Dealing with virtual aggregation – a new index for analysing heterogeneous point patterns

Katja Schiffers, Frank M. Schurr, Katja Tielbörger, Carsten Urbach, Kirk Moloney and Florian Jeltsch

K. Schiffers (katja.schiffers@uni-potsdam.de), Dept of Plant Ecology and Nature Conservation, Univ. of Potsdam, Maulbeerallee 2, DE-14469 Potsdam, Germany, and Dept of Plant Ecology, Univ. of Tübingen, Auf der Morgenstelle 3, DE-72076 Tübingen, Germany. – F. M. Schurr and F. Jeltsch, Dept of Plant Ecology and Nature Conservation, Univ. of Potsdam, Maulbeerallee 2, DE-14469 Potsdam, Germany. – K. Tielbörger, Dept of Plant Ecology, Univ, of Tübingen, Auf der Morgenstelle 3, DE-72076 Tübingen, Germany. – C. Urbach, Inst. of Physics, Humboldt-Univ. Berlin, Newtonstr. 15, DE-12489 Berlin, Germany. – K. Moloney, Dept of Ecology, Evolution and Organismal Biology, Iowa State Univ., 253 Bessey Hall, Ames, IA 50011–1020, USA.

Point pattern analyses such as the estimation of Ripley's K-function or the pair-correlation function g are commonly used in ecology to characterise ecological patterns in space. However, a major disadvantage of these methods is their missing ability to deal with spatial heterogeneity. A heterogeneous intensity of points causes a systematic bias in estimates of the K- and g-functions, a phenomenon termed "virtual aggregation" in the recent literature.

To address this problem, we derive a new index, called K2-index, as an extension of existing point pattern characteristics. The K2-index has a heuristic interpretation as an approximation to the first derivative of the g-function.

We estimate the K-, g- and K2-functions for six different types of simulated point patterns and show that the K2-index may provide important information on point patterns that the other methods fail to detect. The results indicate that particularly the small-scale distributions of points are better represented by the K2-index. This might be important for testing hypotheses on ecological processes, because most of these processes, such as direct neighbour interactions, occur very locally.

When applied to empirical patterns of molehill distribution, the results of the K2-analysis show regularity up to distances between 0.1 and 0.4 m in most of the study areas, and aggregation of molehills up to distances between 0.2 and 1.1 m. The type and scale of these deviations from randomness agree with a priori expectations on the hill-building behaviour of moles. In contrast, the estimated g-functions merely indicate aggregation at the full range from 0 to 7 m (or even above).

Considering the advantages and disadvantages of the different methods, we suggest that the K2-index should be used as a complement to existing approaches, particularly for point patterns generated by processes that act on more than one scale.

The analysis of spatial patterns has become a widely used approach in ecological research (Levin 1992, Grimm et al. 1996, Wiegand et al. 1998, Barot et al. 1999, Jeltsch et al. 1999, Stoyan and Penttinen 2000, Bossdorf et al. 2000, Klaas et al. 2000, Wiegand et al. 2000, Liebhold and Gurevitch 2002, Revilla and Palomares 2002, Wiegand and Moloney 2004). Spatial patterns in natural systems, e.g. the shape of species ranges, the location of territories, or the local distribution of nests, dens, dispersed seeds and disturbances reflect the underlying ecological processes that cause their distribution in space. As the same pattern can be a realisation of different types of processes, it is generally impossible to infer the underlying process from a pattern (Levin 1992, Stoyan and Penttinen 2000, Molofsky et al. 2002). However, in combination with explicit a priori hypotheses, the analysis of emerging patterns can help to

identify the likely generating processes or provide insight into the strength and scale-dependence of known processes (Levin 1992, Schurr et al. 2004, Perry et al. 2006, van Teeffelen and Ovaskainen 2007).

Point pattern analysis is commonly used to investigate patterns of sessile organisms like plants or spatial distributions of other stationary structures such as bird nests or gopher- and molehills (Klaas et al. 2000). Generally speaking, a point pattern is a set of locations of individuals or other ecological entities that are recorded in a defined study region. The density of these points is referred to as the "intensity" of the pattern (Ripley 1981). Since most ecological processes are highly scale dependent (Wiens et al. 1993, Gustafson 1998), approaches are needed that can detect (auto-)correlation of points over a range of spatial scales. Ripley's K-function and the pair-correlation

function g (Ripley 1981) fulfil this requirement and methods based on them have proliferated in recent years (Haase 1995, 2001, Podani and Czaran 1997, Goreaud and Pélissier 1999, Coomes et al. 1999, Dale and Powell 2001, Freeman and Ford 2002, Dungan et al. 2002, Perry et al. 2002, Diggle 2003, Wiegand and Moloney 2004, Wiegand et al. 2006). However, the K- and the g-function are developed for patterns with a homogeneous intensity across the study region (Ripley 1981, Pélissier and Goreaud 2001). Hence, spatial variation in intensity induces a systematic bias in the estimated functions, indicating a stronger positive autocorrelation than actually exists. This effect is termed "virtual aggregation" (Wiegand and Moloney 2004).

As a consequence of virtual aggregation, essential information of the pattern – particularly at smaller scales – may be obscured, so that the understanding of the causative ecological processes might be impeded (Spooner et al. 2004, Strand et al. 2007). A variety of approaches has been developed in an attempt to tackle this problem, either by the use of homogeneous subregions (Pélissier and Goreaud 2001) or by an adaptation of the null-hypothesis to the underlying heterogeneity (Baddeley et al. 2000, Baddeley and Turner 2005, Getzin et al. 2006). But these approaches generally suffer from the fact that pre-analyses are needed and part of the underlying information might be lost, if there is no sufficient a priori knowledge on the pattern available. However, since almost all natural patterns are to some degree heterogeneous (Wagner and Fortin 2005), an analysis that is not affected by virtual aggregation would be highly desirable.

In this work, we introduce the K2-index – a straightforward extension of the g-function – as a new approach to deal with the problem of virtual aggregation. We apply this method to simulated point patterns, discuss the advantages and disadvantages of K, g and K2, and show that the K2-index is less sensitive to heterogeneity than the K- and g-functions. We then apply the K2-index to a three-year dataset on molehill distribution to show its utility in interpreting natural patterns with respect to the underlying processes. Finally, we formulate recommendations for the use of the three addressed approaches in ecological studies focusing on various aspects of point pattern characteristics.

Methods of point pattern analysis

The K-function

Ripley's K-function is the most commonly used approach to characterise point patterns in ecological studies. It is defined as the expected (E) number of additional points within a circle of radius r centred on a randomly chosen point, divided by the overall intensity of the pattern:

$$K(r) = \frac{E \text{ [no. of points with distance} < r]}{n/A}, \qquad (1)$$

where n is the total number of points and A is the area of the study region. Since K(r) not only expresses the expected number of points at distance r but also comprises all points at smaller distances, it can be interpreted as being a cumulative distribution function.

To calculate the estimator of the underlying K-function of a pattern $\hat{K}(r)$, the mean number of events, located within radius r around each event, is evaluated and standardised by the overall intensity:

$$\hat{K}(r) = \frac{A}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}^{-1} I_t(d_{ij}), \qquad (2)$$

where $I_t(d_{ij})$ is an indicator variable that equals 1, if the distance d_{ij} between the points i and j is smaller than r, and 0 otherwise. To correct for edge effects, the counter variable is multiplied by a weighting factor w_{ij} that corresponds to the fraction of the circle centred on i lying inside the study area. The resulting value of $\hat{K}(r)$ can then be compared with the expected number of events under a null-hypothesis (such as complete spatial randomness, CSR). Observed numbers greater than the expected value indicate aggregation relative to the null-hypothesis, whereas observed numbers smaller than the expectation indicate regularity.

For an easier visual interpretation of the results, a transformation of $\hat{K}$, called L-function (Besag 1977), is typically used:

$$\hat{L}(r) = \sqrt{\hat{K}(r)/\pi} - r. \qquad (3)$$

The advantage of $\hat{L}$ is that the square root transformation stabilises the variance and CSR patterns have an expectation of zero at all scales (Ripley 1981).

The g-function and O-ring statistics

In contrast to the K-function (eq. 1), the g-function gives the expected number of points at distance r from a randomly chosen point, divided by the mean intensity of the pattern and by the perimeter of the ring with radius r:

$$g(r) = \frac{E \text{ [no. of points at distance r]}}{n/A * 2\pi r}. \qquad (4)$$

Therefore, the value of the g-function for scale r can be considered to be the finite difference of the number of points between the circle with radius r+d and the circle with radius r (with d being small), normalised by the perimeter of the circle. In the limit of d→0 this is equivalent to the derivative of K, so that the g-function can be written as

$$g(r) = \frac{dK(r)/dr}{2\pi r}, \qquad (5)$$

Unlike the cumulative K-function, the g-function can be interpreted as being a probability density function of r (Ripley 1981, Mattfeldt et al. 2006). An estimate of the g-function is given by

$$\hat{g}(r) = \frac{1}{2\pi r} \frac{A^2}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} k_h(d_{ij}). \qquad (6)$$

k_h is an Epaneçnikov smoothing kernel, putting maximal weight on points with distances d_{ij} that are exactly equal to r, but incorporating points with reduced weight, as well, whose distances do not differ more from r than defined by the bandwidth parameter of the kernel. As recommended

by Stoyan and Stoyan (1994) a bandwidth equal to $h/\sqrt{5}$, where $h = 0.15/\sqrt{A/n}$ was chosen.

The O-ring statistic (Wiegand et al. 1999) is basically the discrete variant of the g-function, using an underlying grid to estimate the respective function, but giving the absolute number of expected additional points in circle with radius r instead of the normalised value.

The K2-index

Just as the g-function shows the expected differences in the number of points between consecutive circles, the K2-index shows the expected differences in the number of points between consecutive distances. Therefore, analogous to the mathematical relation between the K- and the g-function (eq. 1, 4), the K2-index can be heuristically interpreted as the first derivative of the g-function:

$$K2(r) = \frac{dg(r)}{dr} \quad (7)$$

Other than the g-function, the K2-function does not indicate a range of distances but rather the upper limit of a range of distances at which events are not randomly distributed. An estimate of the K2-index for a given pattern can be obtained by calculating the slope of the estimated g-function over a small range of scales between $r+\Delta r$ and $r-\Delta r$.

$$\hat{K}2(r) = \frac{\hat{g}(r + \Delta r) - \hat{g}(r - \Delta r)}{2\Delta r}. \quad (8)$$

Scales where $\hat{K}2$ is different from zero can be interpreted as distances of changing intensities. Note that a positive value indicates a transition from lower to higher intensity and hence – other than a positive value in an L- or g-function, which would indicate attraction of events – the upper limit of a range of scales at which the pattern is regular. A negative value indicates a transition from higher to lower intensity and hence the upper limit of a range of scales at which the pattern is aggregated. This opposite algebraic sign of the K2-index compared to the two other indices reflects its mathematical dependence on the g-function and emphasises the fact that K2-analyses have a somewhat different interpretation.

Calculation of confidence envelopes

We used Monte-Carlo simulations to show at which scales the observed patterns depart significantly from a null-hypothesis of complete spatial randomness (CSR). For each of 399 realisations of a homogeneous CSR-process, the K-, g-, and K2-index were estimated. At each scale r, the 10th highest and 10th lowest value of each statistic were then used to estimate the 95% confidence envelope for the null-hypothesis (Haase 1995, Bailey and Gatrell 1995, Martens et al. 1997).

We used the statistical software package R 2.5.1 (R Development Core Team 2007) and the additional package "spatstat" (ver. 1.11–5, Baddeley and Turner 2005) for implementing the K2-estimator, simulating the hypotheti-

cal patterns, and conducting all analyses. The R-code for the K2-estimator is available as Supplementary material.

Analysis of simulated point patterns

To compare the performance of the K-, the g-, and the new K2-index, we simulated six hypothetical patterns in an area of 20 by 20 units, for which we controlled the scale, intensity, and sign of autocorrelation: a) a realisation of complete spatial randomness with an intensity of 2 events per unit square (Fig. 1a, CSR-pattern). b) The same as (a) but all events outside a circle with a radius of 8 units were eliminated (Fig. 1b, CSR-patch). c) A realisation of a Strauss process with intensity of 2.5 per unit square. This process generates a pattern of complete spatial randomness but with a given probability eliminates all points that are within an inhibition scale of another point. In our case, the inhibition scale was 0.5, and the probability of elimination 0.7 (Fig. 1c, Strauss-pattern). d) The same as (c) but as in pattern (b) all events outside a centred circle with a radius of 8 units were eliminated (Fig. 1d, Strauss-patch). e) A realisation of two co-occurring processes: repulsion with a probability of elimination of 0.75 for an event being located within the radius of 0.6 of another event; and positive autocorrelation with a probability of elimination of 0.95 for events being located outside the radius of 0.9 (Fig. 1e, repulsion-attraction-patches). f) The same as (e) but with a probability of elimination of 0.5 for both processes (Fig. 1f, repulsion-attraction-pattern).

Other than the patterns b, d, and e, the repulsion-attraction-pattern f is not heterogeneous on the scale of the plot size. However, with the inclusion of this pattern, we want to test the performance of the different approaches in analysing a pattern where interactions occur on more than one scale.

Results for the simulated patterns

The results of the L-analysis accurately reflect the scale of the processes generating the two homogeneous patterns (Fig. 1a, c). For the CSR-pattern (Fig. 1a) there is no significant departure from the null-hypothesis at any scale. For the homogeneous Strauss-pattern (Fig. 1c), the scale of repulsion is clearly shown as the scale with the lowest resulting value of the analysis. Patterns with large-scale heterogeneity (Fig. 1b, d, e) are affected by virtual aggregation: the results of the L-analysis indicate an increasing intensity of events relative to the mean intensity of the study area. This is true for the whole range of scales for the CSR-patch (Fig. 1b). For the Strauss-patch and the repulsion-attraction-pattern (Fig. 1d, e), $\hat{L}$ shows no deviation from CSR up to the scale where repulsion occurs (0.5 and 0.6, respectively). Here, the effect of negative autocorrelation outweighs the effect of virtual aggregation. At larger scales, where no repulsion occurs, virtual aggregation dominates and leads to a significant departure from the null-hypothesis. For the repulsion-attraction-pattern (Fig. 1f), which is not heterogeneous on the scale of the plot size but only on smaller scales, the L-analysis indicates higher than average intensities up to a scale of approximately 8.5.
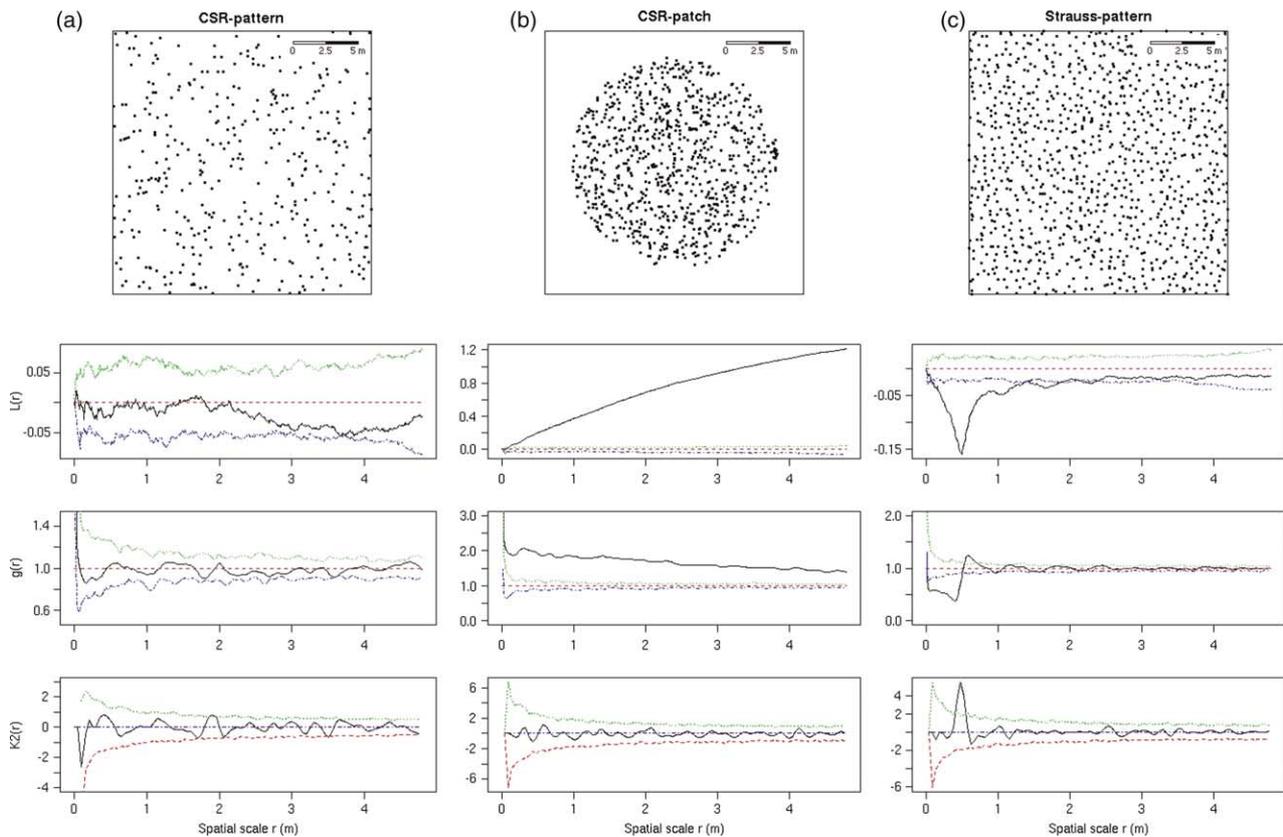
Figure 1. (a–c) Realisations of the following point processes: (a) complete spatial randomness, CSR; (b) the same as (a) but only within a patch with a radius of 8 units; (c) a Strauss process with an inhibition scale of 0.5 and an inhibition strength of 0.7. Dotted lines indicate 95% confidence intervals.

The actual process of attraction only acts up to a scale of 0.9 and the underlying repulsion up to scale 0.6 is not visible at all.

The estimated g-function provides the expected results for the two homogeneous patterns as well: no deviation from the null-hypothesis for the CSR-pattern (Fig. 1a) and a significantly decreased number of events up to scale 0.5 for the Strauss-pattern (Fig. 1c). The curves for the respective heterogeneous patterns (Fig. 1b, d) look almost similar except for two attributes: first, they are consistently shifted towards higher values of $\hat{g}$ and second, they slightly decrease towards higher scales. Again, the results for the repulsion-attraction-patches are similar to the results for the Strauss-patch (Fig. 1e). Even though the repulsion-attraction-pattern (Fig. 1f) is homogeneous on the scale of the plot size, the g-function shows more than average neighbours at distances of 0.3 until 1.3, while the actual scale of positive autocorrelation is only up to a distance of 0.9. As for the L-analysis, the repulsion at small scales is not reflected in a significant deviation from the null-hypothesis.

In general, the K2-index shows a stronger variability in the results compared to the L- and g-function, which is due to the smaller number of points that enter the calculation. Apart from this, the K2-index reflects all underlying processes very accurately so that the interpretation of the results is straightforward. For both CSR-patterns, there is no deviation from the null-hypothesis at any scale (Fig. 1a, b). For the two Strauss-patterns, the results of the K2-analysis indicate an increase in intensity around 0.5 (Fig. 1c,

d), which can be directly attributed to the underlying repulsion in the point process up to this scale. At the observed scales, the existence of the cluster in the two patch-patterns (Fig. 1b, d) has no effect on the results. The advantage of the K2-index is most apparent for the two realisations of the repulsion-attraction process. Where the other approaches overestimate the scale of aggregation and fail to detect significant evidence of repulsion, the K2-index shows a significantly positive value at a scale of 0.6 and a significantly negative value at approximately 0.9 (and 1.7 for the second repulsion-attraction pattern). These are exactly the upper limits of the scales over which the underlying processes of repulsion and attraction occur.

Evaluation of the three approaches

In the following, we discuss the explanatory power of the individual approaches and substantiate our conclusions by comparing the results for the simulated patterns, for which the underlying processes are known.

Ripley's K and the L-function
Ripley's K and the L-function (eq. 1, 3) are now the most commonly used indices for analysing point patterns in ecology (Wiegand and Moloney 2004). We argue, however, that in many studies, the K- and L-function are not optimal for testing ecological hypotheses. The reasons for that are twofold: first, the cumulative character of these indices
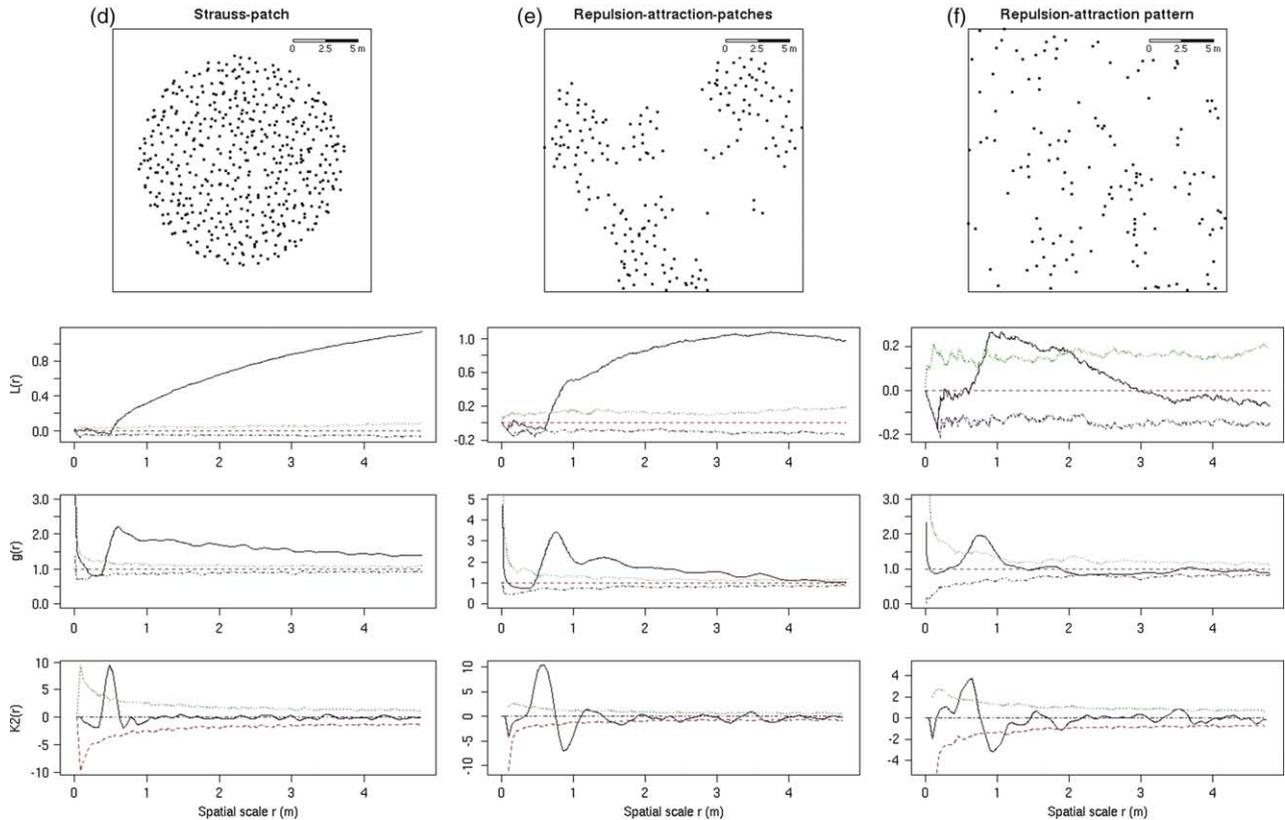
Figure 1. (d–f) Realisations of the following point processes: (d) the same as (c) but only within a patch with a radius of 8 units; (e) two co-occuring processes: repulsion up to scale 0.6 with a probability of elimination of 0.75 within the repulsion radius, and attraction up to scale 0.9 with a probability of elimination of 0.95 outside of the attraction radius; (f) same as (e) but with a probability of elimination of 0.5 for both processes. Dotted lines indicate 95% confidence intervals.

often hampers their ability to detect scale-dependent patterns (Condit et al. 2000, Schurr et al. 2004). For example, if clumping occurs on a relatively small scale (scales A, B, and C in Fig. 2) the point density at larger scales (D and E in Fig. 2) is above average, too. Accordingly, K- and L-function analyses indicate deviation from the null-hypothesis up to scales that are much larger than the radius of point clusters. Secondly, most natural patterns show at least some degree of heterogeneity (Wagner and Fortin 2005), so that the results of the K- and L-analysis are biased due to virtual aggregation. To illustrate this effect, consider the case that the cluster in Fig. 2 is caused by an exogenous factor, such as a spatially varying environmental variable (a first-order effect in the terminology of spatial statistics). In this case, it may be ecologically relevant to test, whether within the cluster the points/individuals show aggregation or repulsion (a second-order effect). Even if the individuals were distributed randomly within the patch, the estimated g-function would detect more events than expected from the global average and the results for these scales would be misinterpreted to indicate attraction. This bias in the results originates from the fact that the expected number of events is derived from the mean intensity in the entire study area rather than from the local intensity in the patch.

Virtual aggregation also influences the results of the simulated patterns: in the two patch-patterns (Fig. 1b, d), the results are significantly shifted towards higher values, so

that the actual scales of the processes are masked. Similarly, the scales of repulsion and attraction in the patterns with two co-occurring processes (Fig. 1e, f) are not accurately
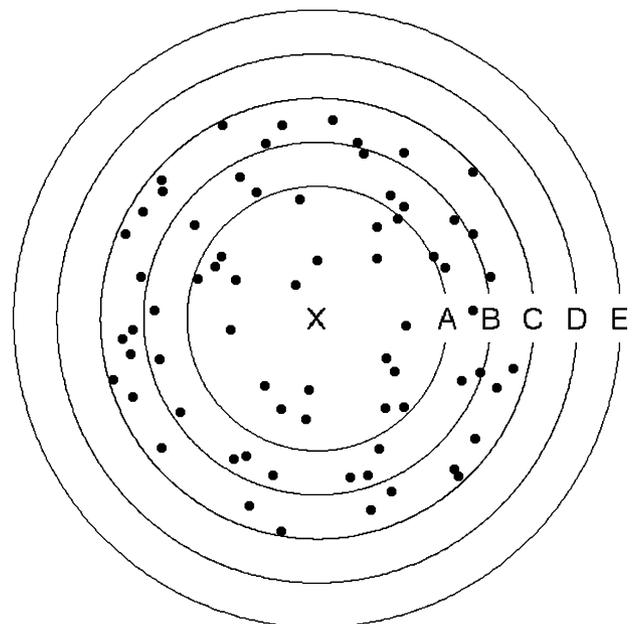

Figure 2. Example of a heterogeneous pattern with different scales of observation (A-E), exemplified for one central event (x).

549

reflected. Particularly the results of the attraction-repulsion pattern (Fig. 1f) show that virtual aggregation does not only occur in the case of large-scale heterogeneity, but as soon as the local intensity of events differs from the global intensity (e.g. because patches are formed by interactions among events or due to environmental heterogeneity). As a consequence, significant departures from the null-hypothesis at a given scale do not necessarily indicate autocorrelation. The actual information the analysis provides is the relation of the average neighbourhood intensity of events to the mean intensity of the whole study area, whereas the neighbourhood is defined by r, i.e. the scale of observation. This might be important e.g. for investigating the minimum size of an experimental area to represent the mean density of events: imagine a patchy distribution of plants. For small scales, the L-function will indicate aggregation since there are more neighbours than would be expected from the mean intensity of the area. As soon as a value of r is reached for which the L-function does not deviate from the null-hypothesis of CSR, a sample area with the radius of r centred on an individual represents the mean intensity of plants in the area. Another example would be to estimate the optimal size of the zone of influence of plant individuals, for which the number of competitors is small in relation to the covered area: the value of r for which the L-function shows the lowest value is the size of neighbourhood for which the density of neighbours is lowest.

The pair-correlation function g and O-ring statistics

The pair-correlation function g (eq. 4) (Ripley 1981, Stoyan and Stoyan 1994, Stoyan and Penttinen 2000, Dale et al. 2002) and the O-ring statistics (Wiegand and Moloney 2004), avoid the cumulative character of the K- and L-analysis by using rings instead of circles. This means that for an evaluation of the pattern in Fig. 2, not all points located within e.g. circle C are counted, but only those lying in-between circles B and C. In that way, clustering at smaller scales does not affect the number of events observed at larger scales (Getis and Franklin 1987, Penttinen et al. 1992, Condit et al. 2000). In contrast to the L-function, which shows up to which scale the pattern appears to be non-random, the intensity of events in relation to the mean intensity at a specific scale is shown. In many cases, this interpretation, similar to a probability density function, is more intuitive than the accumulative measure (Stoyan and Penttinen 2000). However, the results of the g-function are also affected by virtual aggregation in the case of heterogeneity: even if the points in the pattern in Fig. 2 would be randomly distributed within the patch, the density of points in-between scales A and B would be higher than the overall intensity so that the g-function would indicate aggregation. This is also apparent in the g-function analyses of the simulated heterogeneous patterns: neither for the patch-patterns nor for the repulsion-attraction patterns could the g-function filter the actual scales of interactions among the points (Fig. 1b, d–f).

The proper interpretation of the results of a g-analysis for a heterogeneous pattern is that it relates the intensity around the points to the mean intensity of the study area. This information is important e.g. for assessing how strong a heterogeneous distribution of individuals increases competition intensity in comparison to a totally random distribution, i.e. whether the mean-field approximation is an appropriate approach for analysing interactions among individuals (Law and Dieckmann 2000, Stoll and Weiner 2000).

One way to circumvent the problem of virtual aggregation is the use of homogeneous sub-regions of the study area for the analyses (Pélissier and Goreaud 2001, Wiegand and Moloney 2004). For patterns with obvious heterogeneity at large scales, as for the two simulated patch-patterns, this approach would lead to the desired results. But for patterns with processes on two adjacent scales – like repulsion in seedlings due to competition, but aggregation on the scale of seed dispersal distances – the method is not appropriate. First, the sub-regions have to be distinctively larger than the expected scale of interactions: due to edge effects, indices should not be calculated up to the scale of the study area. Recommendations rage from one fourth of the side length (Baddeley and Turner 2005) to one half of the side length of the study area (Pélissier and Goreaud 2007). Moreover a reasonable number of events has to be used in order to calculate the average values. Secondly – and this is particularly difficult, if there are interactions on more than one scale-all information of the pattern for scales larger than the sub-region is lost.

The K2-index

As shown above, the results of the K- as well as the g-function are affected by heterogeneity because observed intensities are related to the "global" intensity rather than the "local" intensity in the surrounding of a point. This dependence on the global intensity is circumvented in the K2-index (eq. 7) by calculating differences in intensities between scales, or – in other terms – by relating the intensities at a given scale to the intensities of the adjacent scales. Therefore, also in heterogeneous point patterns, scales of significant deviations from the null-hypothesis can be interpreted as distances where transitions from low to high intensities occur or vice versa.

An instructive example is the analysis of the realisation of the repulsion-attraction processes (Fig. 1e, f). Where the other approaches fail to detect the repulsion at small scales, the K2-index yields the correct results. A positive peak appears at a distance scale of around 0.6 and a negative one at a distance scale of 0.9, which are exactly the scales up to which repulsion and attraction occur. The additional negative value at approximately 1.7 is due to indirect effects: the presence of a neighbour that itself interacts with surrounding points increases the probability of having further neighbours at a somewhat larger distance (Stoyan and Stoyan 1994).

Since the K2-index can be estimated as the derivative of the g-function and the g-function itself as the derivative of the K-function, the number of points that enter the calculations for each value of r decreases from $\hat{L}$ over $\hat{g}$ to $\hat{K}2$. In the same order, the sensitivity of the approaches increases. On the one hand, this facilitates the detection of interactions and their correct scales, as shown in the above examples. On the other hand, it must be highlighted that for the same reason $\hat{K}2$ is also more susceptible to random noise compared to the other approaches. Some authors even

suggest to only use the K- or L-analysis for detecting significant deviations from the null-hypotheses (Stoyan and Penttinen 2000). The sensitivity of the K2-index depends on two methodological decisions, as well as on attributes of the observed pattern: 1) the bandwidth parameter h of the Epanečnikov kernel k used in the calculation of the g-function (eq. 6), which determines the degree to which results are smoothed, 2) the step width Δr used in estimating K2 (eq. 8) – the larger the steps are, the more fluctuations are ignored, and 3) the total number of events of the pattern, since the average neighbourhood densities become more stable with an increasing number of replicates. These factors do of course also affect the g-function. But since their impact on results of the K2-analysis is considerably stronger, it is even more important to keep them in mind, when interpreting K2-indices. To examine the strength of this impact, we used different values of h and Δr to analyse the pattern in Fig. 1e (Supplemetary material). The sensitivity of K̂2 decreases both with increasing step width Δr and with increasing h, but the response to changing values of h is more pronounced for small step widths. Although the resulting values obviously depend on the above factors, the general conclusions of all K2-analyses are the same: because the confidence envelopes are calculated under the same conditions as the K2-index, they change accordingly and scales of significant departures from the null-hypothesis remain identical.

Besides its high sensitivity, another disadvantage of the K2-index is that effects on large scales might not be detected if the intensity of points changes only slowly with increasing scale (i.e. if the g-function has a shallow slope). This is often the case for larger clusters like the two patch-patterns (Fig. 1b, d), because due to the averaging of all neighbourhood intensities for points with different position within a cluster, the resulting values change only gradually.

A case study: analysis of observed molehill patterns

To show the relevance of the new index for the interpretation of natural patterns, we applied the K2-index to a dataset of molehill occurrence.

Data collection and a priori hypotheses on the molehill patterns

The study region is located in the nature reserve "Untere Havelaue", in western Brandenburg, Germany (52°71′66′′N, 12°21′66′′E). The area represents an island between two arms of the river Havel and it is characterised by slightly undulating, extensively used grassland. Lower elevations are subject to annual flooding of the River Havel during the winter months and exhibit higher soil moisture throughout the whole year. In addition to flooding, moles constitute a common and widespread disturbance agent in the study region.

Within the study site, we identified 6 experimental blocks of 50 by 50 m in size. Blocks were selected based on
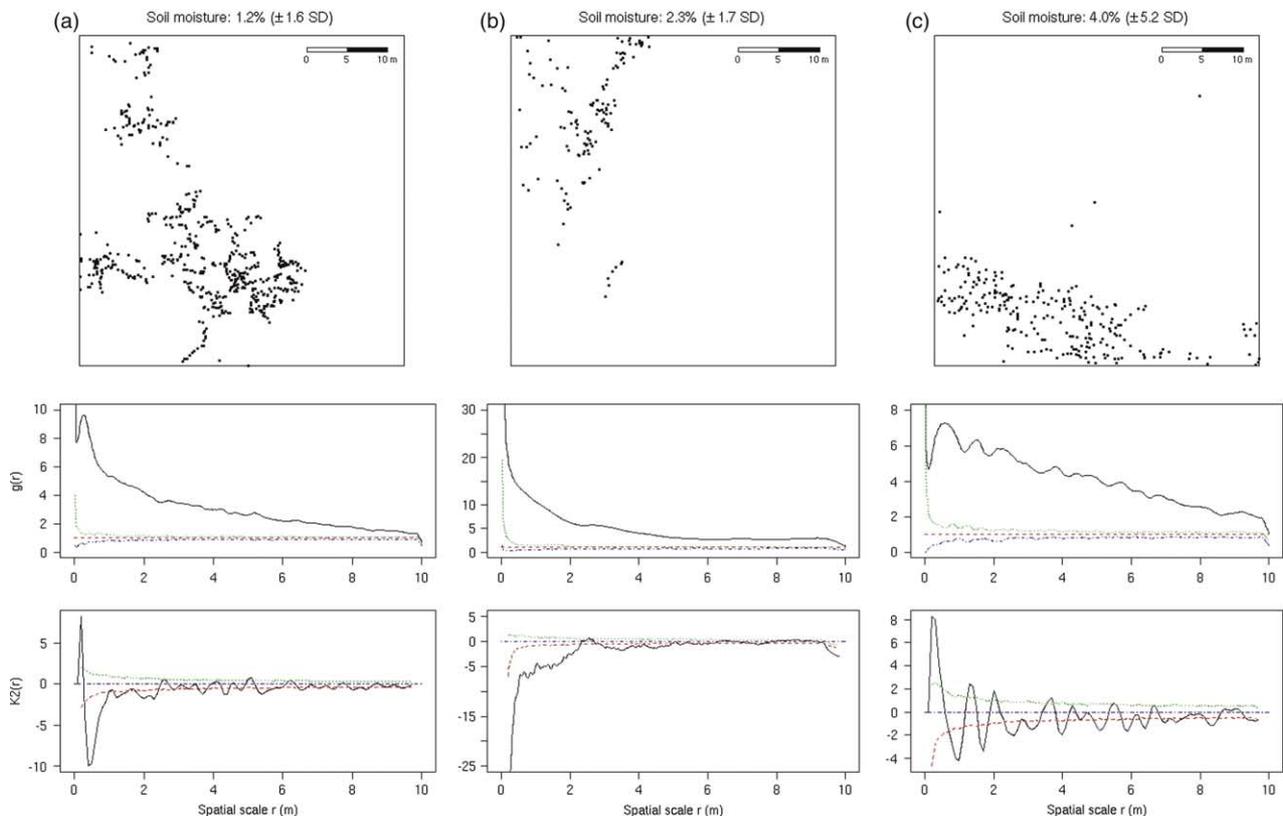

Figure 3. (a–c) Molehill patterns for three study sites with a soil moisture of 1.2% (a), 2.3% (b), and 4.0% (c). The dotted lines in the results for the L-, g-, and K2-analyses indicate 95% confidence intervals.

two criteria: 1) the presence of actively digging moles, and 2) occurrence of a range of different soil elevations allowing investigation of the influence of soil moisture, which is correlated to elevation, on the molehill pattern. In these plots, the exact positions of fresh molehills were recorded almost every week over three years using a tachymeter (Elta-R, Zeiss, Oberkochen), resulting in a total of 6738 hills. The soil moisture was measured with a mobile TDR-probe (Trime-FM, IMKO Micromodultechnik, Ettlingen) with 27 replicates for each plot at one day in June 2006.

Several endo- and exogenous factors are likely to affect the pattern of molehill production: the territoriality of the moles, their limited mobility, the necessity to dispose soil every few decimeters, the structure of the soil, and the availability of earthworms, which are their main food (Mellanby 1967). Our hypotheses were that there is: 1) negative autocorrelation at small scales, since at short intervals the costs for building new molehills are still higher than for shifting the accumulating amount of soil below-ground (Fig. 4: soil shifting), 2) positive autocorrelation among molehills at somewhat larger scales, because neighbouring molehills form part of the same tunnel system (Fig. 4: tunnel systems), and 3) a negative correlation between moisture and the scale up to which negative autocorrelation occurs, since wetter soil is heavier, so that the balance between the advantages of building a new hill and moving an increasing amount of soil shifts towards disposing soil at shorter intervals.

Since the molehill patterns are strongly heterogeneous (Fig. 3) and we expect processes on at least two scales

(hypotheses 1 and 2), we applied the K2-approach to the dataset. However, for a confirmation of the significance of the detected processes, we additionally estimated the g-functions of the six patterns. Prior to the analyses, we selected a period of continuous molehill occurrence from the three-year dataset for each plot in order to avoid effects of overlapping territories over time and to balance the number of molehills per study area. We excluded an area of markedly high molehill aggregation in one plot where a below-ground pipe was installed several years before data recording. To homogenise final plot size all plots were reduced to 40 by 40 m.

Results and discussion for the molehill data

In accordance with hypothesis 1, the K2-index indicates significant regularity of the molehills up to scales between 0.1 and 0.4 m for three out of the six patterns (Fig. 3a, c, d). For one pattern, the results show no significant deviation from the null-hypothesis of CSR but a clear trend towards negative autocorrelation (Fig. 3f). This regularity on small scales is probably due to the hypothesised effect of energy optimisation: at short intervals the costs for building new molehills are still higher than for shifting the accumulating amount of soil (Fig. 4). Early studies state that a mole is able to move up to 800 g of soil, which is more than tenfold its own bodyweight (Skoczeǹ 1958). However, a more recent study shows that on average only 20–30 ml of soil are transported at a time, so that for
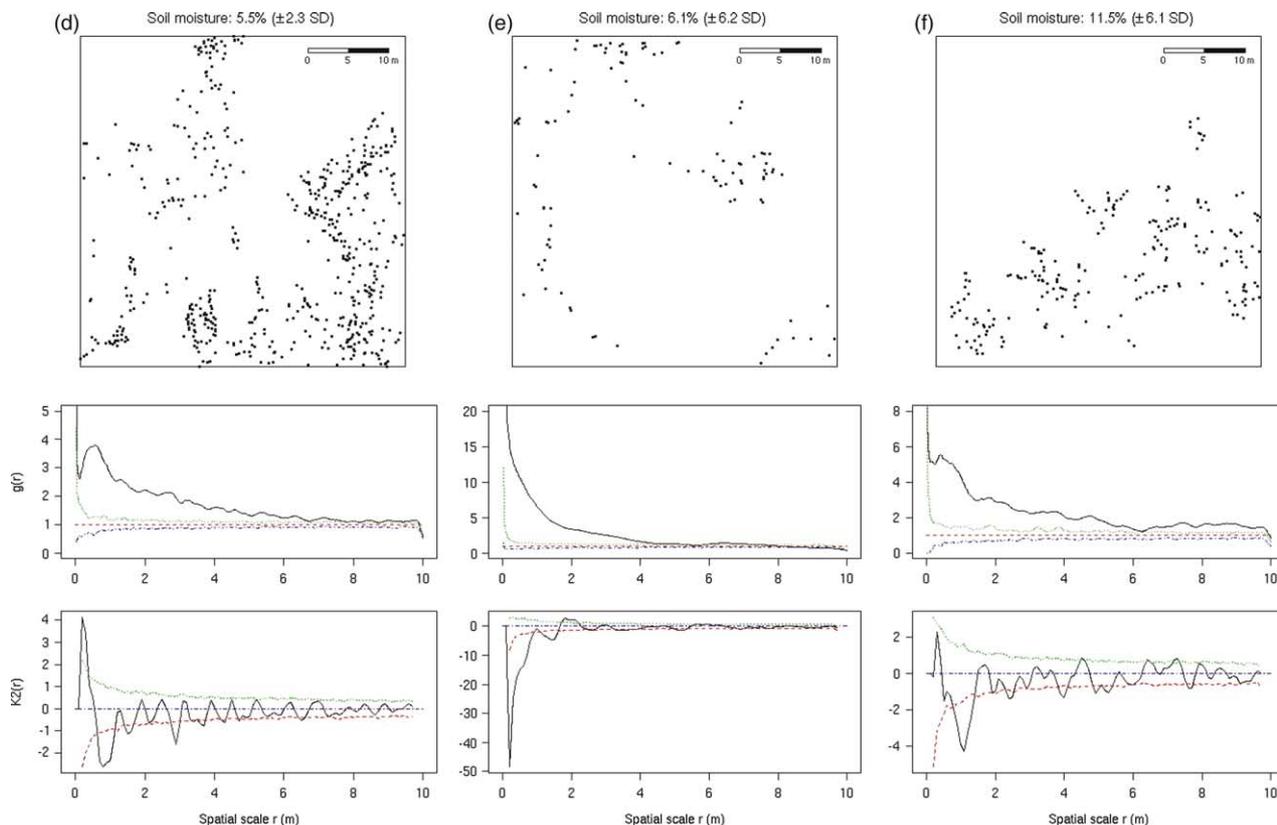

Figure 3. (d–f) Molehill patterns for three study sites with a soil moisture of 5.5% (d), 6.1% (e), and 11.5% (f). The dotted lines in the results for the L-, g-, and K2-analyses indicate 95% confidence intervals.
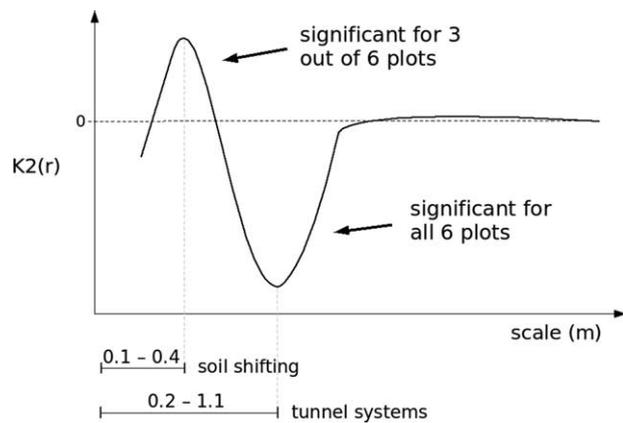
Figure 4. Summary of the results of the K2-analyses for the molehill patterns in relation to the a priori hypotheses.

the construction of one molehill several hundred soil transports have to be carried out (Matthias and Witte 1987). Hence, the average distance to the next molehill should have a strong impact on the time and energy a mole spends for building the tunnel system and an optimisation of the frequency of molehill creation respective to the soil conditions is very likely. In our study plots, the average distance between two neighbouring molehills varies between 0.2 and 1.1 m. At these scales the results of the K2-analysis show a significant transition to higher densities for all plots, as suggested by hypothesis 2 (Fig. 3a–f, and Fig. 4).

The general existence of aggregation is supported by the estimation of the g-function and is caused by the limited movement distance of moles and the sequential construction of hills when building new tunnels. The fact that the scales of significant departures from the null-hypothesis are markedly higher in the g-functions is due to the effect of virtual aggregation and does not contradict the results of the K2-index. The consecutive fluctuations of the values around zero may be due to indirect effects comparable to the results of the simulated repulsion-attraction-pattern (Fig. 2f).

A reasonable explanation of the relatively strong variability in the scale of aggregation would be the different soil moisture of the plots, as suggested in hypothesis 3. However, we found no clear relation between moisture and the scales of autocorrelation so that this explanation does not seem to hold for our study system. This may be because in wetter soils, not only is soil transport more costly

but also the construction of molehills themselves. An alternative explanation for the high variability in the scales of interactions might be that neighbouring molehills do not have to be built in sequence. Often territories of different individuals overlap to some extent and, besides the sharing of common tunnels, this overlap also involves a vertical arrangement of separate tunnels in the soil column (Stone and Gorman 1985). To our knowledge, there are no further investigations on second order characteristics of molehill patterns at all, so that we cannot compare our results to any reference studies. The most closely related work is probably the one by Klaas et al. (2000). They investigated the spatio-temporal pattern of gopher-mounds in Iowa, USA using Ripley's K-analysis for marked point patterns. For time periods of ca 2 weeks they found that mound production was highly clustered at scales of <8 m. However, because of the different biology and social structure of gophers and moles, the ecological interpretation of the underlying processes of the patterns cannot be directly inferred.

In this context, we want to highlight that an analysis of point patterns is not only valuable for accessing information on the underlying process, but also for estimating potential effects of the pattern characteristics on the affected ecosystem. For example, van Hulzen et al. (2006) proved that also relatively small disturbances may have a large effect on the ecosystems diversity, if the distribution of the disturbances leads to "disturbance hot spots".

Conclusions

Our comparison of the L-function, the g-function, and the new K2-index clearly shows that the approaches quantify different aspects of point patterns. Even though all three methods are based on the same information – namely the set of all inter-point distances – they provide insights into complementary aspects of the patterns. The L-analysis shows the neighbourhood intensity up to the observed scale in relation to the intensity of the whole study area, while the g-function illustrates the neighbourhood intensity at the observed scales in relation to the intensity of the study area. However, in heterogeneous point patterns or patterns where processes act on more than one scale, the results of processes acting at small distances are concealed by the "virtual aggregation" effect. The K2-index overcomes this problem by relating the intensity of points at a given scale to the intensity at the next higher scale. In this way, the distribution of events within patches can be tested against a

Table 1. Overview of potential ecological questions and the respective methods for the analysis of heterogeneous point patterns.

Index	Biased by virtual aggregation	Meaning of scale r, where value deviates from CSR	Example of an appropriate application in ecological research
K- and L-function	yes	Radius of neighbourhood area for which neighbourhood intensity is not equal to average intensity.	Which is the optimal size of a zone of influence of a plant in terms of low neighbour density?
g-function	yes	Distance from observed points where neighbourhood intensity is not equal to average intensity.	What is the optimal dispersal distance for species inhabiting habitats with point pattern structure?
K2-index	no	Transition from lower to higher intensities or vice versa/distance up to which attraction or repulsion occurs.	Up to which distances do interactions among individuals occur?

null-hypothesis and the K2-index has the potential to detect the correct scales of interactions also in the case of multiple processes. However, it has to be emphasised that to the same extent as the sensitivity for interactions increases it becomes more susceptible to random noise, as well. Therefore, the decision as to which method to chose should be based on 1) the size of the dataset, since a high number of events can mitigate the effects of the higher susceptibility of the K2-index to random noise, and 2) the ecological phenomenon and exact problem to be investigated.

An overview over potential ecological questions and the respective methods for the analysis of heterogeneous point patterns is given in Table 1.

Since most ecological patterns are heterogeneous to some extent, a broad range of studies will benefit from the use of the K2-index. Nevertheless, we argue for the use of a plurality of approaches – most studies will profit from the application of more than one method and from a deliberate interpretation of the results in the case of heterogeneity.

References

Baddeley, A. J. and Turner, R. 2005. spatstat: An R package for analyzing spatial point patterns. – J. Stat. Soft 12: 1–42.

Baddeley, A. J. et al. 2000. Non and semi-parametric estimation interaction in inhomogeneous point patterns. – Stat. Neerlandica 54: 329–350.

Bailey, T. C. and Gatrell, A. C. 1995. Interactive spatial data analysis. – Longman, Harlow.

Barot, S. et al. 1999. Demography of a savanna palm tree: predicitons from comprehensive spatial pattern analyses. – Ecology 80: 1987–2005.

Besag, J. 1977. Contribution to the discussion of Dr. Ripley's paper. – J. R. Stat. Soc. 39: 193–195.

Bossdorf, O. et al. 2000. Spatial patterns of plant association in grazed and ungrazed shrublands in the semi-arid Karoo, South Africa. – J. Veg. Sci. 11: 253–258.

Condit, R. et al. 2000. Spatial patterns in the distribution of tropical tree species. – Science 288: 1414–1418.

Coomes, D. A. et al. 1999. Identifying aggregation and association in fully mapped spatial data. – Ecology 80: 554–565.

Dale, M. R. T. and Powell, R. D. 2001. A new method for characterizing point patterns in plant ecology. – J. Veg. Sci. 12: 597–608.

Dale, M. R. T. et al. 2002. Conceptual and mathematical relationships among methods for spatial analysis. – Ecography 25: 558–577.

Diggle, P. J. 2003. Statistical analysis of point patterns, 2nd ed. – Arnold.

Dungan, J. L. et al. 2002. A balanced view of scale in spatial statistical analysis. – Ecography 25: 626–640.

Freeman, E. A. and Ford, E. D. 2002. Effects of data quality on analysis of ecological pattern using the (d) statistical function. – Ecology 83: 35–46.

Getis, A. and Franklin, J. 1987. Second-order neighborhood analysis of mapped point patterns. – Ecology 68: 473–477.

Getzin, S. et al. 2006. Spatial patterns and competition of tree species in a Douglas-fir chronosequence on Vancouver Island. – Ecography 29: 671–682.

Goreaud, F. and Pèlissier, R. 1999. On explicit formulas of edge effect correction for Ripley's K-function. – J. Veg. Sci. 10: 433–438.

Grimm, V. et al. 1996. Pattern-oriented modelling in population ecology. – Sci. Total Environ. 183: 151–166.

Gustafson, E. J. 1998. Quantifying landscape spatial pattern: what is the state of the art? – Ecosystems 1: 143–156.

Haase, P. 1995. Spatial pattern analysis in ecology based on Ripley's K-function-introduction and methods of edge correction. – J. Veg. Sci. 6: 575–582.

Haase, P. 2001. Can isotropy vs. anisotropy in the spatial association of plant species reveal physical vs. biotic facilitaiton? – J. Veg. Sci. 12: 127–136.

Jeltsch, F. et al. 1999. Detecting process from snapshot pattern: lessons from tree spacing in the southern Kalahari. – Oikos 85: 451–466.

Klaas, B. A. et al. 2000. The tempo and mode of gopher mound production in a tallgrass prairie remnant. – Ecography 23: 246–256.

Law, R. and Dieckmann, U. 2000. A dynamical system for neighbourhoods in plant communities. – Ecology 8: 2137–2148.

Levin, S. A. 1992. The problem of pattern and scale in ecology. – Ecology 73: 1943–1967.

Liebhold, A. M. and Gurevitch, J. 2002. Integrating the statistical analysis of spatial data in ecology. – Ecography 25: 553–557.

Martens, S. N. et al. 1997. Scales of above-ground and below-ground competition in a semi-arid woodland detected from spatial pattern. – J. Veg. Sci. 8: 655–664.

Mattfeldt, M. et al. 2006. Statistical analysis of reduced pair correlation functions of capillaries in the prostate gland. – J. Microsc. 223: 107–119.

Matthias, G. and Witte, G. R. 1987. Untersuchung zur Genese des Maulwurfhaufens bei *Talpa europaea* (L.). Säugetierkunde. – Paul Parey Verlag.

Mellanby, K. 1967. Food and activity in the mole *Talpa europaea*. – Nature 215: 1128–1130.

Molofsky, J. et al. 2002. Negative frequency dependence and the importance of spatial scale. – Ecology 83: 21–27.

Pélissier, R. and Goreaud, F. 2001. A practical approach to the study of spatial structure in simple cases of heterogeneous vegetation. – J. Veg. Sci. 12: 99–108.

Pélissier, R and Goreaud, F. 2007. ads: spatial point patterns analysis. – R-package version 1.2–7.

Penttinen, A. et al. 1992. Marked point-processes in forest statistics. – For. Sci. 38: 806–824.

Perry, J. N. et al. 2002. Illustrations and guidelines for selecting statistical methods for quantifying spatial patterns in ecological data. – Ecography 25: 578–600.

Perry, L. W. G. et al. 2006. A comparison of methods for the statistical analysis of spatial point pattern in plnat ecology. – Plant Ecol. 187 : 59–82.

Podani, J. and Czárán, T. 1997. Individual-centered analysis of mapped point patterns representing multi-species assemblages. – J. Veg. Sci. 8: 259–270.

R Development Core Team 2007. R: a language and environment for statistical computing. – R foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org >.

Revilla, E. and Palomares, F. 2002. Spatial organization, group living and ecological correlates in low-density populations of Eurasian badgers, *Meles meles*. – J. Anim. Ecol. 71: 497–512.

Ripley, B. D. 1981. Spatial statistics. – Wiley.

Schurr, F. M. et al. 2004. Spatial pattern formation in semi-arid shrubland: a priori predicted versus observed pattern characteristics. – Plant Ecol. 173: 271–282.

Skoczeǹ, S. 1958. Tunnesl digging by the mole (*Talpa europaea* Linne). – Acta Theriol. 2: 235–249.

Spooner, P. G. et al. 2004. Spatial analysis of roadside *Acacia* populations on a road network using the network K-function. – Landscape Ecol. 19: 491–499.

Stoll, P. and Wiener, J. 2000. A neighborhood view of interactions among individual plants. – In: Dieckmann, U. et al. (eds), The geometry of ecological interactions: simplifying spatial complexity. Cambridge Univ. Press, pp. 11–27.

Stone, R. D. and Gorman, M. L. 1985. Social organization of the European mole (*Talpa europaea*) and the Pyrenean desman (*Galemys pyrenaicus*). – Mammal. Rev. 15: 35–42.

Stoyan, D. and Stoyan, H. 1994. Fractals, random shapes and point fields. Methods of geometrical statistics. – Wiley.

Stoyan, D. and Penttinen, A. 2000. Recent application of point process methods in forest statistics. – Stat. Sci. 15: 61–78.

Strand, E. K. et al. 2007. Spatial patterns on the sagebrush steppe/ western juniper ecotone. – Plant Ecol. 190: 159–173.

van Hulzen, J. B. et al. 2006. The significance of spatial and temporal patterns of algal mat deposition in structuring salt marsh vegetation. – J. Veg. Sci 17: 291–298.

van Teeffelen, A. J. A. and Ovaskainen, O. 2007. Can the cause of aggregation be inferred from species distributions? – Oikos 116: 4–16.

Wagner, H. H. and Fortin, M.-J. 2005. Spatial analysis of landscapes: concepts and statistics. – Ecology 86: 1975–1987.

Wiegand, K. et al. 2000. Do spatial effects play a role in the spatial distribution of desert-dwelling *Acacia raddiana*? – J. Veg. Sci. 11: 473–484.

Wiegand, T. and Moloney, K. A. 2004. Rings, circles, and null-models for point pattern analysis in ecology. – Oikos 104: 209–229.

Wiegand, T. et al. 1998. Populations dynamics, disturbances, and pattern evolution: identifying the fundamental scales of organization in a model ecosystem. – Am. Nat. 152: 321–337.

Wiegand, T. et al. 1999. Finding the missing link between landscape structure and population dynamics: a spatially explicit perspective. – Am. Nat. 154: 605–627.

Wiegand, T. et al. 2006. Extending point pattern analysis for objects of finite size and irregular shape. – J. Ecol. 94: 825–837.

Wiens, J. A. et al. 1993. Ecological mechanisms and landscape ecology. – Oikos 66: 369–380.

Download the Supplementary material as file E5374 from <www.oikos.ekol.lu.se/appendix>.