., nearer than 30 cm). This is an effect of geometry: larger shrubs will not be close together.

In this section, we have presented an extension of conventional point-pattern analysis, where objects are approximated as points in categorical raster maps that can represent objects of finite size and irregular shape. This approach facilitates incorporation of real-world structures into point-pattern analysis and provides a powerful tool for the statistical analysis of objects for which the common point approximation would not be capable of revealing important information. Clearly, mapping plants and characterizing them by their real shapes, instead of treating them as dimensionless points, takes considerably more effort, but is necessary if we are to consider real size and shape effects instead of employing crude point approximations.

### 3.2 Replicate Patterns

Mapping a plant community for the purpose of conducting a point-pattern analysis often involves a tradeoff between sampling one large plot or
several smaller plots (Illian et al. 2008, p. 269). Mapping one large plot has the advantage that edge effects will be less severe (i.e., there are fewer plants at the border of the plot impacted by the influence of unknown plants outside the plot), but one large plot may not sufficiently represent the typical conditions occurring within an entire study site. For statistical analysis, it is a common practice to sample several replicate plots under identical conditions within a study site; this can also be done for point patterns. Point-pattern analysis provides methods that allow for the aggregation of results obtained from individual analyses of several replicate plots. Basically, the resulting test statistics of the individual replicate plots are combined to produce an average (Diggle 2003, p. 123; Illian et al. 2008, p. 260f). This is of particular interest when the number of points in each of the replicate plots is relatively low since the envelopes resulting from the analysis of individual plots would be quite wide. By combining the results of several replicate plots into one average test statistic, the sample size is increased and the range of values covered by the simulation envelopes is narrowed. There are several examples in the ecological literature where replicate plots have been combined in a point-pattern analysis (e.g., Riginos et al. 2005; Blanco et al. 2008; De Luis et al. 2008; Raventós et al. 2010, 2011; Jacquemyn et al. 2010).

A straightforward method for combining data from different replicate point patterns would be to place them (artificially) into one large observation window (i.e., combine them into one large data set). This can be done; however, care must be taken to avoid mixing points from one replicate plot with those of another plot in the resulting estimator. We can insure this geometrically by shifting the coordinate systems of the separate replicate plots so that they are “farther apart” than $r_{max}$, the maximal distance at which the summary statistics are evaluated (Figure 3.39). Additionally, the boundaries of the “observation window” have to be set so that only subplots are considered to be “inside” the observation window. The latter condition is necessary to prevent bias due to edge correction (Figure 3.39). The pair-correlation function can then be estimated using the inhomogeneous pair-correlation function based on the estimator given by Equation 3.31, together with an intensity function $\lambda(x)$ that yields the observed intensity $\lambda_m$ within each of the replicate patterns $m$ and a value of zero outside (Figure 3.39). In this case, the null models would distribute the points only inside the replicate windows $m$ with their associated intensity $\lambda_m$. Although this method may work well for the pair-correlation function, and help us in motivating suitable aggregation formulas, there are actually more direct methods for aggregating the results of individual replicate analyses. The specific formula used in combining results from several replicate plots depends on the estimators of the test statistics used. Illian et al. (2008, p. 261f) provide aggregation formulas for the most important summary statistics. In the following Section 3.2.1, we provide aggregation formulas for univariate patterns, and in Section 3.2.2 we present some
example applications using this approach. Note that it is also relatively straightforward to extend the aggregation formulas for univariate patterns to bivariate and marked patterns.

### 3.2.1 Aggregation Formula

In developing an aggregation formula for $M$ replicate plots, we can estimate a summary statistic $S_m$ for each plot $m$. Aggregation of $S_m$ into one summary statistic $S$ can be developed from two basic principles. The first principle is to estimate a weighted mean of the summary statistic $S_m$ of the $M$ individual plots:

$$
\hat{S} = \frac{\sum_{m=1}^{M} c_m \hat{S}_m}{\sum_{i=1}^{M} c_m}
$$

(3.105)

where the $c_m$ are appropriate weights, which depend on the nature of the summary statistic. The second principle for deriving aggregation formulas applies if the estimators of the summary statistic $S$ have a particular functional form, which arise from “adapted” estimators that can be expressed as a ratio of sums over focal points $i$ (we omit the index $m$ for clarity):
The numerator and the denominator in Equation 3.106 are sums of the functions \( f(x_i, r) \) and \( g(x_i, r) \), which depend on the focal point \( i \) and other points within or at distance \( r \) from the focal point. For example, the WM estimator of the O-ring statistic (Equation 3.17) follows this functional form. Here, the numerator is given by \( f(x_i, r) = \sum_{j=1, j \neq i}^{n} k_i^B(\|x_i - x_j\| - r) \), which is a factor that returns the number of points falling at distance \( r \pm dr/2 \) from a focal point \( i \). The denominator \( f(x_i, r) = \nu_{i-1}(W \cap \partial b(x_i, r)) \) is the area of the corresponding ring with radius \( r \) and width \( dr \) centered on point \( i \). The sums are taken over all points \( i \) of the focal pattern. For second-order summary statistics the function \( f(x_i, r) \) counts the number of points in the neighborhoods of point \( i \), whereas the function \( g(x_i, r) \) determines the area of the neighborhood (e.g., Equations 3.17 and 3.36). For mark connection or mark-correlation functions, the function \( f(x_i, r) \) determines the sum of a test function involving the marks of focal point \( i \) and all other points at distance \( r \) from this point, and the function \( g(x_i, r) \) counts the number of points in the neighborhood of point \( i \) (e.g., Equation 3.84).

Basically, these estimators visit each point \( i \) of the pattern and estimate the mean value of the functions \( f(x_i, r) \) and \( g(x_i, r) \), which summarize neighborhood properties of point \( i \) as a function of the radius \( r \). Because the factor \( 1/n \) appears in both the numerator and denominator and cancels out, the estimator given in Equation 3.106 simply sums up the contributions of all points \( i \) of the pattern. We can extend this principle to points of different subplots, which suggests a natural way to aggregate

\[
\hat{S}(r) = \frac{\sum_{i=1}^{n_1} f_1(x_i, r) + \cdots + \sum_{i=1}^{n_M} f_M(x_i, r)}{\sum_{i=1}^{n_1} g_1(x_i, r) + \cdots + \sum_{i=1}^{n_M} g_M(x_i, r)} = \frac{\sum_{m=1}^{M} e_m}{\sum_{m=1}^{M} d_m} \tag{3.107}
\]

Each point \( i \) contributes in the same way to this estimator. Thus, we can apply a simple aggregation formula by saving, not only the estimate of the summary statistics \( \hat{S}_m(r) \), but also the numerator \( e_m = \sum_{i=1}^{n_m} f_m(x_i, r) \) and the denominator \( d_m = \sum_{i=1}^{n_m} g_m(x_i, r) \) for each plot \( m \). Examples for summary statistics that have the functional form of Equation 3.106 are the O-ring statistic, based on the WM edge correction (Equation 3.17), the K-function based on the WM edge correction (Equation 3.36), the distribution function of the distances to the \( k \)th neighbor (Equation 3.47), mark connection functions (Equation 3.80), and non-normalized mark-correlation functions (Equation 3.84). Note that we need to use estimators of \( \lambda g(r) \), \( \lambda K(r) \), and non-normalized mark-correlation functions, otherwise we would add a weighting factor of
$1/\lambda_m$ or $1/c_i^m$ to the corresponding contribution $e_m$ of subplot $m$ in Equation 3.107. Thus, in this case the results of subplots with low intensity (or low value of $c_i^m$) would receive too much weight in the combined estimate.

Note that the aggregation principle of Equation 3.107 can be derived from the “naive” method of aggregating replicates shown in Figure 3.39. The data of the $M$ plots are integrated into one data base by arranging them in the way shown in the figure, which guarantees that the distances between points of the different replicate plots are greater than the maximal distance $r_{\text{max}}$ to be analyzed. It can be easily verified that the estimator of type 3.108 will result in the estimator given in Equation 3.106, as long as the functions $g_m$ and $f_m$ of replicate plot $m$ do not use points from any other replicate plot. For the $O$-ring statistic, this is guaranteed if the distance among replicate plots is greater than $r_{\text{max}}$.

### 3.2.1.1 Intensity

Following Illian et al. (2008, p. 261), the aggregated estimator for the intensity $\lambda$ of $M$ windows $W_m$ yields

$$\hat{\lambda} = \sum_{m=1}^{M} \frac{\hat{\lambda}_m A_m}{A} = \frac{\sum_{m=1}^{M} n_m}{\sum_{m=1}^{M} A_m} = \frac{n}{A}$$

(3.108)

where $\hat{\lambda}_m$ is the estimate of the intensity of points within the $m$th window, $A_m$ is the area of the $m$th window, and $A$ is the sum of the areas of all the windows $A_m$. If the intensity of all of the $m$ windows were estimated by the natural estimator $\hat{\lambda}_m = n_m/A_m$, we find that the aggregated estimator is given by $\hat{\lambda} = n/A$, where $n$ is the total number of points in all $m$ windows and $A$ is the total area (Equation 3.108). Thus, the aggregation formula follows the recipe given in Equation 3.108. However, in the aggregation formula 3.110, the intensity $\hat{\lambda}_m$ is multiplied by the relative area $A_m$ of plot $m$. Thus, the aggregation recipe also follows the approach of Equation 3.105, where the area $A_m$ of plot $m$ is the weighting factor $c_i^m$. This is reasonable because the number of points will be proportional to the area $A_m$.

### 3.2.1.2 Nearest-Neighbor Distribution Function

The aggregation formula for the nearest neighbor (or the $k$th neighbor) distribution function depends on the estimators used (see Section 3.1.3). If no edge correction is used, the estimators for the $m$ individual plots are weighted with the relative number of points in each window $W_m$:

$$\hat{D}(r) = \frac{\sum_{m=1}^{M} n_m \hat{D}_m(r)}{\sum_{m=1}^{M} n_m}$$

(3.109)
where \( n_m \) is the number of points in the \( m \)th window. In this case, the number of points in the \( m \)th window provides the weights \( c_m \) for Equation 3.105. However, if the border estimator is used (see Section 3.1.3; Equation 3.44), only focal points that are located further away than distance \( r \) from the nearest border are used for the evaluation. Therefore, the aggregation formula becomes

\[
\hat{D}(r) = \frac{\sum_{m=1}^{M} n_{r,m} \hat{D}_m(r)}{\sum_{m=1}^{M} n_{r,m}} \tag{3.110}
\]

where the \( n_{r,m} \) is the number of points inside a reduced window \( W_{mr} \) with only points farther than distance \( r \) to the nearest border included as focal points (Illian et al. 2008, p. 261).

We can also apply the second aggregation principle to a calculation of nearest-neighbor distribution functions, since the border or Hanisch estimator (Equation 3.47) has the functional form of Equation 3.106. For example, with the Hanisch estimator we can define \( g(x_i, r) = 1(d_{ik} < \beta) \times w(d_{ik}) \) and \( f(x_i, r) = 1(0 < d_{ik} \leq r) \times g(x_i, r) \) and rewrite Equation 3.45 as

\[
\hat{D}_h(r) = \frac{\sum_{i=1}^{n} \left( 1(0 < d_{ik} \leq r) \times g(x_i, r) \right)}{\sum_{i=1}^{n} g(x_i, r)} = \frac{e}{d} \tag{3.111}
\]

Here, \( d_{ik} \) is the distance to the \( k \)th nearest neighbor and \( \beta_i \) is the distance from the point \( i \) at location \( x_i \) to the nearest border. We can then use the aggregation formula

\[
\hat{D}_h(r) = \frac{\sum_m e_m}{\sum_m d_m} \tag{3.112}
\]

if we save the numerators \( e_m \) and denominators \( d_m \) of estimators for the \( m \) replicate plots. Note that both aggregation receipts (i.e., Equations 3.105 and 3.107) yield basically the same results because the denominator of Equation 3.112 is an estimator of the intensity of points in \( W_m \).

### 3.2.1.3 Spherical Contact Distribution

The spherical contact distribution \( H_s(r) \) yields the probability that the first neighbor of a univariate pattern is distance \( r \) away from an arbitrary test location \( t \). This can be conceptualized as a bivariate nearest-neighbor distribution function \( D_{tp}(r) \) that gives the proportion of test points \( t \) which has the nearest point of pattern \( p \) within distance \( r \). In this case, we can use the
preceding aggregation formulas for the distribution function to the nearest neighbor. Illian et al. (2008, pp. 204, 261) propose a slightly different estimator and aggregation formula based on the weighted mean (Equation 3.105). Their weighting factor is given by the relative area of the mth replicate plot \( A_m/A \).

### 3.2.1.4 Pair-Correlation Function

Illian et al. (2008) and Law et al. (2009) proposed an aggregation formula for the pair-correlation function that considers a weight \( \gamma \) based on the isotropized set covariance \( \gamma_W(r) \) (Section 3.1.2.1; Equation 3.7):

\[
\hat{g}(r) = \frac{\sum_{m=1}^{M} \bar{g}_{W_m}(r) \hat{g}_m(r)}{\sum_{m=1}^{M} \bar{g}_{W_m}(r)}
\]

(3.113)

This is a reasonable approach for the case that all windows \( W_m \) have the same shape, as the weight reduces to the relative area of the mth window due to the fact that \( \bar{g}_{W_m}(r) \) is proportional to the area \( A \) of the window (Equation 3.7). However, this weight also provides an additional correction if the shape of the windows \( W_m \) differ. Regarding this, Law et al. (2009) provide an interpretation of Equation 3.113: “The isotropized set covariance gives a larger weight to estimates of the pair-correlation function in square windows than in rectangular windows at larger distances. This reflects the fact that a larger number of points with larger distances can be expected in a square window than in a rectangular window yielding a better-quality estimation at these distances, p. 625.” Note that the formula given in Equation 3.113 is adapted to the case of Ohser edge correction (Equation 3.21) because it eliminates the individual edge correction terms for replicates \( m \) given by the isotropized set covariance \( \bar{g}_{W_m}(r) \) of window \( W_m \) and replaces them by their average.

If we use the WM estimator of the O-ring statistic (Equation 3.17) we can apply the aggregation formula (3.107)

\[
\hat{O}_{WM}(r) = \frac{\sum_{i=1}^{n} \left( \sum_{j=1}^{n_{i,j}} k_{b}^{\mathcal{B}}(\|x^i_{1} - x^j_{1}\| - r) \right) + \cdots + \sum_{i=1}^{n_{M}} \left( \sum_{j=1}^{n_{M,j}} k_{b}^{\mathcal{B}}(\|x^M_i - x^M_j\| - r) \right)}{\sum_{i=1}^{n} v_{d-1}(W_i \cap \partial b(x^1_i, r)) + \cdots + \sum_{i=1}^{n_{M}} v_{d-1}(W_M \cap \partial b(x^M_i, r))}
\]

(3.114)

where the superscripts refer to the plots. The estimator of the O-ring statistic can then be combined with the estimator of the intensity (Equation 3.108) for the final estimate of the pair-correlation function.
3.2.1.5 K-Function

Diggle (2003, p. 123) and Illian et al. (2008, p. 262) propose the following aggregation formula for the K-function:

\[
\hat{K}(r) = \frac{\sum_{m=1}^{M} n_m \hat{K}_m(r)}{\sum_{m=1}^{M} n_m} \quad (3.115)
\]

This is the weighted average of the K-functions for the individual windows \(W_{m}\) with weights provided by the relative proportion of points in \(W_{m}\). However, this weighting recipe is somewhat inconsistent with that of the pair-correlation function (Equation 3.113). Following the same philosophy as for the pair-correlation function, the cumulative isotropized set covariance

\[
c_m = \frac{2\pi \int_0^r t \mu_{W_{m}}(t) dt}{\pi r^2} \quad (3.116)
\]

can be used as the weighting factor \(c_m\).

If we use the WM estimator of \(\lambda K(r)\) given in Equation 3.36, we can apply the aggregation formula (3.107) with a similar approach to the one used in calculating the pair-correlation function. We then obtain

\[
\lambda K^{WM}(r) = \pi r^2 \frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_{ij}} 1(\|x_i^j - x_j^i\|, r)) + \ldots + \sum_{i=1}^{n_M} \left(\sum_{j=1}^{n_{Mj}} 1(\|x_i^M - x_j^M\|, r))\right)}{\sum_{i=1}^{n_1} v(W_i \cap b(x_i^1, r)) + \ldots + \sum_{i=1}^{n_M} v(W_M \cap b(x_i^M, r))} \quad (3.117)
\]

Again, the estimator of \(\lambda K(r)\) can be combined with the estimator of the intensity (Equation 3.108) to estimate the K function.

3.2.1.6 Mark-Connection Functions and Mark-Correlation Functions

The estimators of mark-connection functions (Equations 3.80 and 3.83) and those of non-normalized mark-correlation functions (Equations 3.84, 3.96, 3.97, 3.99, and 3.102) all have the functional form of Equation 3.100. It is, therefore, straightforward to apply the aggregation formula of Equation 3.107, but, in doing this, the numerator and denominator of the estimators must be saved separately for each plot.
3.2.2 Examples of the Application of Aggregation Formulas

In this section, we provide several practical examples of the application of the aggregation formulas. The basic application is, of course, to combine the results of several replicate plots into a single summary statistic and produce the associated simulation envelopes. However, we also point to the fact that the aggregation formulas can be used to implement very specific null models that otherwise would require very specialized software.

### 3.2.2.1 Increasing the Power of GoF Test

One of the principle applications of the aggregation formulas is to combine the results of replicate plots with a low number of points to produce more sensitive analyses, since the low number of points within individual plots generally results in wider simulation envelopes. The combined data act to improve the results of a point-pattern analysis by yielding a single summary statistic with narrower simulation envelopes, due to increased sample size. To illustrate this, we consider an example of nine 250 \times 250 m^2 plots, with each pattern containing \( n = 50 \) points (Figure 3.40). The patterns were created with a Thomas process (Box 2.5) consisting of 13 clusters with a radius of 44 m. The expectation of the pair-correlation function under this point process is shown as a gray line in Figure 3.40a,b. Figure 3.40a presents the analysis of an individual plot (plot 1), which has very wide simulation envelopes arising from the 199 simulations of the CSR null model. Additionally, the pair-correlation function of plot 1 fluctuates somewhat erratically. Although the observed pair-correlation function lies, at some spatial scales, marginally outside of the simulation envelopes, the GoF test (see Section 2.5.1.2) does not detect a significant departure from the CSR null model over the distance interval of 1–50 m. The P-value of the test yields \( P = 0.12 \). These results are a consequence of the high degree of stochasticity, which does not allow a separation of the real signal of clustering from the stochastic fluctuations. As a consequence, five of nine plots did not yield significant departures from the CSR null model, although they were all created with a point process that incorporated an explicit cluster mechanism. However, when aggregating the results of the nine plots with Equation 3.114, the estimate of the pair-correlation function approximates the expected pair-correlation function of the underlying Thomas process reasonably well, and the simulation envelopes are considerably narrower (Figure 3.40b). Because the sample size was much larger for the combined analysis (\( n = 450 \)), there is a clear departure from CSR (Figure 3.40b).

### 3.2.2.2 Intra- versus Interspecific Interactions

The aggregation formulas are not only an effective way of combining the results of several replicate plots; they can also be used for specific applications that would otherwise require very specific software. One example for such an application is the study of mortality in multi-species communities,
where we may wish to test the “segregation hypothesis” (Pacala 1997; see also Chapter 1). This hypothesis states that most plants in communities that show a particular segregation pattern (i.e., intraspecific aggregation and interspecific segregation) will compete at short distances with con specifics, which prevents (or retards) competitive exclusion and contributes to coexistence (Pacala 1997; Murrell et al. 2001).

One possibility for testing this hypothesis is to use data from repeated censuses of a plant community. In this case, each individual is characterized by its spatial position and species identifier, as well as a mark indicating whether it is alive or dead. We use the subscript 1 to refer to dead individuals and subscript 2 to refer to those surviving. The task is to determine whether

![Figure 3.40](image-url)
dead plants, as well as surviving plants, are in fact mostly surrounded by conspecifics, as suggested by the segregation hypothesis. To this end, we decompose the overall neighborhood density of dead plants (subscript 1) around dead plants (subscript 2) \[ \hat{O}_{11}(r) \] and of surviving plants around surviving plants \[ \hat{O}_{22}(r) \] into two components, one that considers only conspecific pairs \[ \hat{O}_{\text{intra}}^s(r) \] and one that considers only heterospecific pairs \[ \hat{O}_{\text{inter}}^s(r) \].

To derive the estimators of these quantities, we first need to calculate for each species the four possible neighborhood densities \[ \hat{O}_{11}(r) \], \[ \hat{O}_{12}(r) \], \[ \hat{O}_{21}(r) \], and \[ \hat{O}_{22}(r) \] considering both dead and surviving plants. The subscript 1 refers to dead individuals and subscript 2 to surviving individuals. In all cases, the first subscript indicates the focal pattern and the second subscript represents the pattern that is assessed around the focal pattern.

For a given species \( s \), \[ \hat{O}_{11}(r) \] is the density of dead plants at distance \( r \) away from other dead plants; \[ \hat{O}_{22}(r) \] is the density of surviving plants at distance \( r \) away from other surviving plants; \[ \hat{O}_{12}(r) \] is the density of surviving plants at distance \( r \) away from dead plants; and \[ \hat{O}_{21}(r) \] the density of dead plants at distance \( r \) away from surviving plants.

For a given species \( s \), the estimator of the bivariate \( \hat{O}_{ij}(r) \) \((i, j = 1, 2)\) can be written in simplified notation as

\[
\hat{O}_{ij}(r) = \frac{\sum_{k=1}^{n_i^j} \text{Points}_j[R^s_k(r)]}{\sum_{k=1}^{n_i^j} \text{Area}[R^s_k(r)]}
\]

where the summation is done over the \( n_i^j \) type \( i \) points of species \( s \), \( R^s_k(r) \) is a ring segment with radius \( r \) and width \( h \) centered at a point \( k \) (which is of species \( s \) and focal type \( i \)) that overlaps the observation window \( W \). The operator \( \text{Points}_j[X] \) counts the number of points of type \( j \) inside an area \( X \), and the operator \( \text{Area}[X] \) determines the area of \( X \).

Given the previous definitions, the first task is to derive an estimator for the overall neighborhood density function of conspecifics that integrates over all species \( S \). This can be accomplished by considering the neighborhood density \( \hat{O}_{ij}^\text{intra}(r) \) for different species \( s \) as replicates. Thus, the replication here is given by different species that are located within the same plot rather than by one species found in several plots. By pooling the neighborhood densities for the \( S \) species \( s \), we obtain, by using aggregation formula 3.107, the overall (community average) intraspecific neighborhood density of type \( j \) points at distance \( r \) from type \( i \) points:

\[
\hat{O}_{ij}^\text{intra}(r) = \frac{\sum_{k=1}^{n_i^j} \text{Points}_j[R^s_k(r)] + \cdots + \sum_{k=1}^{n_i^S} \text{Points}_j[R^s_k(r)]}{\sum_{k=1}^{n_i^j} \text{Area}[R^s_k(r)] + \cdots + \sum_{k=1}^{n_i^S} \text{Area}[R^s_k(r)]}
\]
Note that the estimator given in Equation 3.119 looks only at conspecific pairs of points within the community.

The next task is to derive the corresponding estimator \( O_{ij}^{\text{inter}}(r) \) that considers only heterospecific pairs of points. This can be done by estimating the neighborhood density function \( O_{ij}^{\text{all}}(r) \) for all individuals in the plot, regardless of species, and subtract the estimator for \( O_{ij}^{\text{intra}}(r) \). By maintaining the same notation as used in the estimator of \( O_{ij}^{\text{intra}}(r) \) we find:

\[
\hat{O}_{ij}^{\text{all}}(r) = \frac{\left( \sum_{k=1}^{n_1} \text{Points}_j[R_1(r)] + \text{mix}_1 \right) + \cdots + \left( \sum_{k=1}^{n_S} \text{Points}_j[R_S(r)] + \text{mix}_S \right)}{\sum_{k=1}^{n_1} \text{Area}[R_k(r)] + \cdots + \sum_{k=1}^{n_S} \text{Area}[R_k(r)]}
\]

(3.120)

where the symbol \( \text{mix}_j \) refers to the count of all points at distance \( r \pm dr/2 \) of points of species 1 that are heterospecific (these pairs of points were disregarded in \( O_{ij}^{\text{intra}}(r) \)). The symbol \( \text{mix}_S \) refers to the count of all points at distance \( r \pm dr/2 \) of points of species \( S \) that are heterospecific. In a more simplified notation we thus obtain:

\[
\hat{O}_{ij}^{\text{all}}(r) = \frac{\text{Points} + \text{mix}}{\text{Area}}
\]

(3.121)

\[
\hat{O}_{ij}^{\text{intra}}(r) = \frac{\text{Points}}{\text{Area}}
\]

(3.122)

\[
\hat{O}_{ij}^{\text{inter}}(r) = \frac{\text{mix}}{\text{Area}} = \hat{O}_{ij}^{\text{all}}(r) - \hat{O}_{ij}^{\text{intra}}(r)
\]

(3.123)

\( \hat{O}_{ij}^{\text{all}}(r) \) can be easily estimated by pooling the individuals of all species, distinguishing only between surviving and dead individuals, and applying Equation 3.118. \( \hat{O}_{ij}^{\text{intra}}(r) \) can be estimated by joining the results of the analysis obtained for individual species with Equation 3.119.

We exemplify this method with a reanalysis of data presented in Raventós et al. (2010). In this study, a long-term data set was analyzed that comprised fully mapped seedling emergence and subsequent survival in a Mediterranean gorse shrubland after experimental fires. The experimental fires killed all plants and restarted the system though the germination processes. The analyses involved surviving and dead individuals of the four most common species Cistus albidus, Helianthemum marifolium, Ulex parviflorus, and Ononis fruticosa obtained one, two, three, and nine years after the fire (Raventós et al. 2010). Figure 3.41 shows results from the first year after fire for the “fire only” treatment (corresponding to Figures 2A and 2I in Raventós et al. 2010). We used
Equation 3.122 to estimate the neighborhood densities of intraspecific pairs of plants and Equation 3.123 to calculate the neighborhood densities for interspecific pairs of plants. The intraspecific neighborhood density $O_{22}(r)$ of surviving plants around surviving plants (Figure 3.41b, closed disks) showed strong aggregation. This is indicated by a neighborhood density at short distances $r$ that is approximately three times higher than at greater distances. However, the interspecific neighborhood density $O_{22}(r)$ of surviving plants around surviving plants (Figure 3.41b, open circles) exhibits almost no dependence on distance. Thus, the neighborhood density of surviving plants around conspecific surviving plants is much higher than the neighborhood density of surviving plants around heterospecific surviving plants. The same pattern is shown by dead individuals (Figure 3.41a), but is somewhat weaker. Interestingly, the mixed pairs of surviving and dead plants showed no clustering for interspecific and intraspecific pairs (Figure 3.41c,d). Thus, surviving plants (and to a lesser extent dead plants) occur in intraspecific clusters, but the clusters of different species rarely overlap. These results point to positive intraspecific association among surviving plants (Raventós et al. 2010).

### 3.2.2.3 Nonparametric Estimation of a Pollen Dispersal Kernel

This example is given by a study by Niggemann et al. (2012) who estimated pollen dispersal kernels and contrasted the observed kernels to those
expected under a random mating null model. The study analyzed a complex quantitatively marked pattern and the aggregation formulas were used to implement specific null models that would otherwise require specialized software. Seeds were harvested from female black poplar trees (*Populus nigra* L.) and genetic paternity analysis was used to determine the pollen donor (i.e., male parent) for each harvested seed from among all male poplar trees in the study area. (The poplar population comprised a dense patch and a few scattered individuals further away.) The basic interest in analyzing this data type was to find out if pollen dispersal occurred in a spatially correlated way (i.e., pollen of nearby males had a higher probability of reaching a female than pollen from males located further away) and to quantify the distance dependence. To answer these questions, parametric dispersal kernels are usually fitted to such data sets (e.g., Robledo-Arnuncio and García 2007); however, for assessment of uncertainty with respect to the random mating null model (no distance dependence), or any other model, it is desirable to calculate nonparametric dispersal kernels. This can be done within the framework of quantitatively marked point patterns and mark-correlation functions (Section 3.1.7).

The data structure involves mother trees *m*, potential father trees *f*, and marks $m_{mf}$, which are the number of seeds of mother tree *m* fathered by male *f* (Figure 3.42). The null model assumes random mating, where all potential pollen donors *f* have the same probability of fathering a seed of mother *m*. To determine how the observed values of the marks $m_{mf}$ depend on the distance *r* to the mother *m*, Niggemann et al. (2012) developed a bivariate mark-correlation function:

$$\hat{k}_{mf}(r) = \frac{1}{\hat{c}_f} \frac{\sum_{m=1}^{n_m} \sum_{f=1}^{n_f} t(m_{mf}) k(\|x_f - x_m\| - r)}{\sum_{m=1}^{n_m} \sum_{f=1}^{n_f} k(\|x_f - x_m\| - r)}$$  (3.124)

where $n_m$ and $n_f$ are the total number of mother trees and potential pollen donors, respectively; $t(m_{mf}) = m_{mf}$ is the test function with associated normalization constant $c_f$ (the mean of the mark $m_{mf}$ taken over all *f*–*m* pairs of female and male trees); $x_f$ and $x_m$ are the location of the pollen donor *f* and the mother tree *m*, respectively; and $k(\|x_f - x_m\| - r)$ is the box kernel (Equation 3.15; Figure 3.7a) with bandwidth $h$. The kernel function $k()$ basically defines a ring with radius *r* and width $2h$ centered on the *m*th mother tree, and the sum $\sum_{f=1}^{n_f} k(\|x_f - x_m\| - r)$ counts the number of potential pollen donors *f* within such a ring (Figure 3.42a).

The interpretation of the mark-correlation function (3.124) is straightforward. All mother trees *m* are visited in sequence, the marks $m_{mf}$ of all potential donors *f* are counted at distance *r* from the mother *m* [i.e., the inner sum $\sum_{f=1}^{n_f} t(m_{mf}) k(\|x_f - x_m\| - r)$] and the number of potential donors *f* at distance *r* from the mother tree *m* are counted [i.e., $\sum_{f=1}^{n_f} k(\|x_f - x_m\| - r)$]. Thus, $k_{mf}(r)$
is proportional to the average number of seeds sired by male trees located at distance $r$ from a representative mother tree and is therefore proportional to the pollen probability density $f(r)$ at distance $r$ from the location of the dispersing father tree (i.e., the dispersal kernel; Clark et al. 1999). Note that no edge correction is required for mark-correlation functions (Illian et al. 2008).

Because the mark $m_{mf}$ depends on the mother tree $m$, an implementation of the random mating null model must randomize the mark individually for each female. In practice, the null model randomly shuffles, for a given female parent $m$, the marks $m_{mf}$ (i.e., number of seeds of female parent $m$ fathered by male $f$) over all males $f$. This procedure must be independently repeated for each female parent $m$. Implementation of this null model would, therefore, require specialized software. However, when using the aggregation formula for mark-correlation functions (based on Equation 3.107), the analysis can be done using the standard procedure for mark-correlation functions.

**FIGURE 3.42**
Nonparametric estimation of pollen dispersal kernels. (a) Data structure for estimation of pollen dispersal kernel, which consists of one focal female tree $m$ (closed disk), a ring of distance $r$ and width $2h$ around the focal female, and potential pollen donors $f$ (open disks) lying in the ring, marked by the number of seeds $m_{mf}$ they sired with the female $m$ (only for trees inside the ring). (b) Estimate of the observed pollen dispersal kernel and simulation envelopes of the random mating null model. (c) The relationship between the width of the simulation envelope and the number of father–mother pairs. (d) The number of father–mother pairs for a ring of distance $r$. (Adapted from Niggemann, M. et al. *Journal of Ecology* 100: 264–276.)
To run the analysis, the data for each mother are formally treated as replicates and the data set comprises the coordinates of the mother (this is the focal pattern) and the coordinates of all male trees with the attached mark \( m_{mf} \) (i.e., number of seeds of female parent \( m \) fathered by male \( f \)). Conducting a separate analysis for each female guarantees that the simulation of the random mating null model keeps the identity of the mother tree for the marks \( m_{mf} \). The analysis for mother tree \( m \) yields the (nonnormalized) estimator

\[
\sum_{f=1}^{n_f} t(m_{mf}) k(\|x_f - x_m\| - r) \\
\sum_{f=1}^{n_f} k(\|x_f - x_m\| - r)
\]  

(3.125)

Subsequent aggregation of the replicates (i.e., females), using the principle in Equation 3.107, yields Equation 3.124 where the normalization constant \( \hat{c}_{mf} \) becomes \( \hat{c}_{mf} = \sum_{m=1}^{n_m} \sum_{f=1}^{n_f} t(m_{mf}) / n_m n_f \). The resulting dispersal kernel estimated using distance intervals of 100 m and a ring width of \( h = 200 \) m is shown in Figure 3.42b. There is a clear signal of distance dependence at distances less than 400 m, but at larger distances the simulation envelopes become quite wide. This is especially the case for distances greater than 2000 m, where the number of data points within each ring (i.e., the number of father–mother pairs) becomes very low (i.e., below 10; Figure 3.42d). Clearly, the data at these distances are not sufficient, prohibiting the discrimination of any reasonable kernel function and the result is completely overpowered by stochasticity. While this simple fact becomes clear when looking at the simulation envelopes, it is not that clear in “traditional” kernel fitting. While lack of data in fitting long-distance dispersal kernels is a well-known problem, it seems to be not fully recognized how serious this problem is. The approach of point-pattern analysis in deriving nonparametric dispersal kernels and the associated simulation envelopes can also be used to estimate the data demand for a certain level of uncertainty (i.e., width of simulation envelopes). In our case, we would need roughly more than 80 father–mother pairs within a given ring of distance \( r \) to make reasonably good inferences. With less data, the envelopes become quite wide (Figure 3.42c).

### 3.2.2.4 Blowouts and Waterholes

Blanco et al. (2008) used the aggregation formula for the O-ring statistic (Equation 3.119) to implement a specific null model. Their study was concerned with an assessment of the impact of grazing on soil erosion processes prevailing in arid rangelands. Blowouts are depressions or hollows formed by wind erosion on a preexisting sand deposit in vegetation-stabilized dune fields. In arid zones, they are among the most common aeolian erosional
landforms in dune landscapes. Vegetated parabolic dunes are leeward extensions of blowouts. Plants colonize, and may eventually stabilize the trailing arms of the dunes, which are formed by the advancing dune apex. Grazing, however, can contribute to destabilization of dunes. The study by Blanco et al. (2008) investigated the neighborhood density of blowouts around waterholes in a vegetated dune field of Peninsula Valdés, in the Patagonia region of Argentina. Their goal was to analyze changes induced by grazing on soil erosion processes. The spatial pattern of blowouts was used as an indicator of erosion intensity. They used the neighborhood density $O_{12}(r)$ of blowouts (i.e., type 2 points) at distance $r$ away from waterholes (i.e., focal points of type 1) as a summary statistic to identify scales at which the spatial pattern of blowouts was significantly aggregated around water points (Figure 3.43a). The hypothesis was that large numbers of livestock concentrating around waterholes might destroy vegetation cover, resulting in dune reactivation and consequently generating blowouts. Accordingly, the impact of livestock on soil erosion processes should be greatest in sites near waterholes relative to those farther away from waterholes (i.e., a utilization gradient). This gradient should result in aggregation of blowouts near waterholes.

The study area comprised eight paddocks, four of which were lightly grazed, four heavily grazed, and each paddock contained one waterhole. The

**FIGURE 3.43**
Use of the aggregation formula for the $O$-ring statistic in implementing specific null models. (a) Partial map of the study area, showing two complete paddocks. Each paddock has one water hole (open circle), the area of the dune crests are indicated by darker gray, and the blowouts are shown as black squares. (b) $O$-ring analysis for lightly grazed paddocks. In a first step, an analysis was done for each paddock separately, using techniques for irregularly shaped study areas. The neighborhood density $O_{12}(r)$ of blowouts at distance $r$ away from the water holes was determined and the null model randomized the blowouts (black squares) over the dune crests (dark gray areas). The results of the four analyses for the individual paddocks were then combined using the aggregation formula for the $O$-ring statistic. Analysis of separate paddocks was necessary because the sheep cannot move from one paddock to another. (c) Same as panel (b), but for the heavily grazed paddocks. (Adapted from Blanco, P.D. et al. 2008. Rangeland Ecology and Management 61: 194–203.)
paddocks had an irregular shape but were adjacent (Figure 3.43a). Because sheep could not move among paddocks the grazing treatment within one paddock had no influence on the dune crests and the formation of blowouts of neighboring paddocks. As a consequence, Blanco et al. (2008) divided the study area into four replicate plots for each treatment and conducted separate analyses for each paddock. The waterhole of a paddock (open disks) was the focal point and the blowouts within ring segments centered on the respective paddock were counted (black squares). Consequently, the null model distributed the blowouts randomly on dune crests of a given paddock (which are the areas susceptible to erosion; dark gray areas). However, because sheep could not move from one paddock to another, the null model needed to randomize the blowouts only inside paddocks.

Similar to the example in the preceding section, the analysis was first done separately for each paddock, and then the results of the four replicate paddocks that were lightly grazed and the four paddocks that were heavily grazed were combined using the principle of Equation 3.107. Aggregation yielded an estimator similar to that shown in Equation 3.119. As expected, the lightly grazed paddocks showed only a weak distance dependence of blowouts around waterholes (Figure 3.43b), whereas the heavily grazed paddocks showed a strong attraction of blowouts up to 400 m away from waterholes, with a clear peak in the neighborhood density at 200 m, which for sheep is the walking distance away from a water hole (Figure 3.43c).

### 3.3 Superposition of Point Processes

Point patterns in ecology may be the result of the independent superposition of two or more different mechanisms. A good example illustrating this can be provided by examining the effects of animal dispersal on the seed distributions of tropical trees. As Wiegand et al. (2009) note, this may have consequences for the resulting spatial pattern of recruits: “Howe (1989) pointed to two contrasting seed deposition patterns of animal dispersed tree species: scatter-dispersal and clump-dispersal. The behavior of frugivores such as small birds or bats that regurgitate, spit, or defecate seeds singly may lead to isolated recruits. Conversely, large frugivores that defecate seeds in masses may cause aggregated recruit patterns (Howe 1989). Additionally, secondary dispersal by central-place foragers such as ants may increase clumping of seed deposition by depositing seeds from a wide area within nest or its refuse piles (Passos and Oliveira 2002). Scattered vs. clustered seed deposition may occur for plant species with a diverse assemblage of dispersers, but can even be caused by two different behavioral patterns in a single frugivore species (e.g., Russo and Augspurger 2004), p. E107.” Thus, point patterns may be a superposition of two (or more) component patterns. In the example of
recruits of tropical tree species given by Wiegand et al. (2009), we may expect a random component pattern generated by scatter-dispersal and a clustered component pattern caused by clump-dispersal. Calculating the summary statistics of superposition patterns is a powerful way to better understand the nature of complex spatial patterns. Illian et al. (2008, p. 371) provide formulas for the superposition of summary statistics. Under certain conditions, the summary statistics that result from the superposition of the two-point processes can be calculated, if the summary statistics of the two independent point processes are known. In the following sections, we present the superposition formulas for important summary statistics and provide examples.

### 3.3.1 Intensity Function

The superposition of two independent (but not necessarily homogeneous) point processes with intensities \( \lambda_1(x) \) and \( \lambda_2(x) \) yields the sum of the intensity functions of two individual intensity functions:

\[
\lambda(x) = \lambda_1(x) + \lambda_2(x)
\]

### 3.3.2 Second-Order Statistics

If two-point processes are independent, stationary, and isotropic, the superposition of the two pair-correlation functions \( g_1(r) \) and \( g_2(r) \) yields

\[
g(r) = \frac{\lambda_1^2 g_1(r) + 2\lambda_1 \lambda_2 + \lambda_2^2 g_2(r)}{(\lambda_1 + \lambda_2)^2}
\]

The above formula can be understood based on product densities. Consider two disjoint disks of infinitesimal magnitude \( dx_1 \) and \( dx_2 \) centered on location \( x_1 \) and \( x_2 \), respectively. The product density \( \rho^{(2)}(r) \ dx_1 dx_2 = (\lambda_1 + \lambda_2) g(r) dx_1 dx_2 \) gives the probability that one point of a given point process occurs within the first disk and a second point occurs within the second disk. In the superposition pattern we have four possibilities: (i) the first point (at \( x_1 \)) is of type 1 and the second point (at \( x_2 \)) is of type 1; (ii) the first point is of type 1 and the second point is of type 2; (iii) the first point is of type 2 and the second point is of type 1; and (iv) the first point is of type 2 and the second point is of type 2. If the patterns are stationary, the (partial) product density of type 1 points around type 1 points yields \( \lambda_1^2 g_1(r) \), that of type 1 points around type 2 points yields \( \lambda_1 \lambda_2 g_{12}(r) \), that of type 2 points around type 1 points yields \( \lambda_2 \lambda_1 g_{21}(r) \), and that of type 2 points around type 2 points yields \( \lambda_2^2 g_{22}(r) \). Because the two patterns are independent we have \( g_{12}(r) = g_{21}(r) = 1 \), and summing up all four cases and dividing by \( \lambda_2 \) yields Equation 3.127.
Integrating Equation 3.127 yields the superposition of the two $K$-functions $K_1(r)$ and $K_2(r)$:

$$K(r) = \frac{\lambda_1^2 K_1(r) + 2\lambda_1\lambda_2 \pi r^2 + \lambda_2^2 K_2(r)}{(\lambda_1 + \lambda_2)^2} \quad (3.128)$$

If the two component processes are homogeneous Poisson processes with $g_1(r) = g_2(r) = 1$ and $K_1(r) = K_2(r) = \pi r^2$ it can be easily verified that the superposition process yields $g(r) = 1$ and $K(r) = \pi r^2$.

### 3.3.3 Spherical Contact Distribution

If two-point processes are independent, stationary, and isotropic, the superposition of the two spherical contact distribution functions $H_{s,1}(r)$ and $H_{s,2}(r)$ yields

$$H_s(r) = 1 - [1 - H_{s,1}(r)][1 - H_{s,2}(r)] \quad (3.129)$$

where $H_{s,1}(r)$ and $H_{s,2}(r)$ are the spherical contact distributions of the two-component processes. The formula can be understood with the following arguments. The quantity $1 - H_s(r)$ is the probability that a test point has no neighbor of pattern 1 within distance $r$ and no neighbor of pattern 2 within distance $r$. If the two patterns are independent, this probability can be separated into the product of the two separate probabilities, that is, the test point has no neighbor of pattern 1 within distance $r$, which yields $1 - H_{s,1}(r)$, and the probability that the test point has no neighbor of pattern 2 within distance $r$, which yields $1 - H_{s,2}(r)$. This leads directly to Equation 3.129.

### 3.3.4 Nearest-Neighbor Distribution Function

If two-point processes are independent, stationary, and isotropic, the superposition of the distances to the nearest neighbors of the two distributions $D_1(r)$ and $D_2(r)$ yields

$$D(r) = 1 - \frac{1}{\lambda_1 + \lambda_2} (\lambda_1(1 - D_1(r))(1 - H_{s,2}(r)) + \lambda_2(1 - H_{s,1}(r))(1 - D_2(r))) \quad (3.130)$$

The formula can be understood analogously to that of the spherical contact distribution when considering the probability $1 - D(r)$ that there is no point of the superposition pattern within distance $r$. In this case, a type 1 point has no type 1 neighbor and no type 2 neighbor within distance $r$. The probability that a type 1 point has no type 1 neighbor within distance $r$ is given by $[1 - D_1(r)]$. 

© 2014 Taylor & Francis Group, LLC
If we think of a type 1 point as a test point for the spherical contact distribution of point type 2, the quantity \((1 - H_2(r))\) gives the probability that the type 1 (test) point has no type 2 neighbor within distance \(r\). Thus, if the patterns are independent, the joint probability yields \([1 - D_1(r)][1 - H_2(r)]\). The argument for type 2 points is analogous. If we now sum up the respective probabilities for all points, Equation 3.130 arises. To do this we need to weight the contribution of each pattern with its respective intensity (e.g., the contribution to the distribution function due to type 1 points needs to be weighted by \(\lambda_1\)).

### 3.3.5 Examples of the Superposition of a Thomas Process with a Random Pattern

Application of formula 3.127 for the independent superposition of a Thomas process, with \(g_1(r)\) comprising a proportion \(c\) of all points of the joined pattern, with a random pattern yields

\[
g_{1+2}(r) = 1 + \frac{c^2 \exp(r^2/4\sigma^2)}{\rho \frac{4\pi\sigma^2}{\pi}} \tag{3.131}
\]

where \(c\) is the proportion of the points associated with the Thomas process. We can understand Equation 3.131 by recalling that the pair-correlation function of the Thomas process (Box 2.5) is given by

\[
g_1(r) = 1 + \frac{1 \exp(r^2/4\sigma^2)}{\rho \frac{4\pi\sigma^2}{\pi}} \tag{3.132}
\]

and the pair-correlation function for a random pattern is \(g_2(r) = 1\). Thus, the resulting pair-correlation function 3.131 is exactly the same as that of the component Thomas process (3.132), except that the parameter \(\rho\) which characterizes the density of cluster centers results in the superposition pattern \(\rho_{1+2} = \rho / c^2\). This means that the pair-correlation function alone cannot distinguish between the pure Thomas process and the superposition of a Thomas process with a random pattern. However, the fit with the pair-correlation function will overestimate the real number of clusters by a factor of \(1/c^2\), where \(c\) is the proportion of the points belonging to the cluster component pattern.

To illustrate the values of the summary statistics under superposition we use the pattern shown in Figure 3.44, which was generated by a Thomas process within a \(500 \times 500 \text{ m}^2\) grid (with cluster radius of 10 m and 100 randomly distributed cluster centers). We then superposed different numbers of random points with this pattern. As expected, the values of the pair-correlation function decrease with the addition of random points (Figure 3.44b). The spherical contact distribution also shows a clear and strong response to the addition of random points. Because the random points fill gaps, the probability of finding a point of the superposition pattern within distance \(r\) from a test location increases substantially with the addition of random points.
The distribution function $D(r)$ of the distances to the nearest neighbor shows a more complex behavior. First, when adding 100 random points, $D(r)$ decreases at larger scales, because many random points will be located within the gaps of the cluster component pattern and will have their nearest neighbor at larger distances. However, addition of more than 100 random points does not change $D(r)$ much, once all the gaps are filled. In this case, additional random points will only rarely be the nearest neighbor of an existing point of the superposition pattern. This indicates that using the spherical contact distribution as an additional summary statistic, to complement the pair-correlation function, can provide important additional information on the nature of the pattern.

Figure 3.44 shows a similar situation, but now the number of points is fixed ($n = 500$) and a certain number of points of the cluster pattern are removed and replaced by random points. As expected, the pair-correlation function decreases steadily. When random points replace 80% of the original clustered points, the signal of clustering almost disappears (Figure 3.45b). The change
in shape of the distribution function $D(r)$ of the distances $r$ to the nearest neighbor is also interesting (Figure 3.45c). A low proportion of random points (i.e., 20%) affects mostly the values of $D(r)$ at intermediate distances $r$ (say between 7 and 12 m). This is due to the replacement of points from clusters (which usually have short nearest-neighbor distances) by random points. The random points are mostly placed in gaps of the cluster pattern and therefore yield larger distances to the nearest neighbor. However, higher proportions of random points do not change $D(r)$ at distances greater than 18 m, but occur mostly at smaller distances (say 5 m). This is because many small neighbor distances disappear due to removal of points from the cluster pattern.

The spherical contact distribution exhibits a more continuous response to replacement of clustered points by random points. In all cases, the maximal change occurs at distances between 15 and 35 m (Figure 3.45d). This is understandable because the gaps of the cluster pattern are continuously filled by random points. As a result, the spherical contact distribution is a better summary statistic to use in supplementing the pair-correlation function than is the nearest-neighbor distribution function.

**FIGURE 3.45**
Impact of randomly displacing points of a Thomas process. (a) Maps of the superposition of a Thomas process (top left), with a random pattern where 20%, 40%, and 80% of the points of the original cluster pattern are replaced by random points. Panels (b)–(d) show the resulting summary statistics.
3.3.6 Examples of Superposition of Two Thomas Process

Another important case of the superposition of patterns is given by the independent superposition of two different Thomas processes. The pair-correlation function can be constructed from the parameters for each Thomas process, that is, $\sigma_1$, $\sigma_2$, $\rho_1$, and $\rho_2$, and information regarding the proportion of points associated with each process (Stoyan and Stoyan 1996):

$$g_{1+2}(r) = 1 + \frac{c^2}{\rho_1} \exp\left(\frac{r^2}{4\sigma_1^2}\right) + \frac{(1-c)^2}{\rho_2} \exp\left(\frac{r^2}{4\sigma_2^2}\right)$$ (3.133)

In this case, $c$ is the proportion of points associated with Thomas process 1. If the two Thomas processes are the same, the random superposition yields

$$g_{1+2}(r) = 1 + \left(\frac{c^2 + (1-c)^2}{\rho}\right) \exp\left(\frac{r^2}{4\sigma^2}\right)$$ (3.134)

For the case where $c = 0.5$ (i.e., both Thomas processes contain the same number of points), we find that $c^2 + (1-c)^2 = 0.5$, thus the resulting pair-correlation function yields a Thomas process with double the number of cluster centers (which is indeed the case since the clusters of a Thomas process are independently placed). Wiegand et al. (2007c, 2009) present applications of the superposition of Thomas processes to recruits of tropical trees. We will treat these and similar examples in detail in the following chapter on univariate point patterns (Section 4.1.4).

3.4 Toolbox

This chapter presents different techniques and methods that are especially useful in a variety of applications, but not specific to any given data type. They include general methods for detecting clusters, the analysis of irregularly shaped observation windows, and pattern reconstruction.

3.4.1 Cluster Detection Algorithm

Identifying and characterizing clusters is a task that is of interest in several applications of point-pattern analysis (e.g., Coomes et al. 1999; Plotkin et al. 2002). Here, we present a simple algorithm that can be used to detect clusters and subsequently describe their properties, such as the “center of mass” or the number of points belonging to each cluster.
The purpose of a cluster algorithm is to join together ecological objects idealized as points into successively larger clusters, using a measure of distance between objects. In this section, we present a point-pattern-based approach of cluster analysis (Sneath and Sokal 1973) that identifies cluster centers and points belonging to these clusters. Initially, we have a univariate point pattern given as a map, with the positions of points within an observation window $W$. Each initial point is called a “cluster,” even if it is initially composed of a single point. In later steps, the points are replaced by clusters, which contain several of the original points of the pattern. Each cluster is characterized by the number of points it contains and by the cluster center, which is the center of gravity of all points of the cluster.

In the first step, the algorithm determines the pair of cluster centers that have the smallest distance among all pairs of cluster centers. The points of the two clusters are then amalgamated into one new cluster, and the two original clusters are deleted. Different rules can be used to amalgamate clusters. Here, we use a rule that maintains the identity of all points, but uses the “center of gravity” of all points of the cluster as coordinates of the new cluster. The algorithm proceeds by searching for the pair of cluster centers of the updated pattern that has the shortest separation distance among all cluster centers. It then amalgamates these two clusters and uses all points belonging to the new “cluster” for determining the new mean position (which is the center of gravity). In this way, the number of clusters is reduced by one in each step. The same procedure is repeated as long as cluster centers can be found that have a separation distance less than a predefined maximal distance $r_c$, which is the basic parameter of the algorithm.

Depending on the value of $r_c$, the size of the clusters will vary because the distance $r_c$ is the minimal distance between two cluster centers. Larger values of $r_c$ yield fewer, but larger clusters. Note that this cluster algorithm uses an amalgamation rule called “unweighted pair-group centroid” because the coordinate (or centroid) of a cluster is the center of gravity of all points contained in the cluster (Sneath and Sokal 1973). Cluster analysis is not restricted to point patterns, it can be used for any objects for which we can define a distance or similarity index. Many other cluster algorithms exist that differ in their amalgamation rules and distance definition. For example, the distance between two clusters may be defined as the distance between the two nearest or furthest points of the two different clusters, or as the average distance between all pairs of objects in the two clusters. Additionally, the number of objects contained in the clusters may be used to weight the distance measure. However, the unweighted pair-group centroid generally works well for the purpose of identifying clusters in point patterns.

The output of the cluster algorithm is a list of cluster centers and points that belong to each corresponding cluster. This information can be used to characterize the pattern with respect to the distribution of cluster sizes and the frequency of points within a given distance to the cluster center. This information can also be used to construct a classification tree.
for specific applications it may be important to replace the artificial cluster centers by a point of the pattern. For example, we may use the point of the cluster, which is closest to the cluster center. In this way, each cluster can be replaced by a representative point of the cluster for analysis.

One application of the cluster detection algorithm is de-clustering of a pattern for nonparametric intensity estimation (Box 2.4). The intensity function reflects the impact of first-order effects on the position of points. For example, if the points are trees of a given species in a tropical forest, the intensity function may provide a good description of habitat suitability: some areas of the plot will be better suited for the tree species and have a higher probability of hosting a tree than other areas. This leads to locally elevated point densities. However, other mechanisms, such as clumped or limited seed dispersal, which are independent of habitat suitability, may partly override the signal of habitat suitability and create additional areas of elevated point densities. Such second-order clustering is not caused by better environmental conditions, but by a mechanism that deposits many seeds at a given location. Due to a “mass effect” a location may host more recruits than expected by its habitat suitability. Thus, second-order clustering may “contaminate” nonparametric estimates of the intensity function. This phenomenon has been known for a long time and is an important issue in studies using parametric estimation techniques to develop habitat and species distribution modeling (e.g., Dorman et al. 2007). The main issue, in the context of parameter estimation, is that second-order clustering causes a form of pseudo-replication, because nearby points contain similar information.

The cluster detection algorithm can be used to remove strong small-scale clustering. This is an important step before applying a nonparametric kernel estimate of the intensity function, if it is suspected that the strong clustering is due to factors other than localized patterns of favorable habitat conditions. In the first step, one should analyze a plot of the pair-correlation function to identify an initial value of the cluster detection parameter \( r_c \) (Figure 3.46a). Several values of \( r_c \) should be examined to find the appropriate parameter value. After application of the cluster algorithm, clusters are replaced by the point of the cluster that lies closest to the cluster center. In this way, each cluster is replaced by one representative point. We call this de-clustering.

Figure 3.46 shows an example of this approach. The full pattern is shown in Figure 3.46b, and the resulting pair-correlation functions for the parameters \( r_c = 0, 1, 2, 3, 4, \) and 5 m are shown in Figure 3.46a. The pair-correlation function of the original pattern (open squares in Figure 3.46a) shows strong clustering at small distances, with a neighborhood density at 1 m being 25 times as high as expected under a random pattern. Most of the small-scale clustering in the pair-correlation functions is removed only when minimal distances of \( r_c = 4 \) (open triangles) or 5 m (closed squares) are used in declustering. The sequence of Figure 3.46c through f shows how the declustering works as the maximal distance \( r_c \) is increased. Indeed, the small clusters are still visible for parameter \( r_c = 2 \) and 3 m (Figure 3.46c,d). Ideally, the selection
of the parameter $r_c$ should be guided by biological information on the scales at which the environment typically varies and on the scale of potential clustering mechanisms. Otherwise, one may remove too much clustering from the pattern and underestimate the variability in habitat suitability.

### 3.4.2 Analyzing Irregular Observation Windows

In most applications of point-pattern analysis the observation window $W$ has a rectangular shape, or sometimes a circular shape. The selection of regular shapes is not entirely driven by esthetic considerations, but is more a

---

**FIGURE 3.46**

Application of a de-clustering algorithm. (a) The pair-correlation functions of the observed and de-clustered patterns, varying the parameter $r_c$, the maximal distance between clusters. (b) Observed pattern. (c–f) De-clustered patterns with different parameter values for $r_c$. See Section 3.4.1 for a detailed explanation.
function of the practical issue of edge correction, which is especially important for calculating the pair-correlation and \( K \)-functions. For example, the popular Ripley edge correction (Section 3.1.2.1) is based on a weighting factor that is given by the circumference of a disk with radius \( r \) divided by the circumference of the disk that overlaps \( W \). Relatively straightforward analytical formulas exist for calculating the Ripley weights for rectangular and circular observation windows. However, their calculation is nontrivial if the observation window has an irregular shape. An analytically tractable method was not introduced until 1999, when Goreaud and Pélissier (1999) presented a method that approximated the real shape of an irregularly shaped observation window \( W \) either by a polygonal observation window or by removing a polygonal area inside a rectangle.

Because of the difficulties produced by having irregularly shaped observation window, most observation windows for point-pattern data are placed so that they can retain a rectangular shape. However, as outlined in Section 2.6.3.2, it might be necessary to exclude some areas from an initially rectangular observation window to retain a homogeneous pattern (e.g., Figures 2.25, 2.28, 2.30). In other cases, the area has an irregular shape that cannot be avoided, such as the paddocks in the example of blowouts around water holes (Figure 3.43). In the following, we present two methods that can deal with edge correction for pair-correlation and \( K \)-functions and irregular observation windows. We also recommend that the interested reader examine the study by Goreaud and Pélissier (1999), which presents an extension of the Ripley edge correction method (Section 3.1.2.1; Figure 3.2) approximating observation windows by polygons.

### 3.4.2.1 Grid Approximation

Perhaps the simplest method in considering an irregular observation window is to use an underlying grid and represent the irregular observation window by small cells (Figure 3.47; Wiegand and Moloney 2004). Each cell can be either in or out of the observation window. The area of a ring or circle located inside the observation window \( W \) is then approximated by the underlying grid (Figure 3.47). We can then estimate the various point-pattern statistics accounting for edge effects in a straightforward fashion. For example, the estimators can account for the number of points lying inside the rings or circles around a focal point and also determine the number of cells within the circle or ring lying inside \( W \). The latter information can be used in calculating an edge correction empirically. In the case of the \( O \)-ring statistic, we can rewrite Equations 3.13 and 3.118 to yield

\[
\hat{O}_h(r) = \frac{\sum_{k=1}^{n} \text{Points}_h[R_k(r)]}{\sum_{k=1}^{n} \text{Area}[R_k(r)]} \quad (3.135)
\]
where $R_k(r)$ is a ring with radius $r$ around point $k$, which is approximated by the underlying grid (Figure 3.47). The operator $\text{Points}[X]$ counts the number of points of type $j$ in $X$, and the operator $\text{Area}[X]$ counts the number of cells in $X$. The corresponding estimator of the $K$-function yields

$$
\lambda K_{ij}(r) = \pi r^2 \frac{\sum_{k=1}^{n} \text{Points}[C_k(r)]}{\sum_{k=1}^{n} \text{Area}[C_k(r)]}
$$

(3.136)

where $C_k(r)$ is a circle of radius $r$ around point $k$, which is approximated by the underlying grid (Figure 3.47). This method allows for efficient approximation of the WM estimators for $\lambda g(r)$ and $\lambda K(r)$ (3.17, 3.36), if the cell size is sufficiently small. However, the calculation of the estimator will be quite slow, if the grid is too fine. Figure 3.48a shows a random pattern inside an irregular observation window where a circular area is missing in the center. We approximated this irregular shape with an underlying grid with a mesh width of 0.5 m. Figure 3.48b shows the results of an analysis with 199 simulations of the CSR null model and the pair-correlation function with a ring width of 2 m.

### 3.4.2.2 Using Inhomogeneous Functions

A second approach for analyzing irregularly shaped observation windows is provided by the estimators of the inhomogeneous pair-correlation and
K-functions, based on the Ohser or WM edge correction methods (Equations 3.29, 3.31, 3.41, 3.42). This approach represents the irregularly shaped observation window by using the intensity function \( \lambda(x) \): if the location is inside the observation window, we have \( \lambda(x) = \lambda \) (where \( \lambda \) is the intensity of the pattern inside the irregularly shaped observation window), otherwise we have \( \lambda(x) = 0 \). Clearly, with this approach we will still approximate the intensity function with an underlying grid, but this grid can be made as fine grained as the resolution of the data and does not slow down the estimation as with the grid-based estimator above. Figure 3.48c shows the results of this approach, which uses an underlying grid with a mesh width of 0.5 m for the intensity function and a ring width of 2.5 m in calculating the WM estimator.
There is basically no difference to the grid-based approach. We also calculated the \( L \)-function with both approaches (Figure 3.48d,e) and found essentially the same results.

### 3.4.3 Pattern Reconstruction

For a number of applications in the study of point patterns, it is useful to generate patterns with predefined properties. For example, null models produce patterns with known properties by holding certain aspects of a pattern fixed, while randomizing other aspects of the data (see Section 2.4). This is especially difficult in the case of bivariate or multivariate patterns. The null model of independence in bivariate patterns should conserve the observed univariate structure of the two component patterns, but only break their (potential) association. Although this has been recognized for quite a while (see, e.g., Lotwick and Silverman 1982 or Berman 1986), no satisfying solution has yet been found. The early solution proposed to solve this problem was the toroidal shift (Lotwick and Silverman 1982; Figure 2.4), where one of the two component patterns is shifted entirely by adding a fixed random vector to each coordinate and then wrapping it on a torus. The toroidal shift null model has two potential problems: first, it maintains most of the observed interpoint distances exactly, with no stochastic variability. Second, it may produce additional artifacts at the edges, were the patterns are wrapped. In contrast, the independence null model should produce stochastic replicates of one or both component patterns that have the same statistical properties as the observed pattern.

Another nonperfect solution suggested to achieve the desired properties is to use parametric, point-process models to fit the second component pattern and use realizations of the fitted, point process as a null model. This method works well if the point process represents a good fit to the pattern, which is, in general, not guaranteed. **Pattern reconstruction provides a solution to developing a bivariate null model of independence** (Wiegand et al. 2013a), because it can produce stochastic replicate patterns, which share predefined properties of the observed pattern. The predefined properties are summary statistics of the observed pattern. The nonparametric algorithm of pattern reconstruction can reproduce patterns that match the selected summary statistics as closely as desired. Thus, pattern reconstruction can be used for generating null model patterns.

Another important application of pattern reconstruction is assessment of the degree of information captured by a given summary statistic or a combination of summary statistics (Wiegand et al. 2013a). In Section 3.1, we presented many different summary statistics that capture certain aspects of the spatial structure of point patterns. Although the interpretation of most of the summary statistics is relatively straightforward, it is not clear how well a given summary statistic (or several summary statistics together) describe the potentially complex properties of point patterns. In
other words, if we use only the pair-correlation function to summarize the properties of a given point pattern, do we miss other properties of the pattern that may be important for our ecological question? We will certainly do so in many cases. For example, it is well known that several different point processes (such as cluster processes) have the same pair-correlation function. In this case, additional summary statistics, such as the distribution function to the nearest-neighbor distances or the spherical contact distribution, are required to differentiate from among the various candidate processes (e.g., Section 3.3.5). In particular, the adequate description of complex point patterns may require the use of several summary statistics simultaneously (e.g., trees in tropical forests can exhibit clustering at multiple scales; Condit et al. 2000; Wiegand et al. 2007c, 2009). While it is common practice for experienced spatial statisticians to employ multiple summary statistics (e.g., Illian et al. 2008, p. 214), ecological studies using point-pattern analysis evaluate their data mostly with only one or two. However, it should be noted that important properties of the spatial patterns may remain undetected, if nonappropriate summary statistics are used, and the inferences made would be much weaker. Thus, the question arises as to how many, and which, summary statistics are needed to fully describe the properties of a point pattern. One elegant way of answering this is to use pattern reconstruction.

The purpose of pattern reconstruction is to create stochastic replicates of observed patterns that closely approximate predefined summary statistics of the observed pattern. For pattern reconstruction, we use an inverse approach originally proposed in materials science (Rintoul and Torquato 1997; Torquato 1998). This approach uses optimization techniques to reconstruct a point pattern only with the limited information provided by a given combination of summary statistics. To this end, a variant of the simulated annealing algorithm (Kirkpatrick et al. 1983) is used to generate patterns that minimize the deviations between the summary statistics of the observed and reconstructed patterns. This idea has been translated by Tscheschel and Stoyan (2006) into point-pattern analysis and is illustrated in detail in Illian et al. (2008, p. 407ff).

The following sections follow largely Wiegand et al. (2013a).

### 3.4.3.1 Annealing Algorithm

Here, we describe the pattern reconstruction algorithm outlined in Tscheschel and Stoyan (2006) and Illian et al. (2008), but expand it for non-stationary patterns by conditioning on the intensity function $\lambda(x)$ (Wiegand et al. 2013a). The reconstruction method is a variation of simulated annealing (Kirkpatrick et al. 1983) that generates a series of patterns by trial and error, approaching the values of the summary statistics of the observed patterns more closely in each simulation step (Yeong and Torquato 1998).

Structural information on the observed pattern $\phi$ is measured by a set of $I$ functional summary statistics $f_i^\phi(x)$ (with index $i = 1\ldots I$), where the variable
represents distance \( r \) or neighborhood rank \( k \). During each simulation step \( t \), we also estimate the corresponding \( f_i^\psi(x) \) of the simulated pattern \( \psi \). This allows us to determine the deviation between \( f_i^\phi(x) \) and \( f_i^\psi(x) \) by means of the “partial energy”

\[
E_i^\phi(\psi_t) = \frac{1}{n_i} \sum_{b=1}^{n_i} \left[ f_i^\phi(x_b) - f_i^\psi(x_b) \right]^2
\]

(3.137)

where the variable \( x \) is evaluated at \( n_i \) discrete values \( x_p \). We divide by the number of measurements \( n_i \) and use the square root to obtain a direct measure of mean error. To combine the \( I \) partial energies \( E_i^\phi \) into a total energy \( E_{\text{total}}^\phi \), we need to normalize the values because their absolute values may vary greatly or depend on the intensity \( \lambda \). One approach to normalization is to give all summary statistics approximately the same importance in the reconstruction. In practice, one has to select the weights \( w_i \) so that the different summary statistics yield approximately the same value of \( E_i^\phi \), if the observed and simulated patterns are in a good agreement. By using the weights \( w_p \), the total energy at simulation step \( t \) yields

\[
E_{\text{total}}^\phi(\psi_t) = \frac{\sum_{i=1}^{I} w_i E_i^\phi(\psi_t)}{\sum_{i=1}^{I} w_i}
\]

(3.138)

The reconstruction of a pattern \( \phi \) starts with a random pattern \( \psi_0 \) that has the same number of points as \( \phi \). In each simulation step \( t \), a randomly selected point is tentatively removed and a new point with random coordinates is proposed instead. This new point is accepted if the total energy \( E_{\text{total}}^\phi(\psi_t) \) decreases, that is, \( E_{\text{total}}^\phi(\psi_t) < E_{\text{total}}^\phi(\psi_{t-1}) \), otherwise another new point is considered (Tscheschel and Stoyan 2006). Thus, the new pattern \( \psi_t \) is slightly more similar to the observed pattern than the previous pattern \( \psi_{t-1} \). The algorithm proceeds until a maximum number of steps is reached (e.g., 40,000 or 80,000) or until the total energy is less than a predefined threshold (e.g., 0.005) (Tscheschel and Stoyan 2006). An additional stopping rule may be added to avoid conducting a large number of steps in a configuration that shows no improvement. An index that controls the improvement in the fit relative to the number of iteration steps already conducted is the proportion of accepted points. If this proportion drops below a given threshold, the simulation stops. This “stochastic improvements-only” algorithm is able to find local minima with very small total energies and each simulation will end up in a different local minimum (the absolute minimum would be the observed pattern \( \phi \)). We, therefore, obtain stochastic replicate patterns with predefined properties provided by the \( I \) summary statistics \( f_i^\phi(x) \) of the observed pattern \( \phi \) (Tscheschel and Stoyan 2006).
Because the pattern reconstruction algorithm requires many simulation steps, an efficient calculation of the summary statistics is necessary. As proposed by Tscheschel and Stoyan (2006) and Illian et al. (2008, p. 414), only the part of an estimator that was affected by the exchange of one point is updated. Note that there is no need to use edge correction for the estimators of the pair-correlation function and the $K$-function. Although this is counterintuitive at first, it becomes understandable if we consider Ohser edge correction (Equation 3.21), which accomplishes edge correction by multiplying the naive estimator $\hat{g}^n(r)$ of the pair-correlation function by a factor $c(r)$, characterizing the expected bias due to edge effects. In this case, the estimator of the pair-correlation function can be written as $\hat{g}(r) = c(r)\hat{g}^n(r)$, where the correction factor depends only on distance $r$ and the geometry of the observation window, but not the pattern. However, because both, the estimator of the pair-correlation function of the observed pattern $\phi$ and that of the reconstructed pattern $\psi_t$ share the same edge correction factor $c(r)$, it can be factored out in Equation 3.137. This means that the edge correction factor $c(r)$ plays only the role of a weight when combining the partial energy of all summary statistics in Equation 3.138. Because the factor $c(r)$ increases with distance $r$, it has the effect of putting slightly more weight on the larger values of $\hat{g}^n(r)$ in the estimation of $E^\phi_{\text{total}}(\psi_t)$. However, this will have little effect on the reconstruction. Thus, reconstructions based on the naive estimator of $g(r)$ and $K(r)$ will produce the same results as reconstructions using the Ohser estimator. However, using the Ripley estimator will produce slightly different patterns than the Ohser or the WM estimator because the numerical values of $K(r)$ may differ (e.g., Figure 3.14). Note that we used the pair correlation function $g(r)$ not directly for pattern reconstruction, but the related quantity $2\pi rd\lambda g(r)$ which gives the mean number of points at rings with radius $r$ and width $dr$ around the points of the pattern. We used this quantity instead of $g(r)$ because it varies over a much reduced range, especially if the pattern shows strong small-scale aggregation (e.g., Figure 3.49).

**FIGURE 3.49**
Pair-correlation function $g(r)$ and the transformation $2\pi rd\lambda g(r)$ used for pattern reconstruction for the highly clustered species $C. insignis$ (shown as open disks in Figure 3.57a).
3.4.3.2 Reconstruction of Homogeneous Patterns

A homogeneous pattern is composed of stochastic repetitions of a typical point configuration within an entire observation window and the variability around the typical point configuration is only due to small stochastic fluctuations. An important aspect of the typical point configuration, the average number of points within or at distance \( r \) from the typical point, is captured by second-order summary statistics. Therefore, the pair-correlation function (and/or the \( K \)-function) will capture important aspects of homogeneous patterns. However, even homogeneous patterns may require additional summary statistics to describe the variability in the typical point configuration around the mean. For example, the same pair correlation function may occur for a pattern with all the points having approximately the same number of neighbors within distance \( r \) as there would be for a pattern where, for example, most points have only a small number of neighbors, but a few points have many neighbors. The variability in the number of neighbors can be captured in the reconstruction by incorporating the nearest to \( k \)th neighbor distribution functions in the algorithm.

The impact of these considerations is illustrated in Figure 3.50, where we reconstructed a pattern that resulted from an independent superposition of a homogeneous cluster pattern (196 points) and a homogeneous hyperdispersed pattern (197 points) (Figure 3.50a). If we reconstruct the pattern using only the pair-correlation function, the resulting pattern is a sort of “average” pattern that yields a perfect reconstruction of the pair-correlation function (Figure 3.50b), but lacks the pronounced clusters and hyperdispersion of the observed pattern (Figure 3.50a). Consequently, the distribution function \( D(r) \) of the distance to the nearest neighbor of the reconstructed pattern shows strong departures from the observed function because it does not represent the mixture of a clustered pattern with a hyperdispersed pattern (Figure 3.50g). However, if we reconstruct the pattern based only on the distribution function \( D(r) \) of the distance to the nearest neighbor, we obtain a perfect fit of \( D(r) \) (Figure 3.50h), but the reconstruction lacks the strong clusters of the original pattern, although it yields a satisfying representation of the hyperdispersed part of the pattern (Figure 3.50c). However, using the pair-correlation function and the distribution function \( D(r) \) together, we obtain a reconstruction that shows both the pronounced clusters and the hyperdispersion (Figure 3.50d). It also now matches both summary statistics used in reconstruction quite well (Figure 3.50i). It is notable that a similarly good reconstruction can be obtained by using the first 50 distribution functions \( D_k(r) \) of the distance to the \( k \)th neighbor (Figure 3.50e). This is reasonable to expect, since the \( D_k(r) \) taken together contain more information than the pair-correlation function and the \( K \)-function and we have the relationship

\[
\lambda K(r) = \sum_{k=1}^{\infty} D_k(r)
\]

(3.139)
In the observed pattern we find that the 40th nearest neighbor was at least 50 m away (which means that $D_k(r) = 0$ for $k > 40$ and $r < 50$ m). Therefore, the $K$-function is completely determined by the first 40 $D_k(r)$ up to distances of 50 m. Finally, using several summary statistics together [i.e., $g(r)$, $K(r)$, $H_s(r)$, and $D_k(r)$ with $k = 1,...,50$], the reconstruction does not improve much (Figure 3.50f). This suggests that each pattern may have, depending on its
degree of complexity, an “optimal” set of summary statistics. We will investigate this issue in a later section (Section 3.4.3.6).

Wiegand et al. (2013a) found that homogeneous patterns are usually well reconstructed by using the pair-correlation function $g(r)$, the spherical contact distribution $H_s(r)$, and some of the distribution functions $D_k(r)$ of the distance to the $k$th neighbor. Note that they weighted the $D_k(r)$ in Equation 3.138 such that all the $D_k(r)$ taken together were weighted to the same degree as each of the other summary statistics. Thus, the $D_k(r)$ were functionally treated in the reconstruction as one summary statistic (see also Tscheschel and Stoyan 2006). The reason for the good performance of the combination of $g(r)$, $D_k(r)$, and $H_s(r)$ in reconstructing homogeneous patterns is that the pair-correlation function controls the average neighborhood density, the distribution functions of the distance to the $k$th neighbor control the variability of the neighborhood density (i.e., the small-scale clustering that determines the number of neighbors individual points have within distance $r$), and the spherical contact distribution controls the “gaps” in the pattern.

### 3.4.3.3 Impact of Spatial Scale on Reconstruction

Successful pattern reconstruction depends not only on the selection of appropriate summary statistics, but also on the scales over which the summary statistics are evaluated. If a pattern contains nonrandom structures at distances greater than say 50 m, but the summary statistics (as in Figure 3.51) capture only information up to 50 m, the reconstruction will not be able to correctly reassemble the pattern at distances greater than 50 m. To demonstrate this point we generated 100 reconstructions of the pattern shown in Figure 3.50a, based on the pair-correlation function and the distribution functions of the $k$th neighbor $D_k(r)$ for $k = 1 \ldots 50$, using information from the summary statistics at distances up to $r = 50$ m. Figure 3.51a shows the results for the quantity $(2\pi r dr) \lambda g(r)$, which was used instead of the pair-correlation function as the functional summary statistic $f_i(r)$ in Equation 3.137. For distances $r$ smaller than 50 m, the values were virtually identical to the observed pattern, but at greater distances there was increasing variability. This is because the reconstruction algorithm did not use information at the larger scales (i.e., $r > 50$ m) and, as a consequence, stochastic effects influenced the reconstruction more with increasing scale. However, because the pattern was homogeneous and showed no larger-scale structures (except stochastic ones) the departures are relatively small and the observed function is the same as the average of the 100 reconstructions (Figure 3.51a). We find a similar result for the $L$-function (Figure 3.51d), but because the $L$-function emphasizes larger-scale effects, the stochastic effects at distances greater than 50 m result in increasingly wider envelopes, although the mean value of all reconstructions is close to the value observed for the actual pattern. The same was true for the mean
distance to the $k$th neighbor, which fits very well up to distances of 50 m, but is subject to stochasticity for the larger neighbor ranks (Figure 3.51f). As expected, the distribution function $D(r)$ to the nearest neighbor is well reconstructed (Figure 3.51b) and even the distribution functions to the 16th neighbor $D_{16}(r)$ are well matched in all reconstructions (Figure 3.51c). The spherical contact distribution $H_s(r)$, which was not used for reconstruction, is only slightly overestimated (Figure 3.51e).

The results shown in Figure 3.51 indicate that a homogeneous pattern can be reconstructed well, even with limited distance information. This is a consequence of being homogeneous and having no larger-scale dependencies. At larger distances we find only differences due to random stochasticity, rather than to larger-scale dependencies.
3.4.3.4 Reconstruction of Heterogeneous Patterns

Although the summary statistics used in pattern reconstruction are developed for homogeneous patterns, they can also be used to represent certain aspects of heterogeneity. As a consequence, the pattern reconstruction algorithm is also able to reconstruct heterogeneous patterns to some extent. First, second-order summary statistics such as $K(r)$ or $g(r)$ that capture the average number of points in a neighborhood can be used to describe large-scale clustering. When used in a reconstruction, they will produce some sort of “average” large-scale clustering, but will have problems in correctly reproducing areas void of points or with low-point density. Second, the nearest-neighbor distribution functions $D_k(r)$ can quantify certain aspects of heterogeneous patterns, which are missed by the pair-correlation function. For example, if the pattern comprises areas of low point density (or isolated points), some points will have their $k$th nearest neighbor at large distances $r$. This will be accounted for by the $D_k(r)$, which will saturate only at large distances $r$ [i.e., $D_k(r) < 1$ also for large values of $r$]. This will force the pattern reconstruction algorithm to produce low-density areas, with isolated points, and create heterogeneities, if the neighborhood rank $k$ and the maximal distance $r_{\text{max}}$ over which $D_k(r)$ is evaluated are large enough. Finally, the spherical contact distribution $H_s(r)$ is useful for characterizing an additional aspect of heterogeneous patterns not captured by the pair-correlation function or the $D_k(r)$. Because $H_s(r)$ measures the distribution of the sizes of gaps in a pattern, it is able to force the pattern reconstruction algorithm to produce areas void of points.

Several summary statistics of a different nature [such as $g(r)$, $D_k(r)$, and $H_s(r)$] can be used together to recreate a heterogeneous spatial structure, but only up to the maximal spatial scale $r_{\text{max}}$ considered in the calculation of the summary statistics. This is demonstrated in Figure 3.52, where we generated a simple heterogeneous pattern by enlarging the homogeneous pattern shown in Figure 3.50a to include a void area of the same size (Figure 3.52a). We use this to explore whether, when used together, the pair-correlation function, the spherical contact distribution, and the distribution functions to the $k$th neighbor are able to produce a reasonable reconstruction of this strongly heterogeneous pattern. Indeed, the pattern reconstruction algorithm deals very well with this pattern. Figure 3.52b shows one reconstruction. Because the intensity was not fixed, the void areas of the observed and the reconstructed pattern do not overlap, but the two patterns show the same typical structure internally, with small clusters and an otherwise hyperdispersed pattern. This demonstrates the power of the “homogeneous” summary statistics to depict heterogeneities in the intensity function.

Some reconstructions, especially patterns with large-scale gradients in the intensity, may be incomplete, as illustrated in Figure 3.53. The example pattern used for reconstruction is characterized by a strong heterogeneity
Estimators and Toolbox

in the intensity, with a curved band of high-point density in a pattern otherwise characterized by low-point density or an absence of points (Figure 3.53a). The points inside the high, point-density area exhibit a strong cluster structure, with larger clusters appearing to be composed of smaller clusters. The reconstruction of this pattern with the summary statistics $K(r)$, $g(r)$, $H_j(r)$, and $D_k(r)$ (with $k = 1, \ldots, 50$) for distances up to 200 m are in good agreement with the small to intermediate scale spatial structures in the original pattern (Figure 3.53b). Although the clusters in the reconstruction resemble those of the original pattern closely, these elements were not assembled in the same way as the original pattern (Figure 3.53b). Instead, they were distributed differently within the plot, leaving approximately the same area void of points, but did not form the characteristic band shown by the original pattern. The reconstruction in Figure 3.53b has a distinctly heterogeneous appearance, which again confirms that summary statistics developed for homogeneous patterns can generate heterogeneous patterns. However, the different small- to medium-scale elements are structured in a somewhat different fashion than those of the original pattern. Clearly, the pattern reconstruction algorithm is not able to recreate larger-scale

FIGURE 3.52
Pattern reconstruction of a heterogeneous pattern. (a) The observed pattern is the same as in Figure 3.50a, but enlarged by a void area from x-coordinates 500 to 1000. (b) Pattern reconstructed using the pair-correlation function, the spherical contact distribution, and the distribution functions to the kth neighbor $D_k(r)$ for $k = 1, \ldots, 50$ for distances up to 200 m.
Figure 3.53
Pattern reconstruction of a heterogeneous pattern. (a) Original pattern that contains a high-intensity band extending from the bottom center of the plot to the top right. (b) Reconstruction using the summary statistics $g(r)$, $K(r)$, $H_s(r)$, and $D_k(r)$ ($k = 1–50$) taken over the distance interval of 0–200 m. (c) Reconstruction as panel (b), but the intensity function shown in panel (d) was also used. (d) Estimate of the intensity function after de-clustering with bandwidth 50 m.
gradients such as the high-density band in the observed pattern (Figure 3.53a). To capture this component of the pattern, we must incorporate the intensity function in the reconstruction.

### 3.4.3.5 Reconstructions Using the Intensity Function

Pattern reconstruction can be conditioned on the location-based intensity function $\lambda(x)$ of the observed pattern, if it is critical to conserve the large-scale structure (e.g., smooth gradients in intensity such as those in Figure 3.53a). However, the estimate of the intensity function should not include spatial structure at small scales, so that the reconstruction can capture smaller-scale features that are independent of the intensity. Intensity functions with this property can be estimated, for example, by using a nonparametric kernel estimate that averages over the small-scale structures up to a bandwidth of $h$ (Box 2.4). However, for patterns with strong small-scale clustering, this approach is somewhat problematic because the strong clustering is a small-scale, autocorrelated phenomenon that can bias the intensity estimate. For example, small clusters may be due to some inherent clustering mechanisms and not caused by a locally elevated intensity function (e.g., better environmental conditions). Therefore, strong small-scale clustering should be removed before the intensity function can be estimated nonparametrically from the observed pattern (see Section 3.3.1).

Once the intensity function is estimated (Figure 3.53d), the reconstruction algorithm starts with an initial pattern based on a heterogeneous Poisson process in which a random point is accepted with a probability proportional to $\lambda(x)$. Similarly, any further tentative point is selected in accordance with the intensity function $\hat{\lambda}(x)$. Conditioning on $\lambda(x)$ forces the reconstructed pattern to have the same large-scale properties as the original pattern, but the spatial structure at smaller scales is entirely driven by the partial energies of the particular summary statistics selected for pattern reconstruction. This is illustrated by the pattern shown in Figure 3.53c, which now reproduces both the small-scale characteristics of the pattern [represented by the summary statistics $g(r)$, $K(r)$, $H_s(r)$, and $D_k(r)$] as well as the large-scale gradient structure of the pattern [represented by the intensity function $\lambda(x)$]. The large-scale properties of the reconstructed pattern approximate those of the original pattern well, but the typical small-scale structures appear at stochastically displaced locations (cf. Figure 3.53a and c).

### 3.4.3.6 Ranking of Summary Statistics

Statisticians often recommend the use of multiple summary statistics in analyzing point patterns (e.g., Iliian et al. 2008, p. 214), since each statistic characterizes a somewhat different aspect of the pattern. However, as noted above, it is a widespread practice in ecological applications to use just one, or occasionally two, summary statistics. Very little is known about the loss
of information that ensues from this practice. As a consequence, Wiegand et al. (2013a) conducted a systematic study to determine which summary statistics(s) are required to accurately characterize the spatial structure of a point pattern. This involved the investigation of three fundamental issues. First, is the utilization of different summary statistics to some extent redundant? While the natural recommendation would be to use combinations of nonredundant summary statistics, it is not clear a priori if the same ranking (and specific combination) of summary statistics would apply for all patterns. A second issue was to determine to what extent such a ranking would be influenced by pattern idiosyncrasies. Finally, real-world patterns often show some aspects of nonstationarity and it would be useful to determine if one, or several, stationary summary statistics taken together can be used to describe key properties of a nonstationary pattern. For example, the density of points may be influenced by environmental covariates or the properties of local point configurations may depend on location.

Wiegand et al. (2013a) used an inverse approach to pattern reconstruction, as originally proposed in materials science, to explore the three issues described above (Rintoul and Torquato 1997; Torquato 1998). Their approach involved the reconstruction of a point pattern using only the limited information provided by a given combination of summary statistics. The degree to which the reconstructed pattern matched the original pattern was then assessed by examining how well the reconstructed pattern reproduced the summary statistics of the observed pattern for statistics not used in the reconstruction (Torquato 1998). They applied this approach to eight simulated point patterns that were homogeneous and to eight, complex, point patterns given by the locations of tree species in the BCI plot (Box 2.6). The patterns obtained from the tree species could exhibit some heterogeneity. These patterns were selected to span a wide range of possible spatial structures.

The summary statistics used in this test were \( g(r) \), \( K(r) \), \( D(r) \), \( D_s(r) \), \( H_s(r) \), and \( E(r) \). In addition, the intensity function \( \lambda(x) \) was used to constrain the reconstruction of the (possibly heterogeneous) tree point patterns. The summary statistic \( E(r) \) is related to the nearest-neighbor distribution function \( D(r) \) and yields the probability that a ring of radius \( r \) and width \( dr \) around a point of the pattern contains no point. Wiegand et al. (2013a) used 32 combinations of the six summary statistics \( g(r), K(r), D(r), D_s(r), H_s(r), \) and \( E(r) \) in exploring the ability to reconstruct the 16 spatial patterns. For each combination of summary statistics \( c \), they calculated an index \( M_c(i) \) that measured the information contained in this combination with respect to summary statistic \( i \)

\[
M_c(i) = \frac{1}{8} \sum_{\varphi=1}^{8} \frac{1}{10} \sum_{\text{rep}=1}^{10} I[E^o_{\varphi}(\psi_{\text{rep}}) < E^\text{match}_i]
\]  

(3.140)

The indicator function \( I(\cdot) \) yields a value of 1 if the condition is true and zero otherwise. The “partial energy” \( E^o_{\varphi}(\psi_{\text{rep}}) \) quantifies the deviation of the
observed pattern $\varphi$ from one reconstruction $\psi^{\text{rep}}$ produced by the summary statistic $i$ (Equation 3.137). If the partial energy was smaller than a predefined threshold $E^{\text{match}}_i$, the observed $\varphi$ and simulated pattern $\psi^{\text{rep}}$ agreed in the summary statistic $i$. The threshold $E^{\text{match}}_i$ was the largest partial energy $E^i(\psi)$ observed in reconstructions $\psi$ where the summary statistic $i$ was used. If summary statistic $i$ was used for reconstruction, it always matched the threshold. When a summary statistic was not used for reconstruction, it matched if the deviation yielded $E^i(\psi) < E^{\text{max}}_i$. The index $M_i(i)$ represents the proportion of 80 reconstructions (of eight different patterns $\varphi$ and their 10 replicate reconstructions), based on a combination of $c$ summary statistics, that yielded a match for summary statistic $i$. Finally, the six “partial” matches $M_c(i)$ of the different summary statistics $i$ were averaged to obtain the final index $M_c = \sum M_c(i)/6$ that measures the information contained in a given combination $c$ of summary statistics.

Wiegand et al. (2013a) found a surprising amount of consistency between the 32 indices $M_i$ estimated for the eight simulated patterns and the eight tree patterns (Figure 3.54). The combination of summary statistics that did well for the homogeneous patterns also did well for the partly heterogeneous (real-world) patterns. The different summary statistics were also redundant, to a certain extent (Figure 3.54). The index $M_c$ should have a value of $sc/6$, with $sc$ being the number of summary statistics used for reconstruction.

FIGURE 3.54
Redundancy and ranking of the different combinations of summary statistics. The figure shows the index $M_i$ that measures the information contained in a given combination of summary statistics for cases of 1, 2, 3, 4, and 5 summary statistics. Pattern redundancy occurs if the value of $M_i$ is smaller than $sc/6$ (indicated as horizontal lines), with $sc$ being the number of summary statistics of the combination. Filled bars: simulated (homogeneous) patterns, open bars: observed patterns of trees at BCI (possibly heterogeneous). The best combination for a given number of summary statistics is indicated by a star. (Modified after Wiegand, T, F. He, and S.P. Hubbell. 2013a. Ecography 36: 92–103.)
The pair-correlation function carried the most information on spatial structure, if it was used alone for pattern reconstruction; in this case, approximately 70% of the six summary statistics [i.e., $g(r)$, $D(r)$, $D_k(r)$, $H_s(r)$, and $E(r)$] were matched on average (Figure 3.54). As expected, the two nearest-neighbor statistics $D(r)$ and $H_s(r)$, when considered alone, contained very little information, and $D_k(r)$ carried notably more information than $D(r)$. Interestingly, for the homogeneous patterns the 50 nearest-neighbor distribution functions $D_k(r)$, when considered together, yielded a match almost as good as that of the pair-correlation function $g(r)$ ($\Delta M_c = 0.021$). However, for the point patterns of trees at BCI, the performance of $g(r)$ was clearly better ($\Delta M_c = 0.14$). The summary statistic that, together with the pair-correlation function, contained the most information was $D_k(r)$, with $M_c = 0.89$ for both pattern types. $g(r)$ together with $H_s(r)$ yielded only slightly poorer matches ($M_c = 0.87$). The combination of three summary statistics that captured the most information was $g(r)$, $D_k(r)$, and $H_s(r)$, yielding $M_c = 1$ for the homogeneous patterns and $M_c = 0.96$ for the recruits pattern. Finally, the combination of four summary statistics $g(r)$, $D_k(r)$, $H_s(r)$, and $E(r)$ yielded a perfect match with $M_c = 1$ for both pattern types. Thus, we found a consistent ranking for the summary statistics, with respect to average pattern match, of first $g(r)$, second $D_k(r)$, third $H_s(r)$, and fourth $E(r)$.

Wiegand et al. (2013a) repeated the analysis of Figure 3.54 to explore if the above ranking was also robust with respect to pattern idiosyncrasies. However, instead of looking at the average match among the eight patterns, they conducted the analysis separately for each pattern $\phi$. A given combination of summary statistics $c$ was defined as appropriate, if the “partial” index $M_c(\phi)$ shown in Equation 3.140 was greater than 0.97. This result means that a combination $c$ contained more than 97% of the information contained in all six summary statistics together. They found that the minimal number of summary statistics required for a satisfying match varied between one and four. In some cases, only one summary statistic [$g(r)$ or $K(r)$] was needed to reconstruct a pattern satisfactorily. These were patterns that contained little spatial structure. However, more complex patterns required four summary statistics to yield satisfying reconstructions. Thus, pattern idiosyncrasy plays an important role in the selection of summary statistics. Some patterns contain little spatial structure, which can be quantified with one or two summary statistics, whereas other patterns contain more structure that can only be captured by employing a larger number of summary statistics. Even so, the sequence of useful summary statistics was, in most cases, defined by the ranking $g(r)$ first, $D_k(r)$ second, $H_s(r)$ third, and $E(r)$ fourth.

Including the intensity function in pattern reconstruction improved the overall match $M_c$ for all combinations $c$ of summary statistics (Figure 3.55). The improvement was greatest for combinations of summary statistics that captured relatively little of the spatial structure [such as $D(r)$ or $H_s(r)$], but low for the best combinations of summary statistics [such as $g(r)$, $g(r) + D_k(r)$,
Estimators and Toolbox

\[ g(r) + D_k(r) + H_s(r) \] that already captured a large degree of spatial structure (Figure 3.55). This result indicates that the summary statistics \( g(r), K(r), D(r), D_k(r), H_s(r), \) and \( E(r) \) capture properties of the pattern, which are largely independent of the intensity function. This is understandable because these summary statistics describe mostly small- to medium-scale structures, whereas the intensity function captures the large-scale structures.

One summary statistic that is able to capture larger-scale properties of a pattern is the mean distance \( nn(k) \) to the \( k \)th neighbor. Wiegand et al. (2013a), therefore, analyzed the match in \( nn(k) \) for three different cases: (i) reconstruction of eight real patterns from the BCI plot (represented by the recruits of eight different species) without use of the intensity function; (ii) reconstruction of the same eight patterns using the intensity function; and (iii) reconstruction of eight simulated homogeneous pattern, without the intensity function. The results were interesting. The eight real patterns, when reconstructed without the intensity function, generally yielded a poor fit for the mean distance \( nn(k) \) to the \( k \)th neighbor (Figure 3.56a). For most combinations of summary statistics, only two of the eight patterns showed a good match. Those were patterns with little spatial structure. For most of the other patterns, the combinations of the six summary statistics \( g(r), K(r), D(r), D_k(r), H_s(r), \) and \( E(r) \) were not able to yield a satisfying fit in \( nn(k) \). Thus, these summary statistics were not able to capture the larger-scale properties of the pattern. However, when the intensity function was included in the reconstruction, there were combinations of summary statistics that yielded satisfying reconstruction of \( nn(k) \) for all eight real patterns (Figure 3.56b). Thus, the large-scale and small- to intermediate-scale properties of the patterns were largely separated. Interestingly, in most cases reconstructions of the eight simulated homogeneous patterns yielded

**FIGURE 3.55**
A comparison of the information captured by pattern reconstruction between the cases where the intensity function is not included and where it is included. We estimated the \( M_i \) indices that measure the information contained in a given combination \( c \) of summary statistics for the tree recruits from BCI using 32 combinations of summary statistics. The figure shows the resulting \( M_i \) indices using (x-axis) and not using (y-axis) the intensity function \( \lambda(x) \). (Modified from Wiegand, T., F. He, and S.P. Hubbell. 2013a. *Ecography* 36: 92–103.)
a good fit to \( mn(k) \), even if no intensity function was used for reconstruction (Figure 3.56c). This is understandable because these patterns do not contain large-scale structure and therefore a suitable combination of the summary statistics \( g(r) \), \( K(r) \), \( D(r) \), \( D_k(r) \), \( H_s(r) \), and \( E(r) \) will yield reconstructions that also match \( mn(k) \).

In summary, the study by Wiegand et al. (2013a) found clear answers to the three issues explored. First, a certain degree of redundancy exists among summary statistics and the redundancy depends on the complexity of the pattern analyzed. Patterns with near random structures can be described well by only one summary statistic, but four or five summary statistics are needed to capture the properties of patterns with more complex spatial structures.

Second, Wiegand et al. (2013a) found a clear and robust ranking of summary statistics in the order of utility of first \( g(r) \), second \( D_k(r) \), third \( H_s(r) \), and fourth \( E(r) \). These four summary statistics were able to capture the spatial structure of simulated stationary patterns and were also able to recover small- to medium-scale heterogeneity in real-world patterns. However, larger-scale properties of nonstationary patterns could only be recovered when using the intensity function \( \lambda(x) \) for pattern reconstruction. The ranking of summary statistics was remarkably robust with respect to pattern idiosyncrasies, in the sense that patterns with little spatial structure could be quantified with the first one or two summary statistics in the ranking, whereas patterns with increasingly complex spatial structures required addition of further summary statistics in the sequence defined by the ranking. Finally, Wiegand et al. (2013a) found that stationary summary statistics can indeed describe key aspects of heterogeneous patterns. The spherical contact distribution and nearest-neighbor distribution functions play important roles in describing gaps and in
Estimators and Toolbox

capturing heterogeneous elements, such as presence of isolated points, or areas of low point density. However, the intensity function is needed to characterize larger-scale structures of heterogeneous patterns. In general, spatial analysis will require the use of several summary statistics at the same time to avoid flawed inferences that leave important properties of a pattern undetected.

### 3.4.3.7 Pattern Reconstruction as a Null Model

We now provide two examples for the analysis of bivariate patterns, where we test for independence using pattern reconstruction. In testing independence, the interest is in the relationship between the two patterns and not in the spatial structure of the composite pattern. Therefore, a null model used in analyzing the pattern needs to condition on the univariate spatial structures of the two component patterns, but must also break apart the possible spatial dependence between the two. Pattern reconstruction can be used to generate this null model. For this purpose, we keep the first component pattern unchanged but randomize the second pattern based on pattern reconstruction. The reconstructions are stochastic realizations of the second component pattern that show all typical structures of the original pattern, but at somewhat displaced locations. The advantage of this method is that it conserves the detailed spatial structure of the original pattern without producing potential artifacts of the torus shift null model. Additionally, pattern reconstruction is more flexible in capturing potentially complex spatial structures than parametric reconstruction methods that use models, such as the Thomas process.

The first example investigates independence of two patterns with complex univariate structures. The patterns represent recruits of two tropical tree species *Cecropia insignis* and *Miconia argentea* as found at the BCI plot (Figure 3.57) (Box 2.6). Recruits are all individuals that exceeded the size threshold of 1 cm diameter at breast height during the preceding five years and therefore entered the census for the first time. They were usually saplings up to 3 m in height. The two species used in this example are typical light-demanding gap species, which exhibit a complex superposition of a random pattern (of isolated recruits) with a cluster pattern, where very strong small-scale clustering was nested within larger clusters. These univariate spatial patterns were analyzed in detail in Wiegand et al. (2009).

The analysis with the pair-correlation function and the null model based on pattern reconstruction of *M. argentea* shows that the recruits of the two species are strongly attracted at scales up to 15 m (Figure 3.57b). Thus, the neighborhood density of *M. argentea* recruits around *C. insignis* recruits is significantly greater than expected under independence. For neighborhoods of 2 m, the neighborhood density is more than 20 times higher than expected under independence (Figure 3.57b). The bivariate distribution function of nearest-neighbor distances $D_{12}(r)$ shows a significant attraction affect up to 40 m. The nearest *M. argentea* recruits were up to 40 m closer to *C. insignis*
recruits than expected by independence. The independence null model only agreed with the observed pattern at distances greater than 60 m (Figure 3.57c). The strong attraction was probably caused by canopy gaps shared by the recruits of the two light-demanding species.

Comparison of the results of independence with those of the CSR null model (cf. Figure 3.57d and b) shows that the CSR null model underestimates
the variability in point configurations and yields simulation envelopes in
the pair-correlation function that are too narrow. This is understandable
because CSR cannot produce the strong clustering observed, since the CSR
null model destroys the clusters. In contrast, pattern reconstruction with the
independence null model may produce many configurations where clusters
of \textit{M. argentea} accidentally overlap those of \textit{C. insignis}. Similarly, configurations
where the clusters of the two species accidently segregate are unlikely under
CSR, but are likely under independence. As a consequence, the independence
null model yields wider simulation envelopes. While CSR and independence
yield the same expectations under the pair-correlation function, the expecta-
tions of the nearest-neighbor distribution function \( D_{12}(r) \) differ for the two
null models (cf. Figure 3.57c and e). Under CSR the nearest \textit{M. argentea} neigh-
bors of \textit{C. insignis} trees is usually much closer than under the independence
null model. The reason for this is that the CSR null model does not maintain
the observed clustering (which concentrates points in space) and therefore
fills the space more thoroughly, thereby placing the \textit{M. argentea} individuals
closer to those of \textit{C. insignis}.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{pattern_reconstruction}
\caption{Pattern reconstruction using the intensity function. (a) Map of large individuals (dbh > 10 cm)
of the species \textit{Shorea trapezifolia} from the Sinharaja forest in Sri Lanka. (b) Intensity function of
\textit{S. trapezifolia} estimated with a nonparametric kernel estimate using a bandwidth of \( R = 50 \) m,
after slight de-clustering using parameter \( r_c = 3 \) m for the minimal distance between cluster
centers. (c) and (d) Reconstructions using the intensity function in panel (b).}
\end{figure}
The second example illustrates how pattern reconstruction can help to assess independence even in the case of heterogeneous patterns. The data are large individuals (i.e., dbh > 10 cm) of the two species *M. tetrandra* (focal species) and *Shorea trapezifolia* in the Sinharaja forest of Sri Lanka (Figures 3.58a, 3.59a). The plot contains strong topographic structuring and both species occur predominately in the western half of the plot. Thus, the

![Pattern reconstruction using the independence null model under heterogeneity.](image)

**FIGURE 3.59**

Pattern reconstruction using the independence null model under heterogeneity. (a) Map of patterns of large individuals (dbh > 10 cm) of the species *Mastixia tetrandra* (focal species, open disks) and *Shorea trapezifolia* (reconstructed pattern, gray closed disks) from the Sinharaja forest. (b) Intensity function of *S. trapezifolia* estimated with a nonparametric kernel estimate using a bandwidth of $R = 50$ m after slight de-clustering using parameter $r_c = 3$ m for the minimal distance between cluster centers. (c) Bivariate pair-correlation function and the independence null model for pattern reconstruction of *S. trapezifolia* using the intensity function shown in panel (b). (d) Same as panel (c), but for the nearest-neighbor distribution function $D_{12}(r)$. (e) Same as panel (c), but incorporating the corresponding heterogeneous Poisson null model based on the intensity function shown in panel (b). (f) Same as panel (e), but for the nearest-neighbor distribution function $D_{12}(r)$. © 2014 Taylor & Francis Group, LLC
pair-correlation function is clearly larger than the expectation under independence (i.e., a value of 1; Figure 3.59a) and testing for independence, as done in the previous example, would most likely depict the effects of habitat association of the two species (i.e., they both co-occur largely within the same area of the plot), but provides no answer to the question of whether there is potential attraction or repulsion of the individuals of the two species within a local neighborhood. To do this we need to condition on the observed large-scale structure of the second pattern (i.e., that of S. trapezifolia). This can be done by using a heterogeneous Poisson process based on a nonparametric intensity estimate of the second pattern (Figure 3.58b), which then randomizes the individuals of this pattern only locally. Here we selected a bandwidth of $R = 50$ m, which is somewhat larger than the interaction range among trees. However, pattern reconstruction using the intensity function allows us to additionally condition on the small-scale structure of the second pattern. In this case, the null model patterns have the same intensity function and show the same small-scale structures as the observed pattern, but they are slightly displaced relative to the observed pattern (Figure 3.58c,d).

Figure 3.59c shows the results of the analysis for the pair-correlation function. At distances of 2–9 m we observe a clear attraction of the two species. At larger distances, the observed pair-correlation function agrees with that expected under the independence null model implemented with pattern reconstruction. Thus, both species co-occur in the same restricted area of the plot (i.e., the heterogeneity) and show additional attraction. This is also evident from visualizing the patterns (Figure 3.59a); there are small, shared clusters, where both species co-occur. The analysis using only the heterogeneous Poisson process (which does not maintain the observed small-scale structure of the pattern of S. trapezifolia) shows basically the same result, but produces a smaller level of attraction in the null model at small distances (Figure 3.59e). However, the nearest-neighbor distribution functions show fewer differences between the two null models. However, the pattern reconstruction null model suggests stronger attraction at distances of 2–9 m. Thus, the nearest neighbors under the heterogeneous Poisson null model (which does not maintain the observed clustering) are in general somewhat closer than under the corresponding pattern reconstruction null model that conserves the observed clustering.