



Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Overall and pairwise segregation tests based on nearest neighbor contingency tables

Elvan Ceyhan*

Department of Mathematics, Koç University, 34450 Sarıyer, Istanbul, Turkey

ARTICLE INFO

Article history:

Available online 6 August 2008

ABSTRACT

Multivariate interaction between two or more classes (or species) has important consequences in many fields and may cause multivariate clustering patterns such as spatial segregation or association. The spatial *segregation* occurs when members of a class tend to be found near members of the same class (i.e., near conspecifics) while spatial *association* occurs when members of a class tend to be found near members of the other class or classes. These patterns can be studied using a nearest neighbor contingency table (NNCT). The null hypothesis is randomness in the nearest neighbor (NN) structure, which may result from – among other patterns – random labeling (RL) or complete spatial randomness (CSR) of points from two or more classes (which is called the *CSR independence*, henceforth). New versions of overall and cell-specific tests based on NNCTs (i.e., NNCT-tests) are introduced and compared with Dixon's overall and cell-specific tests and various other spatial clustering methods. Overall segregation tests are used to detect any deviation from the null case, while the cell-specific tests are post hoc pairwise spatial interaction tests that are applied when the overall test yields a significant result. The distributional properties of these tests are analyzed and finite sample performance of the tests are assessed by an extensive Monte Carlo simulation study. Furthermore, it is shown that the new NNCT-tests have better performance in terms of Type I error and power estimates. The methods are also applied on two real life data sets for illustrative purposes.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Multivariate clustering patterns such as segregation and association may result from multivariate interaction between two or more classes (or species). Such patterns are of interest in ecological sciences and other application areas. See, for example, Pielou (1961), Whipple (1980) and Dixon (1994, 2002). For convenience and generality, we refer to the different types of points as “classes” (which are widely termed as ‘marks’ in the point pattern analysis literature (Cressie, 1993)), but the class can stand for any characteristic of an observation at a particular location. For example, the spatial segregation pattern has been investigated for *species* (Diggle, 2003), *age classes* of plants (Hamill and Wright, 1986), and *sexes* of dioecious plants (Nanami et al., 1999). Many of the epidemiological applications are for a two-class system of case and control labels (Waller and Gotway, 2004). For simplicity, we discuss the spatial clustering patterns between two and three classes only; the extension to the case with more classes is straightforward. The null pattern is usually one of the two (random) pattern types: *random labeling* (RL) or *complete spatial randomness* (CSR). We consider two major types of spatial clustering patterns as alternatives: *segregation* and *association*. Segregation (association) occurs when members of a given class have NNs that are more (less) frequently of the same (other) class than would be expected if there were randomness in the NN structure.

* Tel.: +90 212 338 1845; fax: +90 212 338 1559.

E-mail address: elceyhan@ku.edu.tr.

In statistical and other literature, many univariate and multivariate spatial clustering tests have been proposed (Kulldorff, 2006). These include comparison of Ripley's K and L -functions (Ripley, 2004), comparison of nearest neighbor (NN) distances (Cuzick and Edwards, 1990; Diggle, 2003), and analysis of nearest neighbor contingency tables (NNCTs) which are constructed using the NN frequencies of classes (Pielou, 1961; Meagher and Burdick, 1980). Pielou (1961) proposed various tests and Dixon (1994) introduced an overall test of segregation, cell-specific and class-specific tests based on NNCTs in a two-class setting and extended his tests to multi-class case in (Dixon, 2002). For the two-class case, Ceyhan (2006) discussed these tests and demonstrated that Pielou's test is liberal under CSR independence or RL and only appropriate for a random sample of (base, NN) pairs. Furthermore, he introduced three new overall segregation tests based on NNCTs and suggested an empirical adjustment for Pielou's test (Ceyhan, 2008a).

In this article, we introduce a new overall segregation test and new cell-specific tests of segregation based on NNCTs for testing spatial clustering patterns in a multi-class setting. We compare these tests with Dixon's NNCT-tests which are introduced for testing against the RL of points (Dixon, 1994). We extend the use of these tests for the CSR independence pattern also. We also compare the NNCT-tests with Ripley's K or L -functions and pair correlation function $g(t)$ (Stoyan and Stoyan, 1994), which are methods for second-order analysis of point patterns. We only consider *completely mapped data*; i.e., for our data sets, the locations of all events in a defined area are observed. We show through simulation that the newly proposed cell-specific tests perform better (in terms of empirical size and power) than Dixon's cell-specific tests. Likewise the new overall test tends to have higher power compared to Dixon's overall test under segregation of the classes. Furthermore, we demonstrate that NNCT-tests and Ripley's L -function and related methods (i.e., second-order analysis) answer different questions about the pattern of interest.

We provide the null and alternative patterns in Section 2, describe the NNCTs in Section 3, provide the cell-specific tests in Section 4, overall tests in Section 5, empirical significance levels in the two- and three-class cases in Section 6, rejection rates of the tests under various Poisson processes in Section 7, empirical power comparisons under the segregation in Section 8, under the association alternatives in Section 9, examples in Section 10, and our conclusions and guidelines for using the tests in Section 11.

2. Null and alternative patterns

In the univariate spatial point pattern analysis, the null hypothesis is usually *complete spatial randomness (CSR)* (Diggle, 2003). To investigate the spatial interaction between two or more classes in a multivariate process, usually there are two benchmark hypotheses: (i) *independence*, which implies two classes of points are generated by a pair of independent univariate processes and (ii) *random labeling (RL)*, which implies that the class labels are randomly assigned to a given set of locations in the region of interest (Diggle, 2003). In this article, we will consider two random pattern types as our null hypotheses: CSR of points from two classes (this pattern will be called the *CSR independence*, henceforth) or RL. In the CSR independence pattern, points from each of the two classes independently satisfy the CSR pattern in the region of interest.

Our null hypothesis is

H_0 : randomness in the NN structure.

Although CSR independence and RL are not same, they lead to the same null model in NNCT-tests, since a NNCT does not require spatially-explicit information. That is, when the points from two classes are assumed to be independently uniformly distributed over the region of interest, i.e., under the CSR independence pattern, or when only the labeling (or marking) of a set of fixed points is considered, i.e., under the RL pattern, there is randomness in the NN structure. The distinction between CSR independence and RL is very important when defining the appropriate null model in practice. Goreaud and Pélissier (2003) state that CSR independence implies that the two classes are *a priori* the result of different processes (e.g., individuals of different species or age cohorts), whereas RL implies that some processes affect *a posteriori* the individuals of a single population (e.g., diseased vs. non-diseased individuals of a single species). We provide the differences in the proposed tests under these two patterns. For a more detailed discussion of CSR and RL patterns, see the technical report by Ceyhan (2008b).

However, in the ecological and epidemiological settings, CSR independence is the exception rather than rule. Furthermore, it is conceivable for other models to imply randomness in the NN structure also. We also consider patterns that deviate from stationarity or homogeneity in the point process. In particular, we consider various types of Poisson cluster processes (Diggle, 2003) and other inhomogeneous Poisson processes (Baddeley and Turner, 2005). Randomness in the NN structure will hold if both classes independently follow the same process with points having the same support. For example, in a Poisson cluster process, NN structure will be random if parents are the same for each class which is not realistic in practical situations. If classes have different parent sets, then the Poisson cluster process will imply segregation of the classes. If parents and offspring are treated as two different classes, then Poisson cluster process will imply association of the two classes. Further, if the two classes are from the same inhomogeneous Poisson pattern, again randomness in the NN structure will follow. But when the two classes follow different inhomogeneous Poisson patterns whose point intensities differ in space, it might imply the segregation or association of the classes.

As clustering alternatives, we consider two major types of spatial patterns: *segregation* and *association*. *Segregation* occurs if the NN of an individual is more likely to be of the same class as the individual than to be from a different class; i.e., the members of the same class tend to be clumped or clustered (see, e.g., Pielou (1961)). For instance, one type of plant might not grow well around another type of plant and vice versa. In plant biology, one class of points might represent the coordinates

Table 1
The NNCT for two classes

		NN class		Sum
		Class 1	Class 2	
Base class	Class 1	N_{11}	N_{12}	n_1
	Class 2	N_{21}	N_{22}	n_2
	Sum	C_1	C_2	n

of trees from a species with large canopy, so that other plants (whose coordinates are the other class of points) that need light cannot grow (well or at all) around these trees. See, for instance, Dixon (1994) and Coomes et al. (1999). In epidemiology, one class of points might be the geographical coordinates of residences of cases and the other class of points might be the coordinates of the residences of controls. Furthermore, social and ethnic segregation of residential areas can be viewed as a special type of segregation. Given the locations of the residences, the ethnic identity or social status of the residents can be viewed as class labels assigned randomly or not. For example, the residents of similar social status or same ethnic identity might tend to gather in certain neighborhoods which is an example of segregation as opposed to RL of the residences.

Association occurs if the NN of an individual is more likely to be from another class than to be of the same class as the individual. For example, in plant biology, the two classes of points might represent the coordinates of mutualistic plant species, so the species depend on each other to survive. As another example, the points from one class might be the geometric coordinates of parasitic plants exploiting the other plant whose coordinates are the points of the other class.

The patterns of segregation and association do not only result from multivariate interaction between the classes. It is also conceivable to have either of these patterns without any interaction between the point processes; for example, consider the case where species happen to have the same or different fine-scale habitat preferences. Each of the two patterns of segregation and association are not symmetric in the sense that, when two classes are segregated (or associated), they do not necessarily exhibit the same degree of segregation (or association). For example, when points from each of two classes labeled as X and Y are clustered at different locations, but class X is loosely clustered (i.e., its point intensity in the clusters is smaller) compared to class Y so that classes X and Y are segregated but class Y is more segregated than class X . Similarly, when class Y points are clustered around class X points but not vice versa, classes Y and X are associated, but class Y is more associated with class X compared to the other way around. Although it is not possible to list all of the many different forms of segregation (and association), its existence can be tested by an analysis of the NN relationships between the classes (Pielou, 1961).

3. Nearest neighbor contingency tables

NNCTs are constructed using the NN frequencies of classes. We describe the construction of NNCTs for two classes; extension to multi-class case is straightforward. Consider two classes with labels $\{1, 2\}$. Let N_i be the number of points from class i for $i \in \{1, 2\}$ and n be the total sample size, so $n = N_1 + N_2$. If we record the class of each point and the class of its NN, the NN relationships fall into four distinct categories: $(1, 1)$, $(1, 2)$; $(2, 1)$, $(2, 2)$ where in cell (i, j) , class i is the base class, while class j is the class of its NN. That is, the n points constitute n (base, NN) pairs. Then each pair can be categorized with respect to the base label (row categories) and NN label (column categories). Denoting N_{ij} as the frequency of cell (i, j) for $i, j \in \{1, 2\}$, we obtain the NNCT in Table 1 where C_j is the sum of column j ; i.e., number of times class j points serve as NNs for $j \in \{1, 2\}$. Furthermore, we adopt the convention that random variables are denoted by upper case letters and fixed quantities are denoted by lower case letters. Hence, column sums and cell counts are random, while row sums and the overall sum are fixed quantities in our NNCT.

By construction, if N_{ij} is larger than expected, then class j serves as NN more frequently to class i than expected, which implies segregation if $i = j$ and association of class j with class i if $i \neq j$. On the other hand, if N_{ij} is smaller than expected, then class j serves as NN less frequently to class i than expected, which implies lack of segregation if $i = j$ and lack of association of class j with class i if $i \neq j$.

Pielou (1961) used Pearson's χ^2 test of independence for testing segregation. Due to the ease in computation and interpretation, Pielou's test of segregation is used frequently (Meagher and Burdick, 1980). For example, Pielou's test is used for the segregation between males and females in dioecious species (see, e.g., Herrera (1988) and Armstrong and Irvine (1989)) and between different species (Good and Whipple, 1982). However Pielou's test is not appropriate for completely mapped data (Meagher and Burdick, 1980; Dixon, 1994), since the χ^2 test of independence requires independence between cell-counts (and rows or columns also), which is violated under CSR independence or RL. In fact, independence between cell-counts is violated for spatial data in general and in particular it is violated under the null patterns, so Pielou's test is not of the desired size. In fact, Ceyhan (2006) showed that it is liberal under CSR independence or RL. The inappropriateness of Pielou's test was first noted by Meagher and Burdick (1980) who identify the main source of it to be reflexivity of (base, NN) pairs. A (base, NN) pair (X, Y) is reflexive if (Y, X) is also a (base, NN) pair. As an alternative, they suggest using Monte Carlo simulations for Pielou's test. Dixon (1994) derived the appropriate asymptotic sampling distribution of cell counts using Moran join count statistics (Moran, 1948) and hence the appropriate test which also has a χ^2 -distribution asymptotically.

4. Cell-specific tests of segregation

4.1. Dixon's cell-specific test of segregation

The level of segregation can be estimated by comparing the observed cell counts to the expected cell counts in a NNCT. Dixon demonstrates that under RL, one can write down the cell frequencies as Moran join count statistics (Moran, 1948). He then derives the means, variances, and covariances of the cell counts (i.e., frequencies) (see, Dixon (1994, 2002)).

Under RL, the expected count for cell (i, j) is

$$\mathbf{E}[N_{ij}] = \begin{cases} n_i(n_i - 1)/(n - 1) & \text{if } i = j, \\ n_i n_j/(n - 1) & \text{if } i \neq j, \end{cases} \quad (1)$$

where n_i is the fixed sample size for class i for $i = 1, 2, \dots, q$. Observe that the expected cell counts depend only on the size of each class (i.e., row sums), but not on column sums.

The test statistic suggested by Dixon is given by

$$Z_{ij}^D = \frac{N_{ij} - \mathbf{E}[N_{ij}]}{\sqrt{\mathbf{Var}[N_{ij}]}} \quad (2)$$

where

$$\mathbf{Var}[N_{ij}] = \begin{cases} (n + R) p_{ii} + (2n - 2R + Q) p_{iii} + (n^2 - 3n - Q + R) p_{iiii} - (n p_{ii})^2 & \text{if } i = j, \\ n p_{ij} + Q p_{ijj} + (n^2 - 3n - Q + R) p_{ijij} - (n p_{ij})^2 & \text{if } i \neq j, \end{cases} \quad (3)$$

with p_{xx} , p_{xxx} , and p_{xxxx} are the probabilities that a randomly picked pair, triplet, or quartet of points, respectively, are the indicated classes and are given by

$$\begin{aligned} p_{ii} &= \frac{n_i(n_i - 1)}{n(n - 1)}, & p_{ij} &= \frac{n_i n_j}{n(n - 1)}, \\ p_{iii} &= \frac{n_i(n_i - 1)(n_i - 2)}{n(n - 1)(n - 2)}, & p_{ijj} &= \frac{n_i(n_i - 1)n_j}{n(n - 1)(n - 2)}, \\ p_{iiii} &= \frac{n_i(n_i - 1)(n_i - 2)(n_i - 3)}{n(n - 1)(n - 2)(n - 3)}, & p_{ijij} &= \frac{n_i(n_i - 1)n_j(n_j - 1)}{n(n - 1)(n - 2)(n - 3)}. \end{aligned} \quad (4)$$

Furthermore, R is twice the number of reflexive pairs and Q is the number of points with shared NNs, which occurs when two or more points share a NN. Then $Q = 2(Q_2 + 3Q_3 + 6Q_4 + 10Q_5 + 15Q_6)$ where Q_k is the number of points that serve as a NN to other points k times. Under RL, Q and R are fixed quantities, as they depend only on the location of the points, not the types of NNs. So the sampling distribution is appropriate under RL (see also Remark 5.1) and Z_{ij}^D asymptotically has $N(0, 1)$ distribution. But unfortunately, for $q > 2$ the asymptotic normality of the off-diagonal cells in NNCTs is not rigorously established yet, although Dixon (2002) showed the approximate normality for large samples by extensive Monte Carlo simulations. One-sided and two-sided tests are possible for each cell (i, j) using the asymptotic normal approximation of Z_{ij}^D (Dixon, 1994).

Under CSR independence, the expected cell counts in Eq. (1) and the cell-specific test in Eq. (2) and the relevant discussion are similar to the RL case. The only difference is that under RL, the quantities Q and R are fixed, while under CSR independence they are random. That is, under CSR independence, Z_{ij}^D asymptotically has $N(0, 1)$ distribution conditional on Q and R .

4.2. New cell-specific tests of segregation

In standard cases like multinomial sampling with fixed row totals and conditioning on the column totals, the expected cell count for cell (i, j) in contingency tables is $\mathbf{E}[N_{ij}] = \frac{N_i C_j}{n}$. We first consider the difference $\Delta_{ij} := N_{ij} - \frac{N_i C_j}{n}$ for cell (i, j) . Notice that under RL, $N_i = n_i$ are fixed, but C_j are random quantities and $C_j = \sum_{i=1}^q N_{ij}$, hence $\Delta_{ij} = N_{ij} - \frac{n_i C_j}{n}$. Then under RL, $\mathbf{E}[\Delta_{ij}] = \left(\frac{n_i(n_i - 1)}{(n - 1)} - \frac{n_i^2}{n}\right) \mathbf{I}(i = j) + \left(\frac{n_i n_j}{(n - 1)} - \frac{n_i n_j}{n}\right) \mathbf{I}(i \neq j)$ (Ceyhan, 2008b). Notice that the expected value of Δ_{ij} is not zero under RL. Hence, instead of Δ_{ij} , we suggest the following test statistic:

$$T_{ij} = \begin{cases} N_{ij} - \frac{(n_i - 1)}{(n - 1)} C_j & \text{if } i = j, \\ N_{ij} - \frac{n_i}{(n - 1)} C_j & \text{if } i \neq j. \end{cases} \quad (5)$$

Then we have $\mathbf{E}[T_{ij}] = 0$. For the variance of T_{ij} , we have

$$\mathbf{Var}[T_{ij}] = \begin{cases} \mathbf{Var}[N_{ij}] + \frac{(n_i - 1)^2}{(n - 1)^2} \mathbf{Var}[C_j] - 2 \frac{(n_i - 1)}{(n - 1)} \mathbf{Cov}[N_{ij}, C_j] & \text{if } i = j, \\ \mathbf{Var}[N_{ij}] + \frac{n_i^2}{(n - 1)^2} \mathbf{Var}[C_j] - 2 \frac{n_i}{(n - 1)} \mathbf{Cov}[N_{ij}, C_j] & \text{if } i \neq j, \end{cases} \quad (6)$$

where $\mathbf{Var}[N_{ij}]$ are as in Eq. (3), $\mathbf{Var}[C_j] = \sum_{i=1}^q \mathbf{Var}[N_{ij}] + \sum_{k \neq i} \sum_i \mathbf{Cov}[N_{ij}, N_{kj}]$ and $\mathbf{Cov}[N_{ij}, C_j] = \sum_{k=1}^q \mathbf{Cov}[N_{ij}, N_{kj}]$ with $\mathbf{Cov}[N_{ij}, N_{kl}]$ are as in Eqs. (4)–(12) of Dixon (2002).

As a new cell-specific test, we propose

$$Z_{ij}^N = \frac{T_{ij}}{\sqrt{\mathbf{Var}[T_{ij}]}}. \quad (7)$$

Recall that in the two-class case, each cell count N_{ij} has asymptotic normal distribution (Cuzick and Edwards, 1990). Hence, Z_{ij}^N also converges in law to $N(0, 1)$ as $n \rightarrow \infty$. Moreover, one and two-sided versions of this test are also possible.

Under CSR independence, the distribution of the test statistics above is similar to the RL case. The only difference is that Z_{ij}^N asymptotically has $N(0, 1)$ distribution conditional on Q and R .

Dixon's cell-specific test in (2) depends on the frequencies of (base, NN) pairs (i.e., cell counts), and measures deviations from expected cell counts. On the other hand, the new cell-specific test in (7) can be seen as a difference of two statistics and has expected value is 0 for each cell. For the cell-specific tests, the z-score for cell (i, j) indicates the level and direction of the interaction of spatial patterns of base class i and NN class j . If $Z_{ii}^D > 0$ (or $Z_{ii}^N > 0$) then class i exhibits segregation from other classes and if $Z_{ii}^D < 0$ (or $Z_{ii}^N < 0$) then class i exhibits lack of segregation from other classes. Furthermore, in the multi-class case cell-specific test for cell (i, j) measures the interaction of class j with class i . When $i = j$ this interaction is the segregation for class i , but if $i \neq j$, it is the association of class j with class i . Hence for $i \neq j$ the cell-specific test for cell (i, j) is not symmetric, as interaction of class j with class i could be different from the interaction of class i with class j . The new cell-specific tests use more of the information in the NNCT compared to Dixon's tests, hence they potentially will have better performance in terms of size and power.

In the two-class case, segregation of class i from class j implies lack of association between classes i and j ($i \neq j$) and lack of segregation of class i from class j implies association between classes i and j ($i \neq j$), since $Z_{i1}^D = -Z_{i2}^D$ for $i = 1, 2$. The same holds for the new cell-specific tests, since $Z_{ij}^N = -Z_{2j}^N$ for $j = 1, 2$. In the multi-class case with $q > 2$, a positive z-score for the diagonal cell (i, i) indicates segregation, but it does not necessarily mean lack of association between class i and class j ($i \neq j$), since it could be the case that class i could be associated with one class, yet not associated with another one.

5. Overall tests of segregation

5.1. Dixon's overall test of segregation

In the multi-class case with q classes, combining the q^2 cell-specific tests in Section 4.1, Dixon (2002) suggests the quadratic form to obtain the overall segregation test as follows.

$$C_D = (\mathbf{N} - \mathbf{E}[\mathbf{N}])' \Sigma_D^- (\mathbf{N} - \mathbf{E}[\mathbf{N}]), \quad (8)$$

where \mathbf{N} is the $q^2 \times 1$ vector of q rows of NNCT concatenated row-wise, $\mathbf{E}[\mathbf{N}]$ is the vector of $\mathbf{E}[N_{ij}]$ which are as in Eq. (1), Σ_D is the $q^2 \times q^2$ variance-covariance matrix for the cell count vector \mathbf{N} with diagonal entries being equal to $\mathbf{Var}[N_{ij}]$ and off-diagonal entries being $\mathbf{Cov}[N_{ij}, N_{kl}]$ for $(i, j) \neq (k, l)$. The explicit forms of the variance and covariance terms are provided in Dixon (2002). Moreover, Σ_D^- is a generalized inverse of Σ_D (Searle, 2006) and $'$ stands for the transpose of a vector or matrix. Then under RL, C_D has a $\chi_{q(q-1)}^2$ distribution asymptotically. The test statistics Z_{ij}^D are dependent, hence their squares do not sum to C_D . Under CSR independence, the distribution of C_D is conditional on Q and R .

5.2. A new overall test of segregation

Instead of combining the cell-specific tests in Section 4.1, we can also combine the new cell-specific tests in Section 4.2. Let \mathbf{T} be the vector of $q^2 T_{ij}$ values, i.e.,

$$\mathbf{T} = [T_{11}, T_{12}, \dots, T_{1q}, T_{21}, T_{22}, \dots, T_{2q}, \dots, T_{qq}]',$$

and let $\mathbf{E}[\mathbf{T}]$ be the vector of $\mathbf{E}[T_{ij}]$ values. Note that $\mathbf{E}[\mathbf{T}] = \mathbf{0}$. Hence to obtain a new overall segregation test, we use the following quadratic form:

$$C_N = \mathbf{T}' \Sigma_N^- \mathbf{T}, \quad (9)$$

where Σ_N is the $q^2 \times q^2$ variance-covariance matrix of \mathbf{T} . Under RL, C_N has a $\chi_{(q-1)^2}^2$ distribution asymptotically, since rank of Σ_N is $(q - 1)^2$. Furthermore, the test statistics Z_{ij}^N are dependent, hence their squares do not sum to C_N .

Under RL, the diagonal entries in the variance–covariance matrix Σ_N are $\mathbf{Var}[T_{ij}]$ which are provided in Eq. (6). For the off-diagonal entries in Σ_N , i.e., $\mathbf{Cov}[T_{ij}, T_{kl}]$ with $(i, j) \neq (k, l)$, there are four cases to consider:

Case 1: $i = j$ and $k = l$, then

$$\begin{aligned} \mathbf{Cov}[T_{ii}, T_{kk}] &= \mathbf{Cov}\left[N_{ii} - \frac{(n_i - 1)}{(n - 1)}C_i, N_{kk} - \frac{(n_k - 1)}{(n - 1)}C_k\right] \\ &= \mathbf{Cov}[N_{ii}, N_{kk}] - \frac{(n_k - 1)}{(n - 1)}\mathbf{Cov}[N_{ii}, C_k] - \frac{(n_i - 1)}{(n - 1)}\mathbf{Cov}[N_{kk}, C_i] + \frac{(n_i - 1)(n_k - 1)}{(n - 1)^2}\mathbf{Cov}[C_i, C_k]. \end{aligned} \quad (10)$$

Case 2: $i = j$ and $k \neq l$, then

$$\begin{aligned} \mathbf{Cov}[T_{ii}, T_{kl}] &= \mathbf{Cov}\left[N_{ii} - \frac{(n_i - 1)}{(n - 1)}C_i, N_{kl} - \frac{n_k}{(n - 1)}C_l\right] \\ &= \mathbf{Cov}[N_{ii}, N_{kl}] - \frac{n_k}{(n - 1)}\mathbf{Cov}[N_{ii}, C_l] - \frac{(n_i - 1)}{(n - 1)}\mathbf{Cov}[N_{kl}, C_i] + \frac{(n_i - 1)n_k}{(n - 1)^2}\mathbf{Cov}[C_i, C_l]. \end{aligned} \quad (11)$$

Case 3: $i \neq j$ and $k = l$, then $\mathbf{Cov}[T_{ij}, T_{kk}] = \mathbf{Cov}[T_{kk}, T_{ij}]$, which is essentially Case 2 above.

Case 4: $i \neq j$ and $k \neq l$, then

$$\begin{aligned} \mathbf{Cov}[T_{ij}, T_{kl}] &= \mathbf{Cov}\left[N_{ij} - \frac{n_i}{(n - 1)}C_j, N_{kl} - \frac{n_k}{(n - 1)}C_l\right] \\ &= \mathbf{Cov}[N_{ij}, N_{kl}] - \frac{n_k}{(n - 1)}\mathbf{Cov}[N_{ij}, C_l] - \frac{n_i}{(n - 1)}\mathbf{Cov}[N_{kl}, C_j] + \frac{n_i n_k}{(n - 1)^2}\mathbf{Cov}[C_j, C_l]. \end{aligned} \quad (12)$$

In all the above cases, $\mathbf{Cov}[N_{ij}, N_{kl}]$ are as (Dixon, 2002), $\mathbf{Cov}[N_{ij}, C_l] = \sum_{k=1}^q \mathbf{Cov}[N_{ij}, N_{kl}]$ and $\mathbf{Cov}[C_i, C_j] = \sum_{k=1}^q \sum_{l=1}^q \mathbf{Cov}[N_{ki}, N_{lj}]$.

Under CSR independence, the distribution of C_N is as in the RL case, except that it is conditional on Q and R .

Dixon's overall test combines Dixon's cell-specific tests in one compound summary statistic, while new overall test combines the new cell-specific tests. Hence the new overall test might have better performance in terms of size and power, as it depends on the new cell-specific tests.

Remark 5.1 (*The Status of Q and R under RL and CSR Independence*). Under RL, Q and R are fixed quantities, but under CSR independence they are random. The variances and covariances $\mathbf{Var}[N_{ij}]$ and $\mathbf{Cov}[N_{ij}, N_{kl}]$ and all the quantities depending on these quantities also depend on Q and R . Hence under CSR independence, they are variances and covariances conditional on Q and R . The unconditional variances and covariances can be obtained by replacing Q and R by their expectations.

Unfortunately, given the difficulty of calculating the expectations of Q and R under CSR independence, it is reasonable and convenient to use test statistics employing the conditional variances and covariances even when assessing their behavior under CSR independence. Alternatively, one can estimate the expected values of Q and R empirically and substitute these estimates in the expressions. For example, for the homogeneous planar Poisson process, we have $\mathbf{E}[Q/n] \approx .632786$ and $\mathbf{E}[R/n] \approx 0.621120$ (estimated empirically based on 1000 000 Monte Carlo simulations for various values of n on unit square). When Q and R are replaced by $0.63n$ and $0.62n$, respectively, we obtain the so-called *QR-adjusted* tests. However, as shown in Ceyhan (2008c), QR-adjustment does not improve on the unadjusted versions. \square

6. Empirical significance levels under CSR independence

6.1. The two-class case

First, we consider the two-class case with classes X and Y . We generate n_1 points from class X and n_2 points from class Y both of which are independently uniformly distributed on the unit square, $(0, 1) \times (0, 1)$. Hence, all X points are independent of each other and so are Y points; and X and Y are independent data sets. Thus, we simulate the CSR independence pattern for the performance of the tests under the null case. We generate X and Y points for some combinations of $n_1, n_2 \in \{10, 30, 50, 100\}$ and repeat the sample generation $N_{mc} = 10\,000$ times for each sample size combination in order to obtain sufficient precision of the results in reasonable time. At each Monte Carlo replication, we compute the overall and cell-specific tests. Out of these 10 000 samples the number of significant outcomes by each test at $\alpha = .05$ is recorded. The empirical sizes are calculated as the ratio of number of significant results to the number of Monte Carlo replications, N_{mc} . For example empirical size for Dixon's overall test, denoted by $\hat{\alpha}_D$, for a particular (n_1, n_2) , is calculated as $\hat{\alpha}_D := \sum_{i=1}^{N_{mc}} \mathbf{I}(C_{D,i} \geq \chi_2^2(.05))$ where $C_{D,i}$ is the value of Dixon's overall test statistic for iteration i , $\chi_2^2(.05)$ is the 95th percentile of χ_2^2 distribution, and $\mathbf{I}(\cdot)$ is the indicator function.

We present the empirical significance levels for the NNCT-tests in Table 2, where $\hat{\alpha}_{i,j}^D$ and $\hat{\alpha}_{i,j}^N$ are the empirical significance levels of Dixon's and the new cell-specific tests, respectively, $\hat{\alpha}_D$ is for Dixon's and $\hat{\alpha}_N$ is for the new overall tests of

Table 2

The empirical significance levels for Dixon's and new cell-specific and overall tests in the two-class case under H_0 : CSR independence with $N_{mc} = 10\,000$, n_1, n_2 in $\{10, 30, 50, 100\}$ at the nominal level of $\alpha = .05$

Empirical significance levels under CSR independence						
Sizes	Dixon's		New		Overall	
(n_1, n_2)	$\hat{\alpha}_{1,1}^D$	$\hat{\alpha}_{2,2}^D$	$\hat{\alpha}_{1,1}^N$	$\hat{\alpha}_{2,2}^N$	$\hat{\alpha}_D$	$\hat{\alpha}_N$
(10, 10)	.0454 ^c	.0465	.0452 ^c	.0459 ^c	.0432 ^c	.0484
(10, 30)	.0306 ^c	.0485	.0413 ^c	.0420 ^c	.0440 ^c	.0434 ^c
(10, 50)	.0270 ^c	.0464	.0390 ^c	.0396 ^c	.0482	.0408 ^c
(30, 30)	.0507	.0505	.0443 ^c	.0442 ^c	.0464	.0453 ^c
(30, 50)	.0590 ^ℓ	.0522	.0505	.0510	.0443 ^c	.0512
(50, 50)	.0465	.0469	.0500	.0502	.0508	.0506
(50, 100)	.0601 ^ℓ	.0533	.0514	.0515	.0560 ^ℓ	.0525
(100, 100)	.0493	.0463 ^c	.0485	.0486	.0504	.0489

^c: empirical size significantly less than .05; i.e., the test is conservative. ^ℓ: empirical size significantly larger than .05; i.e., the test is liberal. $\hat{\alpha}_{i,i}^D$ and $\hat{\alpha}_{i,i}^N$ are for the empirical significance levels of Dixon's and the new cell-specific tests, respectively, for $i = 1, 2$; $\hat{\alpha}_D$ is for Dixon's and $\hat{\alpha}_N$ is for the new overall tests of segregation.

segregation. The empirical sizes significantly smaller (larger) than .05 are marked with ^c (^ℓ), which indicate that the corresponding test is conservative (liberal). The asymptotic normal approximation to proportions is used in determining the significance of the deviations of the empirical sizes from .05. For these proportion tests, we also use $\alpha = .05$ as the significance level. With $N_{mc} = 10\,000$, empirical sizes less than .0464 are deemed conservative, greater than .0536 are deemed liberal at $\alpha = .05$ level. Notice that in the two-class case $\hat{\alpha}_{1,1}^D = \hat{\alpha}_{1,2}^D$ and $\hat{\alpha}_{2,1}^D = \hat{\alpha}_{2,2}^D$, since $N_{12} = n_1 - N_{11}$ and $N_{21} = n_2 - N_{22}$. Similarly $\hat{\alpha}_{1,1}^N = \hat{\alpha}_{2,1}^N$ and $\hat{\alpha}_{1,2}^N = \hat{\alpha}_{2,2}^N$, since $T_{11} = -T_{21}$ and $T_{12} = -T_{22}$. So only $\hat{\alpha}_{1,1}^D, \hat{\alpha}_{2,2}^D, \hat{\alpha}_{1,1}^N, \hat{\alpha}_{2,2}^N, \hat{\alpha}_D,$ and $\hat{\alpha}_N$ are presented in Table 2.

Observe that Dixon's cell-specific test for cell (2, 2) (i.e., the diagonal entry with base and NN classes are from the larger class) is about the desired level for almost all sample size combinations. On the other hand, Dixon's cell-specific test for cell (1, 1) (i.e., the diagonal entry for the smaller class) is about the desired level for equal and large samples (i.e., $n_1 = n_2 \geq 30$). is conservative or liberal when sample sizes are different. The new cell-specific tests yield very similar empirical sizes for both cells (1, 1) and (2, 2) and are both conservative when $n_1 \leq 30$ and about the desired level otherwise. However, new cell-specific test for cell (1, 1) is less conservative than that of Dixon's, since T_{11} is less likely to be small because it also depends on the column sum.

Dixon's overall test is about the desired level for equal and large samples (i.e., $n_1 = n_2 \geq 30$), is conservative when at least one sample is small (i.e., $n_i \leq 10$), liberal when sample sizes are large but different (i.e., $30 \leq n_1 < n_2$). It is most conservative for $(n_1, n_2) = (10, 50)$. The new overall segregation test is conservative for small samples and has the desired level for moderate to large samples.

For Dixon's cell-specific tests, if at least one sample size is small, the normal approximation is not appropriate. Dixon (1994) recommends Monte Carlo randomization instead of the asymptotic approximation for the corresponding cell-specific tests when cell counts are ≤ 10 ; and for the overall test when some cell counts are ≤ 10 . When sample sizes are small, $n_i \leq 10$ or large but different $30 \leq n_1 < n_2$ it is more likely to have cell count for cell (1, 1) to be < 5 , however for cell (2, 2) cell counts are usually much larger than 5, hence normal approximation is more appropriate for cell (2, 2). Moreover, we not only vary sample sizes but also the relative abundance of the classes in our simulation study. The differences in the relative abundance of classes seem to affect Dixon's tests more than the new tests. See for example cell-specific tests for cell (1, 1) for sample sizes (30, 50) and (50, 100), where Dixon's test suggests that class X (i.e., class with the smaller size) is more segregated which is only an artifact of the difference in the relative abundance. Likewise, Dixon's overall test seems to be affected more by the differences in the relative abundance. On the other hand, the new tests are more robust to differences in the relative abundance, since they depend on both row and column sums.

6.2. The three-class case

The symmetry in cell counts for rows in Dixon's cell-specific tests and columns in the new cell-specific tests occur only in the two-class case. Therefore, in order to better evaluate the performance of cell-specific tests in the absence of such symmetry, we also consider the three-class case with classes X, Y, and Z under CSR independence. We generate n_1, n_2, n_3 points distributed independently uniformly on the unit square $(0, 1) \times (0, 1)$ from classes X, Y, and Z, respectively, at each of some combinations of $n_1, n_2, n_3 \in \{10, 30, 50, 100\}$ for $N_{mc} = 10\,000$ times. The empirical sizes and the significance of their deviation from .05 are calculated as in Section 6.1.

We plot the empirical significance levels for the tests in Fig. 1. The actual values of empirical significance levels are tabulated in the technical report by Ceyhan (2008b). Notice that when at least one class is small (i.e., $n_i \leq 10$) tests are usually conservative, with the Dixon's cell-specific tests being the most conservative. The empirical sizes for the new cell-specific tests are closer to the nominal level for all sample size combinations, while Dixon's cell-specific tests fluctuate around .05 with larger deviations. In the three-class case, both of the overall tests exhibit similar performance in terms of

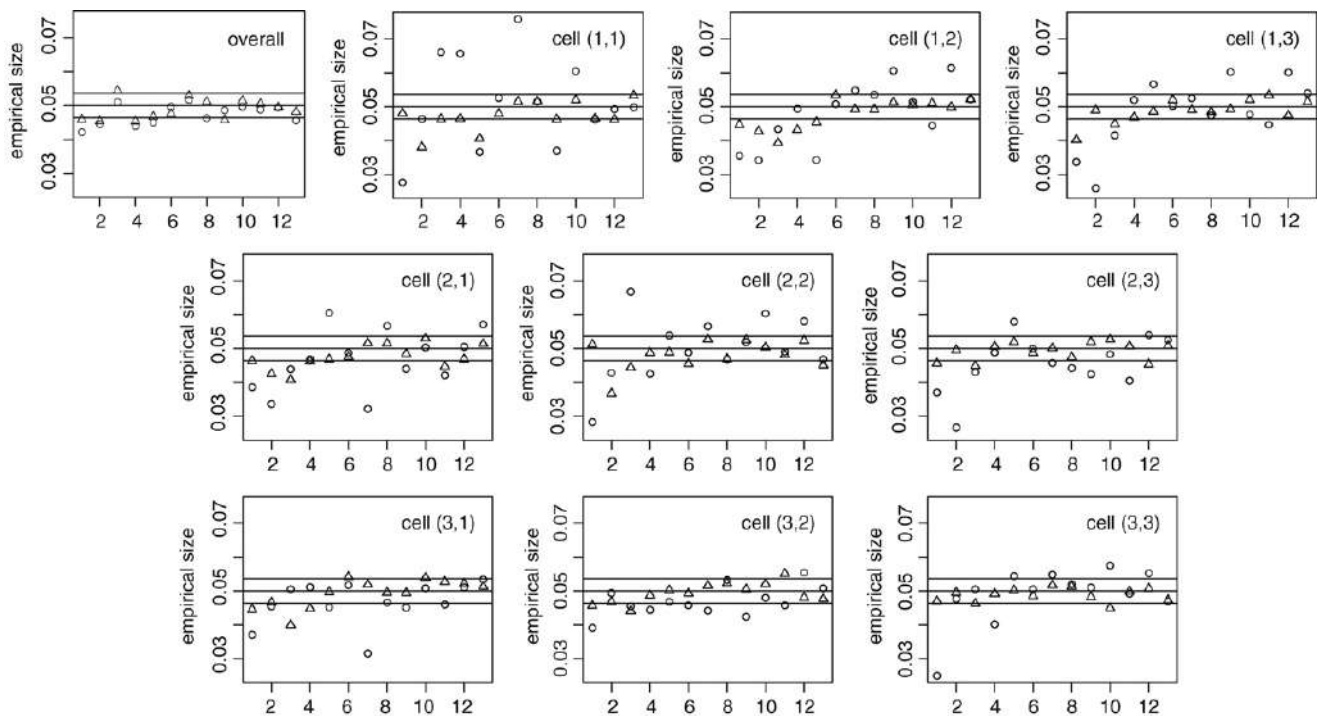


Fig. 1. The empirical size estimates of Dixon's overall and cell-specific tests (circles (o)) and the new overall and cell-specific tests (triangles (Δ)) for cells (1, 1)–(3, 3) under the CSR independence pattern in the three-class case. The horizontal lines are located at .0464 (upper threshold for conservativeness), .0500 (nominal level), and .0536 (lower threshold for liberalness). The horizontal axis labels are: 1 = (10, 10, 10), 2 = (10, 10, 30), 3 = (10, 10, 50), 4 = (10, 30, 30), 5 = (10, 30, 50), 6 = (30, 30, 30), 7 = (10, 50, 50), 8 = (30, 30, 50), 9 = (30, 50, 50), 10 = (50, 50, 50), 11 = (50, 50, 100), 12 = (50, 100, 100), 13 = (100, 100, 100).

empirical size, with Dixon's test being slightly more conservative for small samples. Thus, the new cell-specific tests are more robust to the differences in sample sizes (i.e., differences in relative abundance) and have empirical size closer to .05 compared to Dixon's cell-specific tests.

Remark 6.1 (*Main Result of Monte Carlo Simulations for Empirical Sizes*). Dixon (1994) recommends Monte Carlo randomization test when some cell count(s) are smaller than 10 in a NNCT for his cell-specific tests and when some cell counts are less than 5 for his overall tests and we concur with his suggestion and extend it to the new cell-specific test for cell (i, j) when sum of column j is <10 which happens less frequently than N_{ij} being <10. Dixon's and new overall tests exhibit similar performance in terms of empirical sizes: they are usually conservative for small samples and are about the nominal level otherwise. Furthermore, Dixon's cell-specific and overall tests are confounded by the differences in the relative abundance of the classes. On the other hand, the new cell-specific tests are more robust to differences in relative abundance of the classes. □

Remark 6.2 (*Empirical Significance Levels under RL*). The NNCT-tests we consider are conditional under CSR independence but not under RL. To evaluate their empirical size performance better, we also perform Monte Carlo simulations under three RL cases in the two-class case and two RL cases in the three-class case in Ceyhan (2008b). In these RL cases, we apply RL procedure to points from a homogeneous Poisson process or from various cluster processes. We find that for large samples the new cell-specific tests have better empirical size performance compared to Dixon's cell-specific tests and are more robust to differences in relative abundance. The new and Dixon's overall tests exhibit similar performance under each RL Case. □

7. Finite sample performance of NNCT-tests under various Poisson and inhomogeneous point processes

We provide the finite sample performance of the NNCT-tests under point patterns that are different from CSR independence or RL. In particular, we will consider various versions of Poisson cluster processes and some other inhomogeneous processes (Diggle, 2003).

First Version of (Bivariate) Poisson Cluster Process ($PCP1(n_p, n_1, n_2, \sigma)$): In this process, first we generate n_p parents iid on the unit square, $(0, 1) \times (0, 1)$. Then for each parent n_1/n_p offspring are generated for sample X and n_2/n_p for sample Y from radially symmetric Gaussian distribution with parameter σ . Hence we generate n_1X and n_2Y points, respectively. In the first case, we use the same parent set for both X and Y points. In the second case, we use different parent sets for each of X and Y points.

Second Version of (Bivariate) Poisson Cluster Process ($PCP2(n_p, n_1, n_2, \sigma)$): In this process, we generate n_p parents and n_1X and n_2Y offspring as in the first version PCP1, except the offspring are randomly allocated amongst the parents.

For both versions of the above Poisson cluster processes, we take $\sigma \in \{0.05, .10, .20\}$ and $(n_1, n_2) \in \{(30, 30), (30, 50), (50, 50)\}$.

(Bivariate) Matern Cluster Process ($MCP(\kappa, n_1, n_2, r, \mu)$): In this process, first we generate a Poisson point process of “parent” points with intensity κ . Then each parent point is replaced by a random cluster of points. The number of points in each cluster are random with a $Poisson(\mu)$ distribution, and the points are placed independently and uniformly inside a disc of radius r centered on the parent point. The parent points are not restricted to lie in the unit square; the parent process is effectively the uniform Poisson process on the infinite plane. We consider $\kappa = 5, r \in \{.05, .10, .20\}$ for both X and Y points and $\mu = n_1/5$ for X points and $\mu = n_2/5$ for Y points. In case 1, we use the same parents for both X and Y offspring, while in case 2, we generate different sets of parents with $\kappa = 5$. For each of the above cases, we take $(n_1, n_2) \in \{(50, 50), (50, 100), (100, 100)\}$. For more on Matern cluster processes, see Mat’ern (1986) and Waagepetersen (2007).

(Bivariate) Inhomogeneous Poisson Cluster Process ($IPCP(\lambda_1(x, y), \lambda_2(x, y))$): In this process, the intensity of the Poisson process for the first sample (i.e., sample X) is set to be $\lambda_1(x, y)$ which is a function of (x, y) . We generate a realization of the inhomogeneous Poisson process with intensity function $\lambda_1(x, y)$ at spatial location (x, y) inside the unit square by random “thinning”. That is, we first generate a uniform Poisson process of intensity $\lambda(x, y)$, then randomly delete or retain each point, independently of other points, with retention probability $p(x, y) = \lambda_1(x, y)/\ell_{\max}$ where $\ell_{\max} = \sup_{(x,y) \in (0,1) \times (0,1)} \lambda_1(x, y)$. The second sample (i.e., sample Y) is generated similarly. We take $\lambda_1(x, y) = n_1\sqrt{x+y}$ for sample X for all the three cases we will consider. Then for sample Y , we take $\lambda_2(x, y) = n_2\sqrt{x+y}$ in case 1, $\lambda_2(x, y) = n_2\sqrt{xy}$ in case 2, and $\lambda_2(x, y) = n_2|x-y|$ in case 3. That is, in case 1 X and Y points are from the same inhomogeneous Poisson process; in cases 2 and 3, they are from different processes. For each of the above cases, we take $(n_1, n_2) \in \{(50, 50), (50, 100), (100, 100)\}$. For more on inhomogeneous Poisson cluster processes, see Diggle (2003) and Baddeley and Turner (2005).

The empirical rejection rates of the NNCT-tests are provided in Table 7. Observe that under PCP1 with same parents, the rejection rates are slightly (but significantly) larger than 0.05. Hence under PCP1 with the same parents, the two classes are slightly segregated. Under PCP1 with different parents, the two classes are strongly segregated. Under PCP2 with the same parents, the two classes satisfy randomness in the NN structure, while under PCP2 with different parents, the two classes are strongly segregated. Notice that under these implementations of PCP, the rejection rates decrease as σ increases; i.e., the level of segregation is inversely related to σ . Under MCP with the same parents, the two classes satisfy randomness in NN structure; but with different parents, the classes are strongly segregated. Furthermore, as r increases, the level of segregation decreases under MCP with different parents. Under IPCP patterns, the two classes satisfy randomness in NN structure as long as the density functions are same or similar (see cases 1 and 2); but if the density functions are very different, we observe moderate segregation between the two classes. Notice also that this segregation is detected better by the new NNCT-tests.

8. Empirical power analysis under segregation

8.1. The two-class case

For the segregation alternatives, we generate $X_i \stackrel{iid}{\sim} \mathcal{U}((0, 1-s) \times (0, 1-s))$ and $Y_j \stackrel{iid}{\sim} \mathcal{U}((s, 1) \times (s, 1))$ for $i = 1, \dots, n_1$ and $j = 1, \dots, n_2$. Notice the level of segregation is determined by the magnitude of $s \in (0, 1)$. We consider the following three segregation alternatives:

$$H_S^I: s = 1/6, \quad H_S^{II}: s = 1/4, \quad \text{and} \quad H_S^{III}: s = 1/3. \tag{13}$$

Observe that, from H_S^I to H_S^{III} (i.e., as s increases), the segregation gets stronger in the sense that X and Y points tend to form one-class clumps or clusters more and more frequently. We calculate the power estimates using the asymptotic critical values based on the standard normal distribution for the cell-specific tests and the corresponding χ^2 -distributions for the overall tests.

The power estimates based on the asymptotic critical values are plotted in Fig. 2. The actual values of empirical power estimates are provided in the technical report by (Ceyhan, 2008b). We omit the power estimates of the cell-specific tests for cells (1, 2) and (2, 1), since $\hat{\beta}_{1,1}^D = \hat{\beta}_{1,2}^D$ and $\hat{\beta}_{2,1}^D = \hat{\beta}_{2,2}^D$; and $\hat{\beta}_{1,1}^N = \hat{\beta}_{2,1}^N$ and $\hat{\beta}_{1,2}^N = \hat{\beta}_{2,2}^N$. Observe that, for both cell-specific tests, as $n = (n_1 + n_2)$ gets larger, the power estimates get larger; for the same $n = (n_1 + n_2)$ values, the power estimate is larger for classes with similar sample sizes; and as the segregation gets stronger, the power estimates get larger at each sample size combination. For both cells (1, 1) and (2, 2), the new cell-specific tests have higher power estimates compared to those of Dixon’s. Furthermore, the new overall test has higher power estimates compared to Dixon’s overall test.

8.2. The three-class case

For the segregation alternatives, we generate $X_i \stackrel{iid}{\sim} \mathcal{U}((0, 1-2s) \times (0, 1-2s))$, $Y_j \stackrel{iid}{\sim} \mathcal{U}((2s, 1) \times (2s, 1))$, and $Z_\ell \stackrel{iid}{\sim} \mathcal{U}((s, 1-s) \times (s, 1-s))$ for $i = 1, \dots, n_1, j = 1, \dots, n_2$, and $\ell = 1, \dots, n_3$. Notice that the level of segregation is

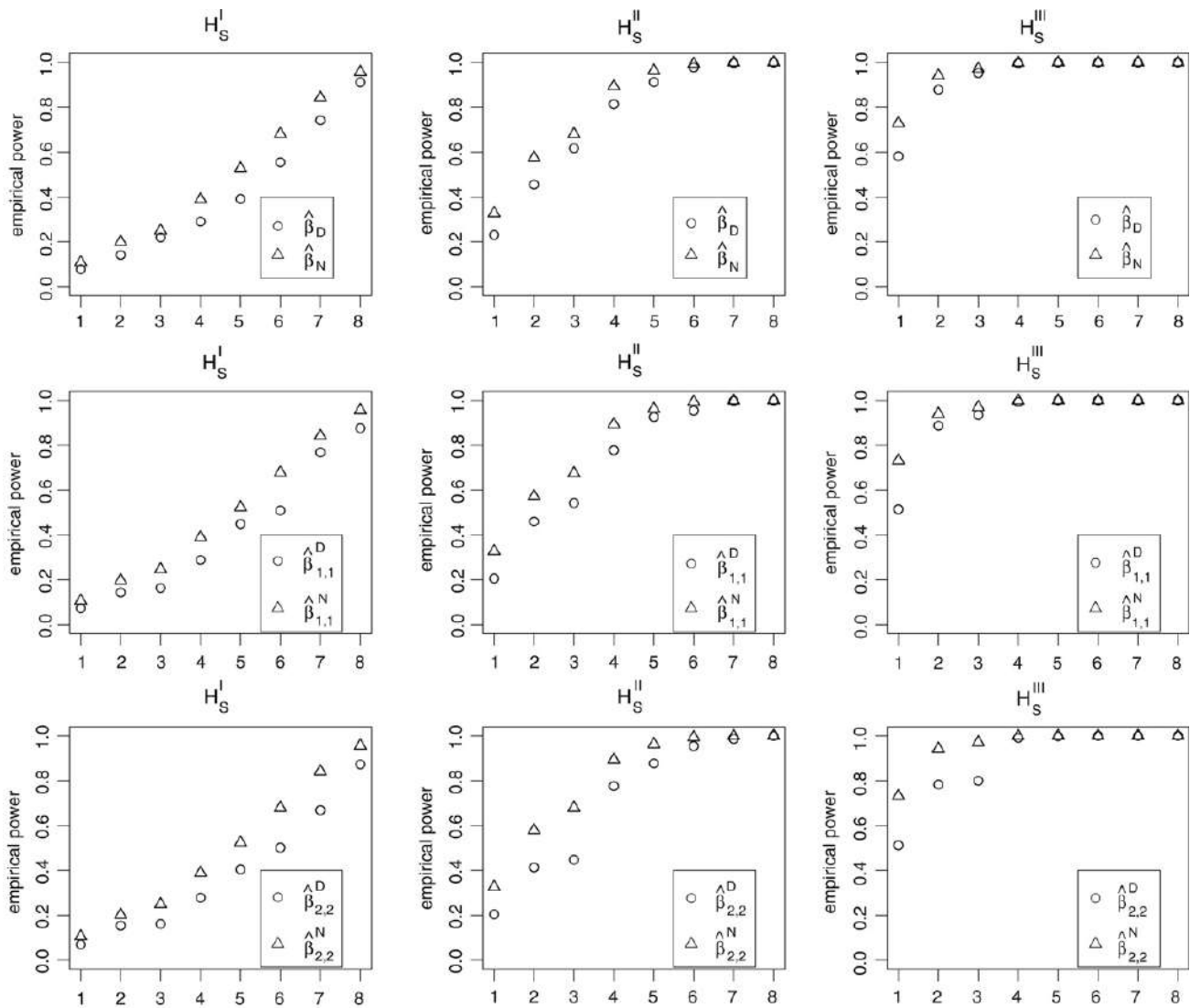


Fig. 2. The empirical power estimates for Dixon's tests (circles (o)) and new tests (triangles (Δ)) under the segregation alternatives in the two-class case. The horizontal axis labels: 1 = (10, 10), 2 = (10, 30), 3 = (10, 50), 4 = (30, 30), 5 = (30, 50), 6 = (50, 50), 7 = (50, 100), 8 = (100, 100). $\hat{\beta}_D$ and $\hat{\beta}_N$ stand for empirical power estimates for Dixon's and new overall tests, respectively. $\hat{\beta}_{ii}^D$ and $\hat{\beta}_{ii}^N$ stand for empirical power estimates for Dixon's and cell-specific tests, respectively, for cell (i, i) with $i = 1, 2$.

determined by the magnitude of $s \in (0, 1/2)$. We consider the following three segregation alternatives:

$$H_{S_1}: s = 1/12, \quad H_{S_2}: s = 1/8, \quad \text{and} \quad H_{S_3}: s = 1/6. \tag{14}$$

Observe that, from H_{S_1} to H_{S_3} (i.e., as s increases), the segregation gets stronger. Furthermore, for each segregation alternative, X and Y are more segregated compared to Z and X or Z and Y .

We plot the empirical power estimates for the cell-specific tests in Fig. 3 and for the overall tests in Fig. 4. The test statistics are mostly positive for diagonal cells which implies segregation of classes and are mostly negative for off-diagonal cells which implies lack of association between classes. For both cell-specific tests for the diagonal cells (i, i) for $i = 1, 2, 3$, as equal sample sizes get larger, the power estimates get larger under each segregation alternative; and as the segregation gets stronger, the power estimates get larger for each cell at each sample size combination. The higher degree of segregation between X and Y is reflected in cells (1, 1) and (2, 2). Furthermore, since the sample sizes satisfy $n_1 \leq n_2$ in our simulation study, cell (2, 2) power estimates tend to be larger. Since class Z is less segregated from the other two classes, cell (3, 3) power estimates tend to be lower than the other diagonal cell statistics. Notice also that off-diagonal cells are more severely affected by the differences in relative abundance of the classes.

The higher degree of segregation between classes X and Y can also be observed in power estimates for cells (1, 2) and (2, 1), since more segregation of these classes imply higher negative values in these cells' test statistics. The lesser degree of segregation between classes X and Z can be seen in cells (1, 3) and (3, 1), as they yield much lower power estimates compared to the other off-diagonal cells. Although Y and Z are segregated in the same degree as X and Z , the power estimates for cells (2, 3) and (3, 2) are larger than those for cells (1, 3) and (3, 1), since $(n_1 + n_3) \leq (n_2 + n_3)$ in our simulation study

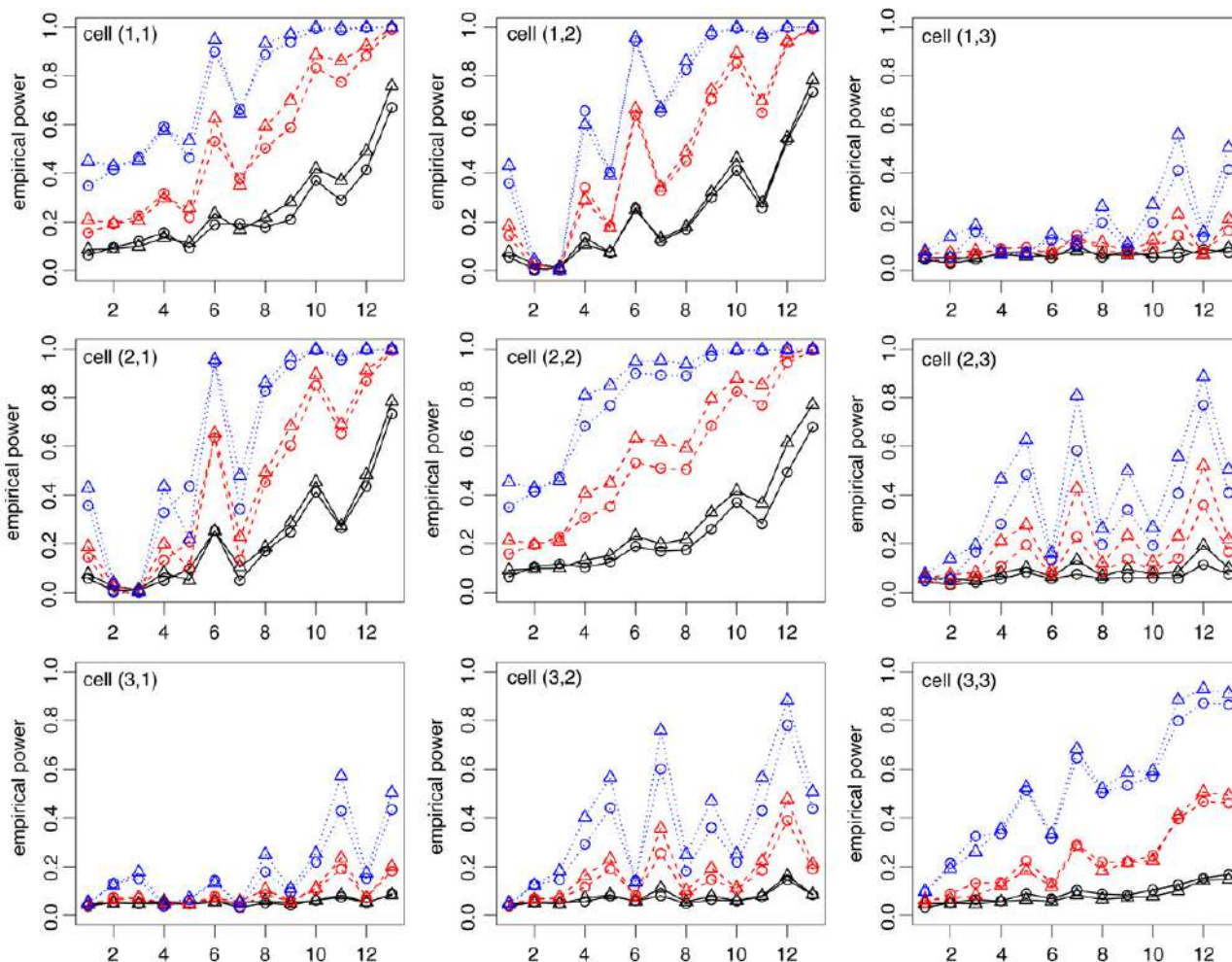


Fig. 3. The empirical power estimates of Dixon's cell-specific tests (circles (o)) and the new cell-specific tests (triangles (Δ)) under the segregation alternatives H_{S_1} (solid lines), H_{S_2} (dashed lines), and H_{S_3} (dotted lines) in the three-class case. The horizontal axis labels are: 1 = (10, 10, 10), 2 = (10, 10, 30), 3 = (10, 10, 50), 4 = (10, 30, 30), 5 = (10, 30, 50), 6 = (30, 30, 30), 7 = (30, 30, 50), 8 = (30, 50, 50), 9 = (50, 50, 50), 10 = (50, 50, 100), 11 = (50, 50, 100), 12 = (50, 100, 100), 13 = (100, 100, 100). Notice that they are arranged in the increasing order for the first and then the second and then the third entries. The size values for discrete sample size combinations are joined by piecewise straight lines for better visualization.

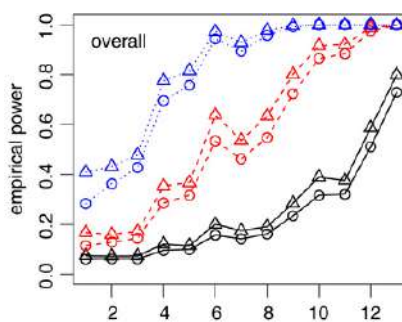


Fig. 4. The empirical power estimates of Dixon's overall test (circles (o)) and the new overall test (triangles (Δ)) under the segregation alternatives H_{S_1} (solid lines), H_{S_2} (dashed lines), and H_{S_3} (dotted lines) in the three-class case. The horizontal axis labels are as in Fig. 3.

and larger sample sizes imply higher power under the same degree of segregation. Furthermore, the power estimates for the new cell-specific tests tend to be higher for each cell under each segregation alternative for each sample size combination. In summary, in the three-class case, new cell-specific tests have better performance in terms of power. The performance of the overall tests are similar to the performance of cell-specific tests for the diagonal cells: power estimates increase as the segregation gets stronger; power estimates increase as the sample sizes increase; and new overall test has higher power than Dixon's overall test.

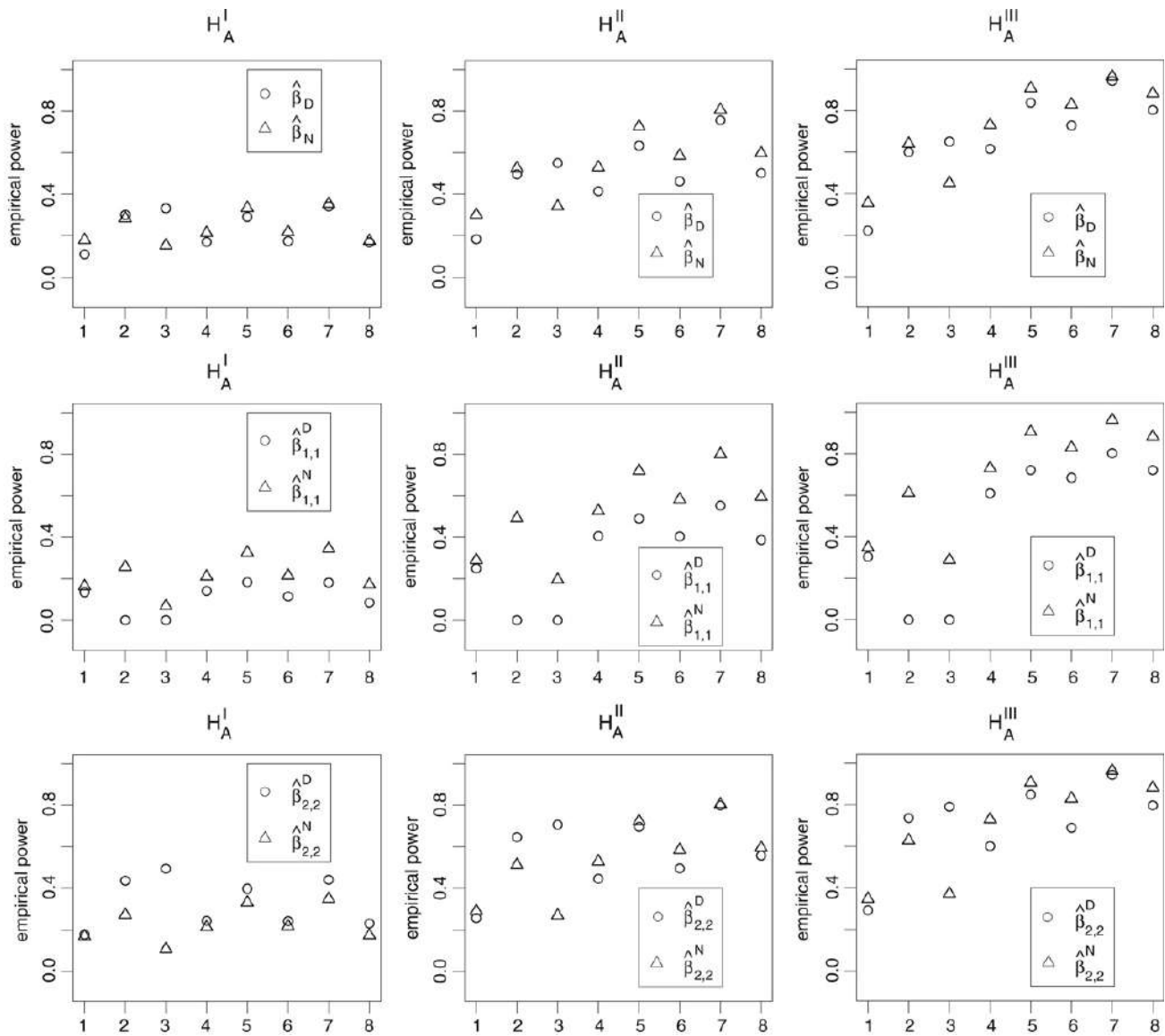


Fig. 5. The empirical power estimates for Dixon's tests (circles (o)) and new tests (triangles (Δ)) under the association alternatives in the two-class case. The horizontal axis labels and power estimate notation are as in Fig. 2.

9. Empirical power analysis under association

9.1. The two-class case

For the association alternatives, we consider three cases also. In each case, first we generate $X_i \stackrel{iid}{\sim} \mathcal{U}((0, 1) \times (0, 1))$ for $i = 1, 2, \dots, n_1$. Then we generate Y_j for $j = 1, 2, \dots, n_2$ as follows. For each j , we pick an i randomly, then generate Y_j as $X_i + R_j^Y (\cos T_j, \sin T_j)'$ where $R_j^Y \stackrel{iid}{\sim} \mathcal{U}(0, r_y)$ with $r \in (0, 1)$ and $T_j \stackrel{iid}{\sim} \mathcal{U}(0, 2\pi)$. In the pattern generated, appropriate choices of r_y will imply association between classes X and Y . That is, it will be more likely to have (X, Y) or (Y, X) NN pairs than same-class NN pairs (i.e., (X, X) or (Y, Y)). The three values of r_y we consider constitute the following association alternatives;

$$H_A^I: r_y = 1/4, \quad H_A^{II}: r_y = 1/7, \quad \text{and} \quad H_A^{III}: r_y = 1/10. \tag{15}$$

Observe that, from H_A^I to H_A^{III} (i.e., as r_y decreases), the association gets stronger in the sense that X and Y points tend to occur together more and more frequently. By construction, for similar sample sizes the association between X and Y are at about the same degree as association between Y and X . For very different samples, larger sample is associated with the smaller but the abundance of the larger sample confounds its association with the smaller.

The empirical power estimates are plotted in Fig. 5. For the actual values of empirical power estimates, see Ceyhan (2008b). Observe that the power estimates increase as the association gets stronger at each sample size combination and

the power estimates increase as the equal sample sizes increase and as the very different sample sizes increase under each association alternative.

Dixon's cell-specific test for cell (1, 1) has extremely poor performance for very different small samples (i.e., $n_1 \leq 10$ and $n_1 \neq n_2$). On the other hand, for larger samples, the empirical power estimates get larger as association gets stronger at each sample size combination. When samples are large, class Y is more associated with class X if $n_2 > n_1$ and this is reflected in the empirical power estimates. The power estimates for the new cell-specific test for cell (1, 1) increase as the association gets stronger and equal sample sizes increase. Both tests have the lowest power estimates for $(n_1, n_2) = (10, 50)$, since cell counts and column sums could be very small for this sample size combination. Dixon's cell-specific test for cell (2, 2) has higher power estimates under weak association compared to those of the new cell-specific test. When association gets stronger, power estimates for Dixon's cell-specific test for cell (2, 2) has higher power for smaller samples and lower power for larger samples compared to the new cell-specific tests. The new cell-specific test has the worst performance for $(n_1, n_2) = (10, 50)$, in which case, column sums could be small. Dixon's overall test has similar power as the new overall test for smaller samples; and new overall test has higher power estimates for larger samples.

9.2. The three-class case

For the association alternatives, we also consider three cases. In each case, first we generate $X_i \stackrel{iid}{\sim} \mathcal{U}((0, 1) \times (0, 1))$ for $i = 1, 2, \dots, n_1$. Then we generate Y_j and Z_ℓ for $j = 1, 2, \dots, n_2$ and $\ell = 1, 2, \dots, n_3$ as in Section 9.1 with parameters r_y and r_z , respectively. In the pattern generated, appropriate choices of r_y (and r_z) values will imply association between classes X and Y (and X and Z). The three association alternatives are

$$H_{A_1}: r_y = 1/7, r_z = 1/10, \quad H_{A_2}: r_y = 1/10, r_z = 1/20, \quad H_{A_3}: r_y = 1/13, r_z = 1/30. \quad (16)$$

Observe that, from H_{A_1} to H_{A_3} (i.e., as r_y and r_z decrease), the association between X and Y gets stronger in the sense that Y points tend to be found more and more frequently around the X points. The same holds for X and Z points. Furthermore, by construction, classes X and Z are more associated compared to classes X and Y.

We plot the power estimates for the cell-specific tests in Fig. 6 and for the overall tests in Fig. 7. The test statistics tend to be negative for the diagonal cells, which implies lack of segregation for the classes; positive for cells (1, 2), (2, 1), (1, 3), and (3, 1), which implies association between classes X and Y and between classes X and Z; negative for cells (2, 3) and (3, 2), which implies lack of association (and perhaps mild segregation) between classes Y and Z.

At each sample size combination, as the association gets stronger, the power estimates increase. The higher degree of association between X and Z compared to that of X and Y are reflected in higher power estimates for cell (1, 3) compared to cell (1, 2). For the same reason, power estimates for cell (3, 1) are higher than those for cell (2, 1). For cells (1, 2), (2, 1), (1, 3), and (3, 1) the power estimates get larger as the equal sample sizes increase. The new cell-specific tests have higher power for cells (1, 2) and (1, 3) and Dixon's cell-specific tests have higher power for cells (2, 1) and (3, 1). By construction the classes Y and Z are not associated, instead they can be viewed as mildly segregated from each other. The test statistics for cells (2, 3) and (3, 2) are negative to indicate such segregation or lack of association between classes Y and Z. However, cell (3, 2) power estimates are much larger than cell (3, 2), which implies that class Z can be viewed as more segregated from class Y. As for the diagonal cells, the higher power estimates for cell (1, 1) for larger samples are indicative of high degree of association of classes Z and Y with class X. Such association, by construction, is barely reflected in cells (2, 2) and (3, 3).

The power estimates for the overall tests tend to increase as the association gets stronger at each sample size combination; as the sample sizes tend to increase, except for the sudden decrease at $(n_1, n_2, n_3) = (10, 50, 50)$, in which case cell (1, 1) counts and column 1 sums tend to be very small. Furthermore, Dixon's overall test has higher power compared to the new overall test. So Dixon's overall test is recommended for the association alternative over the new overall test.

Remark 9.1 (*Main Result of Monte Carlo Power Analysis*). Considering the empirical significance levels and power estimates, for small samples we recommend Monte Carlo randomization tests. For large samples, when testing against the segregation alternatives, we recommend the new overall and cell-specific tests over the use of Dixon's tests, as the new tests either have about the same power as or have larger power than Dixon's tests. Furthermore, if one wants to see the level of segregation between pairs of classes, we recommend using the diagonal cells, (i, i) for $i = 1, 2, 3$ as they are more robust to the differences in class sizes (i.e., relative abundance) and more sensitive to the level of segregation. When testing against the association alternatives, we recommend both Dixon's and new overall and cell-specific tests. \square

Remark 9.2 (*Power Comparison Based on Monte Carlo Critical Values*). When some cell counts or column sums are < 5 (or < 10), then it will not be appropriate to use the asymptotic approximation hence the asymptotic critical values for the overall (or cell-specific) tests of segregation. See Remark 6.1 for further detail. In order to better evaluate the empirical power performance of the tests, for each sample size combination, we generate two and three sets of points under the CSR independence cases of Sections 6.1 and 6.2, respectively. We find the 95th percentiles of the recorded test statistics at each sample size combination and use them as "Monte Carlo critical values" for the power estimation in Ceyhan (2008b). We find that the power estimates based on the Monte Carlo critical values under segregation or association alternatives are very similar to – but tend to be slightly larger – compared to the ones using the asymptotic critical values. Furthermore,

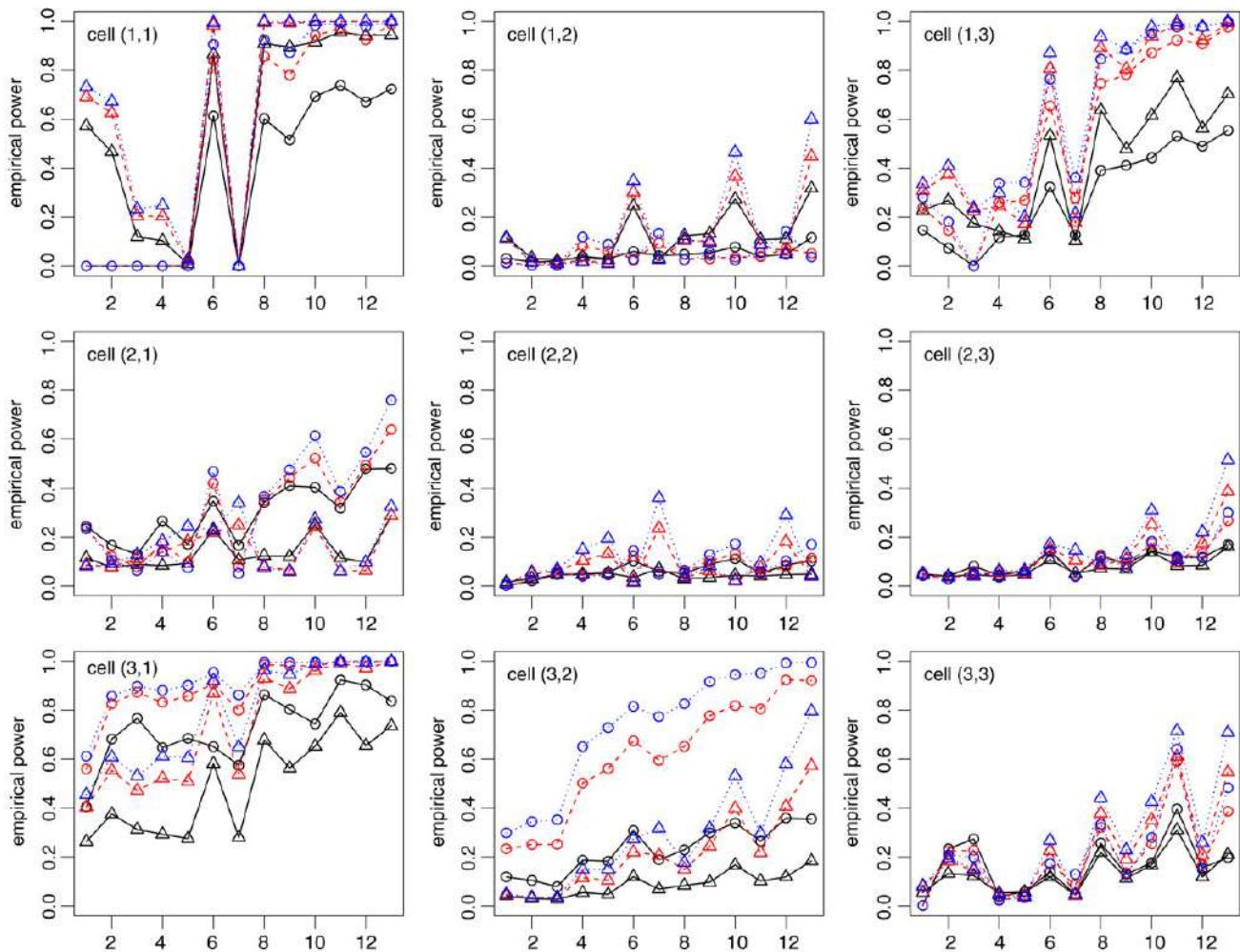


Fig. 6. The empirical power estimates of Dixon's cell-specific tests (circles (o)) and the new cell-specific tests (triangles (Δ)) under the association alternatives H_{A_1} (solid lines), H_{A_2} (dashed lines), and H_{A_3} (dotted lines) in the three-class case. The horizontal axis labels are as in Fig. 3.

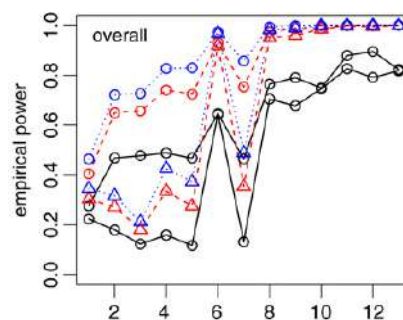


Fig. 7. The empirical power estimates of Dixon's overall test (circles (o)) and the new overall test (triangles (Δ)) under the association alternatives H_{A_1} (solid lines), H_{A_2} (dashed lines), and H_{A_3} (dotted lines) in the three-class case. The horizontal axis labels are as in Fig. 3.

these differences do not influence the trend in the power estimates, e.g., new tests tend to have higher power based on either asymptotic critical values or Monte Carlo critical values under segregation. □

Remark 9.3 (Edge Correction for NNCT-Tests). The CSR independence pattern assumes that the study region is unbounded for the analyzed pattern, which is not the case in practice. So the edge (or boundary) effects might confound the test results in the analysis of empirical (i.e., bounded) data sets if the null pattern is the CSR independence and much effort has gone into the development of edge corrections methods (Yamada and Rogersen, 2003). Two correction methods for the edge effects on NNCT-tests, namely buffer zone correction and toroidal correction, are investigated in Ceyhan (2006, 2007) where it is shown that the empirical sizes of the NNCT-tests are not affected by the toroidal edge correction under CSR independence. However, toroidal correction is biased for non-CSR patterns. In particular if the pattern outside the plot (which is often unknown) is not the same as that inside it it yields questionable results (Haase, 1995; Yamada and Rogersen, 2003). The bias is more severe

Table 3

The NNCT for Pielou's data and the corresponding percentages (left), where the cell percentages are with respect to the row sums (i.e., species sizes) and the marginal percentages are with respect to the total sample size; the overall and cell-specific test statistics for Dixon's tests (middle) and the new tests (right) and the corresponding p -values (in parentheses) based on asymptotic approximation, denoted by p_{asy} , for Pielou's data

		NN			Sum
		D.F.	P.P.		
Base	D.F.	137	23		160
	P.P.	38	30		68
	Sum	175	53		228
		NN			
		D.F.	P.P.		
Base	D.F.	86%	15%		70%
	P.P.	56%	44%		30%
		77%	23%		100%
Dixon's tests					
Overall test					
$C_D = 19.67$ ($p_{asy} = .0001$)					
Cell-specific tests					
		D.F.	P.P.		
D.F.		4.36 (<.0001)	-4.36 (<.0001)		
P.P.		-2.29 (.0221)	2.29 (.0221)		
New tests					
Overall test					
$C_N = 13.11$ ($p_{asy} = .0003$)					
Cell-specific tests					
		D.F.	P.P.		
D.F.		3.63 (.0003)	-3.61 (.0003)		
P.P.		-3.63 (.0003)	3.61 (.0003)		

D.F. = Douglas-firs, P.P. = ponderosa pines; C_D and C_N stand for the value of Dixon's and new overall test statistic, respectively.

especially when there are clusters around the edges. The (outer) buffer zone edge correction method seems to have slightly stronger influence on the tests compared to toroidal correction. But for these tests, buffer zone correction does not change the sizes significantly for most sample size combinations. This is in agreement with the findings of Barot et al. (1999) who say NN methods only require a small buffer area around the study region. A large buffer area does not help much since one only needs to be able to see far enough away from an event to find its NN. Once the buffer area extends past the likely NN distances (i.e., about the average NN distances), it is not adding much helpful information for NNCTs. Hence we recommend inner or outer buffer zone correction for NNCT-tests with the width of the buffer area being about the average NN distance. We do not recommend larger buffer areas, since they are wasteful with little additional gain. For larger distances, the gain might not be worth the effort. Empirical edge corrections are available for other nearest neighbor tests (Sinclair, 1985). They are potentially applicable to the NNCT-tests, however the buffer zone correction analysis in Ceyhan (2007) suggests that such a correction will not make much difference. □

10. Examples

We illustrate the tests on two ecological data sets: Pielou's Douglas-fir/Ponderosa Pine data (Pielou, 1961) and a swamp tree data (Good and Whipple, 1982).

10.1. Pielou's data

Pielou (1961) used a completely mapped data set that is comprised of two tree species: Douglas-fir trees (*Pseudotsuga menziesii* formerly *P. taxifolia*) and Ponderosa Pine (*Pinus ponderosa*) from a region in British Columbia. Her data set was also used by Dixon as an illustrative example (Dixon, 1994). The question of interest is the type of spatial interaction between the two tree species. The corresponding 2 × 2 NNCT and the percentages for each cell are provided in Table 3. The cell percentages are with respect to the sample sizes of each species, for example, 86% of Douglas-firs have NNs from Douglas-firs and remaining 15% of Douglas-firs have NNs from Ponderosa Pines. The row and column percentages are marginal percentages with respect to the total sample size. The percentage values are suggestive of segregation for both species.

The raw data are not available, hence we can not perform Monte Carlo simulation nor randomization versions of the tests. Fortunately, Pielou (1961) provided $Q = 162$ and $R = 134$, hence we can calculate the test statistics and use the asymptotic approximation for these tests. The overall and cell-specific test statistics and the corresponding p -values (in parentheses) based on the asymptotic approximation, denoted by p_{asy} , are also provided in Table 3. Although the locations of the trees are not known, they can be viewed a priori resulting from different processes rather than some process affecting a posteriori the individuals of a single population. So the more appropriate null hypothesis is CSR independence of the trees.

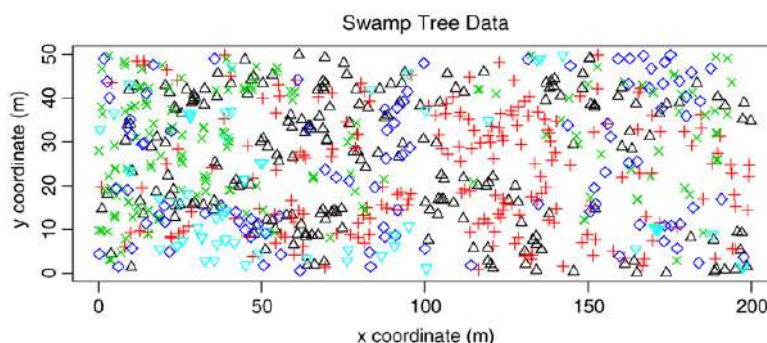


Fig. 8. The scatter plot of the locations of water tupelos (triangles Δ), black gum trees (pluses $+$), Carolina ashes (crosses \times), bald cypress trees (diamonds \diamond), and other trees (inverse triangles ∇).

Table 4

The NNCT for swamp tree data (left) and the corresponding percentages (right), where the cell percentages are with respect to the row sums and marginal percentages are with respect to the total size

		NN					Sum
		W.T.	B.G.	C.A.	B.C.	O.T.	
Base	W.T.	112 (52%)	40 (19%)	29 (13%)	20 (11%)	14 (9%)	215 (29%)
	B.G.	38 (19%)	117 (57%)	26 (13%)	16 (8%)	8 (4%)	205 (28%)
	C.A.	23 (15%)	23 (15%)	82 (53%)	22 (14%)	6 (4%)	156 (21%)
	B.C.	19 (19%)	29 (30%)	29 (30%)	14 (14%)	7 (7%)	98 (13%)
	O.T.	7 (12%)	8 (13%)	5 (8%)	7 (12%)	33 (55%)	60 (8%)
Sum		199 (27%)	217 (30%)	171 (23%)	79 (11%)	68 (9%)	734 (100%)

W.T. = water tupelos, B.G. = black gum, C.A. = Carolina ashes, B.C. = bald cypresses, and O.T. = other tree species.

Hence our inference will be a conditional one (see Remark 5.1). Observe that Dixon's and new overall test statistics yield significant p -values, implying some sort of deviation from CSR independence. In order to see the type of deviation, we apply the cell-specific tests. Both versions of the cell-specific tests for each cell are significant, implying significant deviation from CSR independence. The cell-specific test statistics are positive for the diagonal cells (1, 1) and (2, 2) (and negative for the off-diagonal cells (1, 2) and (2, 1)), implying segregation for both species. This is in agreement with what the NNCT suggests and the findings of Dixon (1994). However, Dixon's cell (1, 1) statistics are much larger than cell (2, 2) statistics, which may be interpreted as clustering of Douglas-firs is stronger than the clustering of Ponderosa Pines. Our simulation study indicates that this might be an artifact of the relative abundance of the tree species. On the other hand, new cell (1, 1) and cell (2, 2) statistics are very similar, hence the segregation of both tree species are at about the same degree.

10.2. Swamp tree data

Good and Whipple (1982) considered the spatial interaction between tree species along the Savannah River, South Carolina, USA. From this data, Dixon (2002) used a single 50 m \times 200 m rectangular plot to illustrate his NNCT-tests. All live or dead trees with 4.5 cm or more dbh (diameter at breast height) were recorded together with their species. The plot contains 13 different tree species, four of which comprises over 90% of the 734 tree stems. The remaining tree stems were categorized as "other trees". The plot consists of 215 water tupelos (*Nyssa aquatica*), 205 black gums (*Nyssa sylvatica*), 156 Carolina ashes (*Fraxinus caroliniana*), 98 bald cypresses (*Taxodium distichum*), and 60 stems from 8 additional species (i.e., other species). A 5 \times 5 NNCT-analysis is conducted for this data set. If segregation among the less frequent species is important, a more detailed 12 \times 12 NNCT-analysis should be performed. The locations of these trees in the study region are plotted in Fig. 8 and the corresponding 5 \times 5 NNCT together with percentages based on row and grand sums are provided in Table 4. Observe that the percentages and Fig. 8 are suggestive of segregation for all tree species, especially for Carolina ashes, water tupelos, black gums, and the "other" trees since the observed percentages of species with themselves as the NN are much larger than the row percentages.

The locations of the tree species can be viewed a priori resulting from different processes, so the more appropriate null hypothesis is the CSR independence pattern. We calculate $Q = 472$ and $R = 454$ for this data set. We present Dixon's overall test of segregation and cell-specific test statistics and the associated p -values in Table 5, where p_{mc} is the p -value based on 10 000 Monte Carlo replication of CSR independence in the same plot and p_{rand} is based on Monte Carlo randomization of the labels on the given locations of the trees 10 000 times. Notice that p_{asy} , p_{mc} , and p_{rand} are very similar for each test. We present the new overall test of segregation and cell-specific test statistics and the associated p -values in Table 6, where p -values are calculated as in Table 5. Again, all three p -values in Table 6 are similar for each test.

Dixon's and the new overall test of segregation are both significant implying significant deviation from the CSR independence pattern for at least one pair of the tree species. Then to determine which pairs exhibit segregation or

Table 5

Test statistics and *p*-values for Dixon's overall and cell-specific tests and the corresponding *p*-values (in parentheses)

		Dixon's overall test				
		$C_D = 275.64$ ($p_{asy} < .0001, p_{mc} < .0001, p_{rand} < .0001$)				
		Dixon's cell-specific tests				
		W.T.	B.G.	C.A.	B.C.	O.T.
W.T.	p_{asy}	6.39 ($< .0001$)	-3.11 (.0019)	-2.87 (.0041)	-1.82 (.0682)	-0.94 (.3484)
	p_{mc}	($< .0001$)	(.0014)	(.0043)	(.0702)	(.3489)
	p_{rand}	($< .0001$)	(.0015)	(.0048)	(.0670)	(.3286)
B.G.	p_{asy}	-3.44 (.0006)	8.05 ($< .0001$)	-3.09 (.0020)	-2.43 (.0150)	-2.34 (.0194)
	p_{mc}	(.0004)	($< .0001$)	(.0014)	(.0179)	(.0192)
	p_{rand}	(.0005)	($< .0001$)	(.0012)	(.0172)	(.0198)
C.A.	p_{asy}	-4.05 (.0001)	-3.73 (.0002)	8.08 ($< .0001$)	0.28 (.7820)	-2.04 (.0410)
	p_{mc}	($< .0001$)	(.0004)	($< .0001$)	(.7810)	(.0430)
	p_{rand}	($< .0001$)	(.0001)	($< .0001$)	(.7580)	(.0409)
B.C.	p_{asy}	-2.18 (.0295)	-0.36 (.7180)	2.04 (.0418)	0.25 (.8011)	-0.38 (.7008)
	p_{mc}	(.0292)	(.7129)	(.0410)	(.7601)	(.6739)
	p_{rand}	(.0293)	(.6861)	(.0419)	(.7910)	(.6555)
O.T.	p_{asy}	-3.02 (.0025)	-2.54 (.0112)	-2.47 (.0135)	-0.39 (.6952)	10.77 ($< .0001$)
	p_{mc}	(.0028)	(.0112)	(.0115)	(.6582)	($< .0001$)
	p_{rand}	(.0021)	(.0121)	(.0158)	(.6490)	($< .0001$)

W.T. = water tupelos, B.G. = black gums, C.A. = Carolina ashes, B.C. = bald cypress and O.T. = other tree species. p_{asy} , p_{mc} , and p_{rand} stand for the *p*-values based on the asymptotic approximation, Monte Carlo simulation, and randomization of the tests, respectively.

Table 6

Test statistics and *p*-values for the new overall and cell-specific tests and the corresponding *p*-values (in parentheses)

		New overall test				
		$C_N = 263.10$ ($p_{asy} < .0001, p_{mc} < .0001, p_{rand} < .0001$)				
		New cell-specific tests				
		W.T.	B.G.	C.A.	B.C.	O.T.
W.T.	p_{asy}	7.55 ($< .0001$)	-4.08 ($< .0001$)	-4.06 (.0001)	-0.74 (.4584)	-1.74 (.0819)
	p_{mc}	($< .0001$)	($< .0001$)	($< .0001$)	(.4564)	(.0860)
	p_{rand}	($< .0001$)	($< .0001$)	($< .0001$)	(.4645)	(.0824)
B.G.	p_{asy}	-3.04 (.0023)	8.16 ($< .0001$)	-4.25 ($< .0001$)	-1.45 (.1479)	-3.27 (.0011)
	p_{mc}	(.0028)	($< .0001$)	($< .0001$)	(.1550)	(.0013)
	p_{rand}	(.0018)	($< .0001$)	($< .0001$)	(.1493)	(.0008)
C.A.	p_{asy}	-3.71 (.0002)	-4.52 ($< .0001$)	7.96 ($< .0001$)	1.36 (.1745)	-2.77 (.0056)
	p_{mc}	(.0001)	(.0001)	($< .0001$)	(.1776)	(.0064)
	p_{rand}	(.0001)	($< .0001$)	($< .0001$)	(.1806)	(.0064)
B.C.	p_{asy}	-1.78 (.0754)	0.00 (.9977)	1.61 (.1081)	0.89 (.3725)	-0.82 (.4097)
	p_{mc}	(.0702)	(.9952)	(.1098)	(.3771)	(.4105)
	p_{rand}	(.0723)	(.9958)	(.1114)	(.3796)	(.4071)
O.T.	p_{asy}	-2.72 (.0066)	-2.90 (.0037)	-2.94 (.0033)	0.21 (.8335)	10.71 ($< .0001$)
	p_{mc}	(.0060)	(.0031)	(.0027)	(.8375)	($< .0001$)
	p_{rand}	(.0070)	(.0036)	(.0026)	(.8354)	($< .0001$)

The labeling of the species and *p*-values are as in Table 5.

association, we perform the cell-specific tests. Dixon's and the new cell-specific tests agree for all cells (i.e., pairs) in term of significance at .05 level except for (B.G., B.C.), (B.C., W.T.), and (B.C., C.A.) pairs. The statistics are all negative for the off-diagonal cells, except for (B.C., C.A.) and (C.A., B.C.) pairs. Based on the Monte Carlo simulation analysis, the new test is more reliable to attach significance to these situations. So using the new tests, the spatial interaction is significant between each pair which does not contain bald cypresses. That is, the new cell-specific test statistics are positive for the diagonal cells (i, i) for $i = 1, 2, \dots, 5$ and are significant for $i = 1, 2, 3, 5$ at .05 level (which also holds for Dixon's tests); and are negative

Table 7
The empirical rejection rates for the NNCT-tests under various patterns different from CSR independence and RL
Rejection rates of the NNCT-tests under various PCP and inhomogeneous patterns

Sizes (n_1, n_2)	Case 1: $\sigma = .025$			Case 2: $\sigma = .05$			Case 3: $\sigma = .10$			
	Dixon's $\hat{\alpha}_{1,1}^D$	New $\hat{\alpha}_{1,1}^N$	Overall $\hat{\alpha}_D$	Dixon's $\hat{\alpha}_{1,1}^D$	New $\hat{\alpha}_{1,1}^N$	Overall $\hat{\alpha}_D$	Dixon's $\hat{\alpha}_{1,1}^D$	New $\hat{\alpha}_{1,1}^N$	Overall $\hat{\alpha}_D$	
PCP1(n_p, n_1, n_2, σ) with $n_p = 5, n_c = n_1/n_p$ for sample X and $n_c = n_2/n_p$ for sample Y (same parent set for X and Y)										
(30, 30)	.0814	.0716	.0709	.0735	.0701	.0717	.0719	.0656	.0619	.0631
(30, 50)	.0702	.0728	.0571	.0637	.0684	.0600	.0627	.0634	.0537	.0543
(50, 50)	.0624	.0614	.0615	.0691	.0602	.0677	.0556	.0541	.0598	.0600
(different parent sets for X and Y)										
(30, 30)	.9993	1.000	1.000	1.000	.9880	.9958	.9957	.7641	.8527	.8522
(30, 50)	1.000	.9997	1.000	1.000	.9967	.9985	.9985	.8702	.9058	.9064
(50, 50)	.9999	.9999	1.000	1.000	.9987	.9996	.9996	.8984	.9525	.9532
PCP2(n_p, n_1, n_2, σ) with $n_p = 5, n_c = n_1/n_p$ for sample X and $n_c = n_2/n_p$ for sample Y (same parent set for X and Y)										
(30, 30)	.0509	.0498	.0475	.0470	.0523	.0470	.0512	.0485	.0428	.0440
(30, 50)	.0587	.0545	.0483	.0499	.0522	.0494	.0609	.0573	.0502	.0510
(50, 50)	.0479	.0499	.0536	.0537	.0482	.0526	.0476	.0466	.0516	.0529
(different parent sets for X and Y)										
(30, 30)	.9993	.9998	.9999	1.000	.9884	.9953	.9953	.7779	.8624	.8627
(30, 50)	1.000	.9999	1.000	1.000	.9976	.9991	.9991	.8854	.9169	.9187
(50, 50)	1.000	1.000	1.000	1.000	.9987	.9991	.9997	.9182	.9607	.9602
Case 1: $r = .05$										
MCP(κ, n_1, n_2, r, μ) with $\kappa = 5$ for both X and Y samples (same parent set for X and Y)										
(50, 50)	.0494	.0487	.0483	.0487	.0530	.0580	.0514	.0487	.0527	.0524
(50, 100)	.0495	.0463	.0471	.0473	.0452	.0485	.0502	.0516	.0489	.0497
(100, 100)	.0458	.0508	.0499	.0501	.0527	.0534	.0507	.0496	.0500	.0504
(different parent sets for X and Y)										
(50, 50)	.9983	.9985	.9997	.9997	.9887	.9954	.9954	.8019	.8873	.8881
(50, 100)	.9992	.9981	.9994	.9994	.9963	.9979	.9979	.9087	.9674	.9674
(100, 100)	.9992	.9994	.9998	.9998	.9988	.9998	.9998	.9581	.9804	.9805
Case 2: $f_1(x, y) = \sqrt{x+y}, f_2(x, y) = \sqrt{xy}$										
Case 2: $f_1(x, y) = \sqrt{x+y}, f_2(x, y) = \sqrt{xy}$										
Case 3: $f_1(x, y) = \sqrt{(x+y)}, f_2(x, y) = x-y $										
IPCP($\lambda_1(x, y) = n_1 f_1(x, y), \lambda_2(x, y) = n_2 f_2(x, y)$) with $i = 1, 2$ for X and Y points, respectively										
(50, 50)	.0541	.0495	.0507	.0511	.0487	.0518	.0534	.0518	.0527	.0534
(50, 100)	.0471	.0465	.0458	.0440	.0515	.0540	.0561	.0524	.0563	.0564
(100, 100)	.0469	.0516	.0491	.0490	.0513	.0572	.0591	.0592	.0592	.0592

PCP: Poisson Cluster Process, MCP: Matern Cluster Process, IPCP: Inhomogeneous Poisson Cluster Process. See Section 7 for details on these point processes.

for the off-diagonal cells (i, j) with $i, j \in \{1, 2, 3, 5\}$ and $i \neq j$ and significant for most of them. Hence each tree species except bald cypresses exhibits significant segregation from each other. These findings are mostly in agreement with the results of Dixon (2002). Hence except for bald cypresses, each tree species seem to result from a (perhaps) different first order inhomogeneous Poisson process.

The results based on NNCT-tests pertain to small scale interaction at about the average NN distances. We might also be interested in the causes of the segregation and the type and level of interaction between the tree species at different scales (i.e., distances between the trees). To answer such questions, we also present the second-order analysis of the swamp tree data (Diggle, 2003) using the functions (or some modified version of them) provided in spatstat package in R (Baddeley and Turner, 2005). We use Ripley's univariate and bivariate L -functions which are modified versions of his K -functions. The estimator $\widehat{K}(t)$ of $K(t)$ is approximately unbiased for $K(t)$ at each fixed t . Bias depends on the geometry of the study area and increases with t . For a rectangular region it is recommended to use t values up to $1/4$ of the smaller side length of the rectangle. So we take the values $t \in [0, 12.5]$ in our analysis, since the rectangular region is 50×200 m. But Ripley's K -function is cumulative, so interpreting the spatial interaction at larger distances is problematic (Wiegand et al., 2007). The (accumulative) pair correlation function $g(t)$ is better for this purpose (Stoyan and Stoyan, 1994). The pair correlation function of a (univariate) stationary point process is defined as

$$g(t) = \frac{K'(t)}{2\pi t}$$

where $K'(t)$ is the derivative of $K(t)$. Ripley's univariate L -functions and the pair correlation functions for all trees combined and each species for the swamp tree data are presented in Ceyhan (2008b) where it is observed that each species exhibit spatial aggregation at different scales (i.e., at different ranges of the interpoint distances). Hence, segregation of the species might be due to different levels and types of aggregation of the species in the study region.

We also calculate Ripley's bivariate L -function $L_{ij}(t)$. By construction, $L_{ij}(t)$ is symmetric in i and j in theory, that is, $L_{ij}(t) = L_{ji}(t)$ for all $i \neq j$. But in practice edge corrections will render it slightly asymmetric, i.e., $\widehat{L}_{ij}(t) \neq \widehat{L}_{ji}(t)$. But the corresponding estimates are pretty close in our example, so only one bivariate plot per pair is presented. Under CSR independence, we have $L_{ij}(t) - t = 0$. If the bivariate pattern is segregation, then $L_{ij}(t) - t$ tends to be negative; if it is association then $L_{ij}(t) - t$ tends to be positive. See Diggle (2003) for more detail. In Fig. 9, we present the bivariate plots of $\widehat{L}_{ij}(t) - t$ functions together with the upper and lower (pointwise) 95% confidence bounds for each pair of species (due to the symmetry of $\widehat{L}_{ij}(t)$ there are only 10 different pairs). The same definition of the pair correlation function can be applied to Ripley's bivariate K or L -functions as well. The benchmark value of $K_{ij}(t) = \pi t^2$ corresponds to $g(t) = 1$; $g(t) < 1$ suggests segregation of the classes; and $g(t) > 1$ suggests association of the classes. The bivariate pair correlation functions for the species in swamp tree data are plotted in Fig. 10. However the pair correlation function estimates might have critical behavior for small t if $g(t) > 0$ since the estimator variance and hence the bias are considerably large. This problem gets worse especially in cluster processes (Stoyan and Stoyan, 1996). See for example Fig. 10 where the confidence bands for smaller t values are much wider compared to those for larger t values. So pair correlation function analysis is more reliable for larger distances and it is safer to use $g(t)$ for distances larger than the average NN distance in the data set. Comparing Figs. 9 and 10 we see that Ripley's L and pair correlation functions usually detect the same large-scale pattern but at different ranges of distance values. Ripley's L suggests that the particular pattern is significant for a wider range of distance values compared to $g(t)$, since values of L at small scales confound the values of L at larger scales where $g(t)$ is more reliable to use (Wiegand and Moloney, 2004; Loosmore and Ford, 2006).

So, in Fig. 9, we only consider distances up to the average NN distance, which is about $t = 2$ m. Water tupelos and black gums exhibit significant segregation ($\widehat{L}_{12}(t) - t$ is below the lower confidence bound); water tupelos and Carolina ashes are significantly segregated; water tupelos and bald cypresses do not have significant deviation from the CSR independence pattern; water tupelos and the other trees do not deviate significantly from CSR independence. Black gums and Carolina ashes do not deviate significantly from CSR independence; neither do the black gums and bald cypresses; black gum and other trees are significantly segregated. Carolina ashes and bald cypresses do not deviate significantly from CSR independence; and neither do Carolina ashes and the other trees. Likewise bald cypresses and other trees do not deviate significantly from CSR independence.

In Fig. 10 we consider distances larger than 2 m, since pair correlation function is unreliable for smaller distances. Observe that water tupelos and black gums do not deviate from CSR independence; water tupelos and Carolina ashes are segregated for about 2.5 m and are associated for values about 6 m; water tupelos and bald cypresses are segregated for 5.5, 9.5, and 11 m and are associated for 6.5 m; water tupelos and other trees are segregated for about 7 m and are associated for 8 m; black gums and Carolina ashes are segregated for 2–2.5, 3.5–4.5, 6–8.5, and 9.5–12 m; black gums and bald cypresses are segregated for 3.5, 5.5–6.5, 7, and 9.5 m; black gums and other trees are segregated for 5 and 6–7.5 m; Carolina ashes and bald cypresses are associated for 2–3, 5.5, and 7 m; Carolina ashes and other trees are associated for 5 and 9–10 m; and bald cypresses and other trees are segregated for 4 m and are associated for 3–4 and 6.5–7.5 m.

While second order analysis (using Ripley's K and L -functions or pair correlation function) provides information on the univariate and bivariate patterns at all scales (i.e., for all distances), NNCT-tests summarize the spatial interaction for the smaller scales (for distances about the average NN distance between the points). In particular, for the swamp tree data average NN distance (\pm standard deviation) is about 1.8 (± 1.04) meters and notice that Ripley's L -function and NNCT-tests usually yield similar results for distances about 2 m.

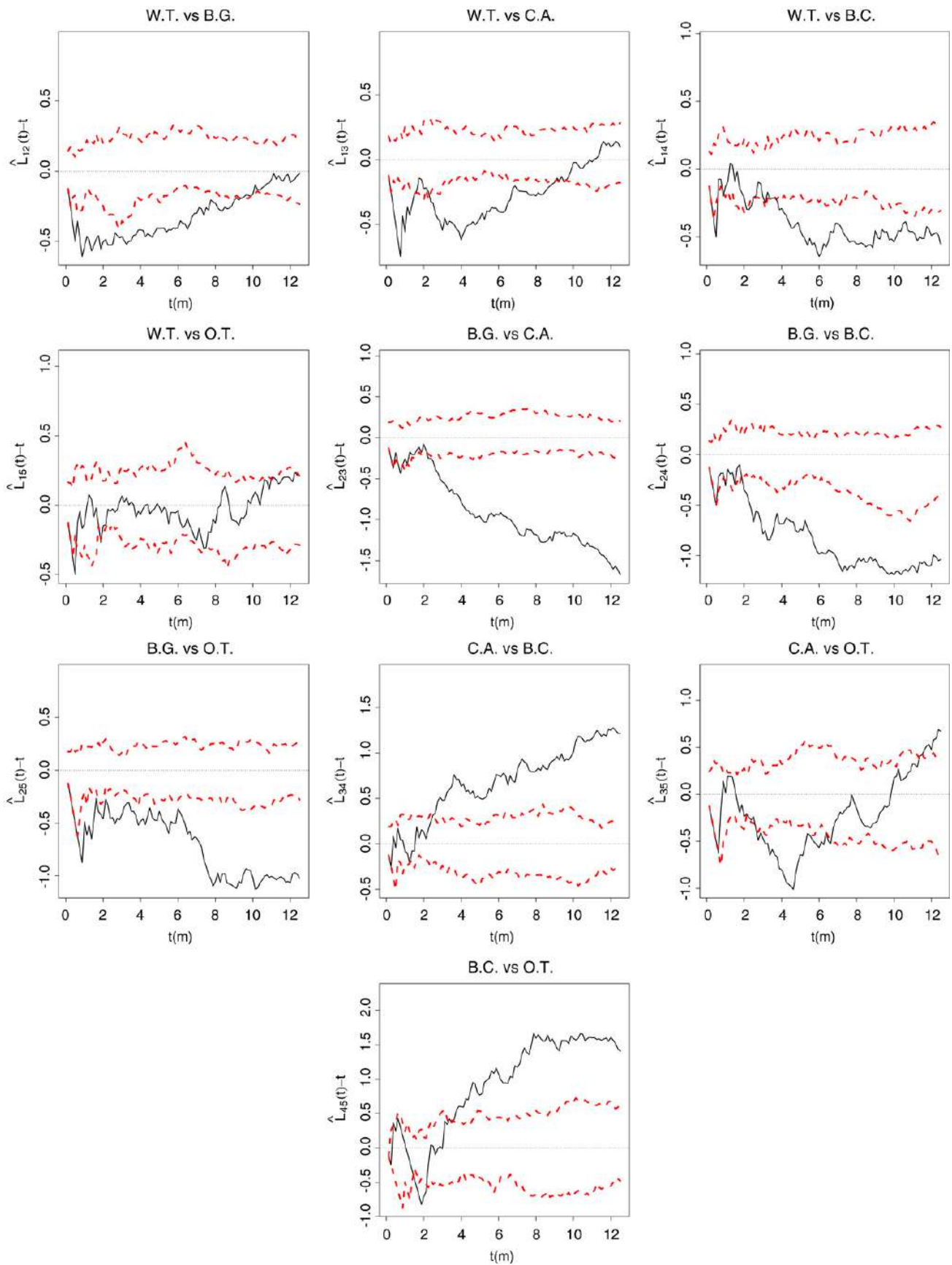


Fig. 9. Second-order analysis of swamp tree data by Ripley's bivariate L -functions $\hat{L}_{ij}(t) - t$ for $i, j = 1, 2, \dots, 5$ and $i \neq j$ where $i = 1$ stands for water tupelos, $i = 2$ for black gums, $i = 3$ for Carolina ashes, $i = 4$ for bald cypresses, and $i = 5$ for other trees. Wide dashed lines around 0 (which is the theoretical value) are the upper and lower (pointwise) 95% confidence bounds for the L -functions based on Monte Carlo simulations under the CSR independence pattern. W.T. = water tupelos, B.G. = black gums, C.A. = Carolina ashes, B.C. = bald cypresses, and O.T. = other tree species.

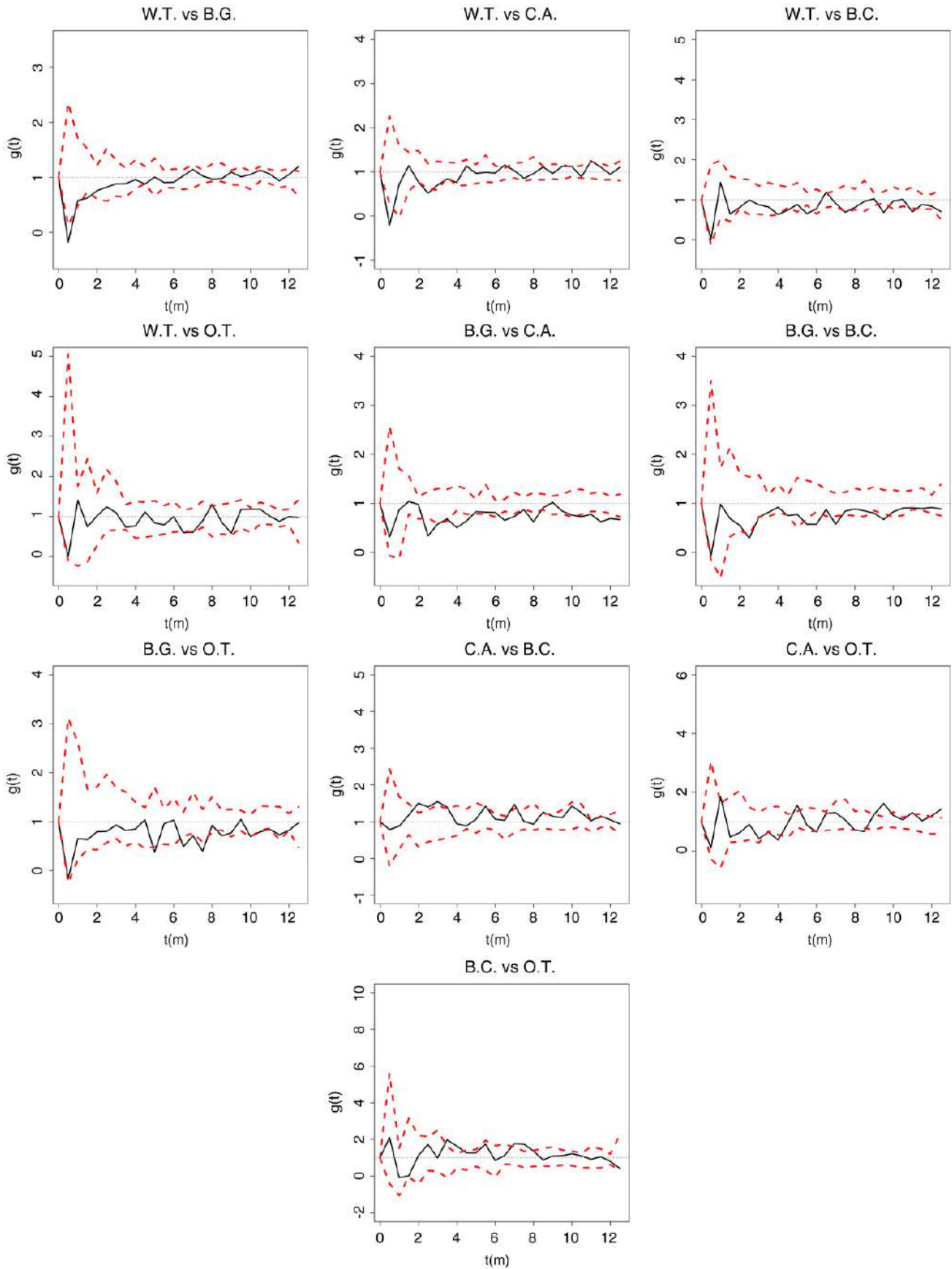


Fig. 10. Pair correlation functions for each pair of species in the swamp tree data. Wide dashed lines around 1 (which is the theoretical value) are the upper and lower (pointwise) 95% confidence bounds for the L -functions based on Monte Carlo simulations under the CSR independence pattern. W.T. = water tupelos, B.G. = black gums, C.A. = Carolina ashes, B.C. = bald cypresses, and O.T. = other tree species.

11. Discussion and conclusions

In this article we introduce new overall and cell-specific tests of segregation based on nearest neighbor contingency tables (NNCTs) and compare them with Dixon's NNCT-tests and exploratory methods such as Ripley's L -function and pair correlation function. We demonstrate that NNCT-tests are conditional on Q and R under CSR independence, but unconditional under RL. Based on our Monte Carlo simulations, we conclude that the asymptotic approximation for the cell-specific-tests is appropriate only when the corresponding cell count in the NNCT is larger than 10; and for the overall test when all cell counts are larger than 4. For smaller samples, Monte Carlo randomization versions of the tests are recommended. For larger samples the new versions of the segregation tests have empirical significance levels closer to the nominal level. Further, Type I error rates (empirical significance levels) of the new cell-specific tests are more robust to the differences in relative abundance. Based on the size and power estimates, we recommend the new cell-specific and overall tests for the segregation alternatives. For the association alternatives, we recommend both versions (i.e., Dixon's and the new versions) of the overall and cell-specific tests.

NNCT-tests summarize the pattern in the data set for small scales around the average NN distance between all points. On the other hand, pair correlation function $g(t)$ and Ripley's classical K or L -functions and other variants provide information on the pattern at various scales. Ripley's classical K or L -functions can be used when the null pattern can be assumed to be CSR independence; that is, when the null pattern assumes first-order homogeneity for each class. When the null pattern is the RL of points from an inhomogeneous Poisson process they are not appropriate (Kulldorff, 2006); Cuzick–Edward's k -NN tests are designed for testing bivariate spatial interaction and mostly used for spatial clustering of cases in epidemiology; Diggle's D -function is a modified version of Ripley's K -function (Diggle, 2003) and adjusts for any inhomogeneity in the locations of, e.g., cases and controls. Furthermore, there are variants of K -function that explicitly correct for inhomogeneity (see Baddeley et al. (2000)). Our example illustrates that for distances around the average NN distance, NNCT-tests and Ripley's bivariate L -function yield similar results.

The NNCT-tests and Ripley's L -function provide similar information in the two-class case at small scales. For q -class case with $q > 2$ classes, overall tests provide information on the (small-scale) multivariate spatial interaction in one compound summary measure; while the Ripley's L -function requires performing all bivariate spatial interaction analysis. The cell-specific tests can serve as pairwise post hoc analysis only when the overall test is significant. Furthermore, the cell-specific tests are testing the spatial clustering of one class or bivariate interaction between two classes as part of the multivariate interaction between all the classes. On the other hand, Ripley's univariate K - or L -functions are restricted to one class and bivariate K - or L -functions are restricted to two classes they pertain to, ignoring the potentially important multivariate interaction between all classes in the study area. However, there are forms of the J -function which is derived from the well-known G and F functions (van Lieshout and Baddeley, 1999) and deal with this multi-type setting (i.e., consider the pattern of type i in the context of the pattern of all other types). van Lieshout and Baddeley (1999) define two basic types of J -functions. First is a type- i -to-type- j function which considers the points of type i in the context of the points of type j . The second one is the type- i -to-any-type function which considers the points of type i in the context of points of all types including type i . Other forms can be derived from them by re-defining the types. For example, if we want to consider the points of type i in the context of points of all other types, then we collapse all the other types j (i.e., all j which are not equal to i) into a single type i' and then use the type- i -to-type- i' function. Several authors have written about the bivariate K -function, which is of the type- i -to-type- j form (Diggle and Chetwynd, 1991; Haase, 1995; Diggle, 2003). Type- i -to-type- j K -function can easily be modified to type- i -to-any-type K -function. Thus, essentially there is only one family of multi-type K -functions in literature. But type- i -to-type- j K -function is comparable with a NNCT analysis based on a 2×2 NNCT restricted to the classes i and j . Similarly, type- i -to-type- i' K -function is comparable with the NNCT analysis based on a 2×2 NNCT with classes i and the rest of the classes labeled as i' . Since pairwise analysis of q classes with 2×2 NNCTs might yield conflicting results compared to $q \times q$ NNCT analysis (Dixon, 2002), Ripley's L -function and NNCT-tests might also yield conflicting results at small distances. Hence Ripley's L -function and NNCT-tests may provide similar but not identical information about the spatial pattern and the latter might detect small-scale interaction that is missed by the former. Since the pair correlation functions are derivatives of Ripley's K -function, most of the above discussion holds for them also, except $g(t)$ is reliable only for large scale interaction analysis. Hence NNCT-tests and pair correlation function are not comparable but provide complimentary information about the pattern in question.

Ripley's univariate and bivariate K - or L -functions (and hence the corresponding pair correlation functions) are symmetric (in theory) in the classes they pertain to. Cell-specific tests for diagonal cells in a NNCT are also symmetric, but cell-specific tests for two different classes (i.e., off-diagonal cells in the NNCT) are not symmetric. Hence, at small scales, the cell-specific test for an off-diagonal cell, provides the type and different levels of spatial interaction for the corresponding two classes, while Ripley's L -function and pair correlation function provide only the type of spatial interaction, but can not distinguish the class-specific level of interaction for each of the two classes in question.

For a data set for which CSR independence is the reasonable null pattern, we recommend the overall segregation test if the question of interest is the spatial interaction at small scales (i.e., about the mean NN distance). If it yields a significant result, then to determine which pairs of classes have significant spatial interaction, the cell-specific tests can be performed. One can also perform Ripley's K or L -function and only consider distances up to around the average NN distance and compare the results with those of NNCT analysis. If the spatial interaction at higher scales is of interest, pair correlation function is recommended (Loosmore and Ford, 2006), due to the cumulative nature of Ripley's K - or L -functions for larger distances. On

the other hand, if the RL pattern is the reasonable null pattern for the data, we recommend the NNCT-tests if the small-scale interaction is of interest and Diggle's D -function if the spatial interaction at higher scales is also of interest.

Acknowledgments

I would like to thank an anonymous associate editor and two referees, whose constructive comments and suggestions greatly improved the presentation and flow of the paper. Most of the Monte Carlo simulations presented in this article were executed on the Hattusas cluster of Koç University High Performance Computing Laboratory.

References

- Armstrong, J.E., Irvine, A.K., 1989. Flowering, sex ratios, pollen-ovule ratios, fruit set, and reproductive effort of a dioecious tree, *Myristica insipida* (*Myristicaceae*), in two different rain forest communities. *American Journal of Botany* 76, 75–85.
- Baddeley, A., Møller, J., Waagepetersen, R., 2000. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* 54 (3), 329–350.
- Baddeley, A., Turner, R., 2005. spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software* 12 (6), 1–42.
- Barot, S., Gignoux, J., Menaut, J.C., 1999. Demography of a savanna palm tree: predictions from comprehensive spatial pattern analyses. *Ecology* 80, 1987–2005.
- Ceyhan, E., 2006. On the use of nearest neighbor contingency tables for testing spatial segregation. *Environmental and Ecological Statistics*. Also available as Technical Report # KU-EC-08-4, Koç University, Istanbul, Turkey or online as [arXiv:0807.4236](https://arxiv.org/abs/0807.4236) [stat.ME] (in press).
- Ceyhan, E., 2007. Edge correction for cell- and class-specific tests of segregation based on nearest neighbor contingency tables. In: *Proceedings of the International Conference on Environment: Survival and Sustainability*, Near East University.
- Ceyhan, E., 2008a. New tests of spatial segregation based on nearest neighbor contingency tables. Under Review. (Available as Technical Report No. KU-EC-08-6, Koç University, Istanbul, Turkey.)
- Ceyhan, E., 2008b. Overall and pairwise segregation tests based on nearest neighbor contingency tables. Technical Report # KU-EC-08-1, Koç University, Istanbul, Turkey. Also available online as [arXiv:0805.1629v2](https://arxiv.org/abs/0805.1629v2) [stat.ME].
- Ceyhan, E., 2008c. QR-adjustment for clustering tests based on nearest neighbor contingency tables. Also available as Technical Report # KU-EC-08-5, Koç University, Istanbul, Turkey or online as [arXiv:0807.4231v1](https://arxiv.org/abs/0807.4231v1) [stat.ME] (submitted for publication).
- Coomes, D.A., Rees, M., Turnbull, L., 1999. Identifying aggregation and association in fully mapped spatial data. *Ecology* 80 (2), 554–565.
- Cressie, N.A.C., 1993. *Statistics for Spatial Data*. Wiley, New York.
- Cuzick, J., Edwards, R., 1990. Spatial clustering for inhomogeneous populations (with discussion). *Journal of the Royal Statistical Society, Series B* 52, 73–104.
- Diggle, P.J., 2003. *Statistical Analysis of Spatial Point Patterns*. Hodder Arnold Publishers, London.
- Diggle, P.J., Chetwynd, A.G., 1991. Second-order analysis of spatial clustering for inhomogeneous populations. *Biometrics* 47, 1155–1163.
- Dixon, P.M., 1994. Testing spatial segregation using a nearest-neighbor contingency table. *Ecology* 75 (7), 1940–1948.
- Dixon, P.M., 2002. Nearest-neighbor contingency table analysis of spatial segregation for several species. *Ecoscience* 9 (2), 142–151.
- Good, B.J., Whipple, S.A., 1982. Tree spatial patterns: South Carolina bottomland and swamp forests. *Bulletin of the Torrey Botanical Club* 109, 529–536.
- Goreaud, F., Pélissier, R., 2003. Avoiding misinterpretation of biotic interactions with the intertype K_{12} -function: Population independence vs. random labelling hypotheses. *Journal of Vegetation Science* 14 (5), 681–692.
- Haase, P., 1995. Spatial pattern analysis in ecology based on Ripley's K -function: Introduction and methods of edge correction. *The Journal of Vegetation Science* 6, 575–582.
- Hamill, D.M., Wright, S.J., 1986. Testing the dispersion of juveniles relative to adults: A new analytical method. *Ecology* 67 (2), 952–957.
- Herrera, C.M., 1988. Plant size, spacing patterns, and host-plant selection in *Osyris quadripartita*, a hemiparasitic dioecious shrub. *Journal of Ecology* 76, 995–1006.
- Kulldorff, M., 2006. Tests for spatial randomness adjusted for an inhomogeneity: A general framework. *Journal of the American Statistical Association* 101 (475), 1289–1305.
- Loosmore, N., Ford, E., 2006. Statistical inference using the g or k point pattern spatial statistics. *Ecology* 87, 1925–1931.
- Mat'ern, B., 1986. Spatial variation. In: Brillinger, D., et al. (Eds.), *Lecture Notes in Statistics*, vol. 36. Springer-Verlag, Berlin.
- Meagher, T.R., Burdick, D.S., 1980. The use of nearest neighbor frequency analysis in studies of association. *Ecology* 61 (5), 1253–1255.
- Moran, P.A.P., 1948. The interpretation of statistical maps. *Journal of the Royal Statistical Society, Series B* 10, 243–251.
- Nanami, S.H., Kawaguchi, H., Yamakura, T., 1999. Dioecy-induced spatial patterns of two codominant tree species, *Podocarpus nagi* and *Neolitsea aciculata*. *Journal of Ecology* 87 (4), 678–687.
- Pielou, E.C., 1961. Segregation and symmetry in two-species populations as studied by nearest-neighbor relationships. *Journal of Ecology* 49 (2), 255–269.
- Ripley, B.D., 2004. *Spatial Statistics*. Wiley-Interscience, New York.
- Searle, S.R., 2006. *Matrix Algebra Useful for Statistics*. Wiley-Intersciences.
- Sinclair, D.F., 1985. On tests of spatial randomness using mean nearest neighbor distance. *Ecology* 66 (3), 1084–1085.
- Stoyan, D., Stoyan, H., 1994. *Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics*. John Wiley and Sons, New York.
- Stoyan, D., Stoyan, H., 1996. Estimating pair correlation functions of planar cluster processes. *Biometrical Journal* 38 (3), 259–271.
- van Lieshout, M.N.M., Baddeley, A.J., 1999. Indices of dependence between types in multivariate point patterns. *Scandinavian Journal of Statistics* 26, 511–532.
- Waagepetersen, R.P., 2007. An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics* 63 (1), 252–258.
- Waller, L.A., Gotway, C.A., 2004. *Applied Spatial Statistics for Public Health Data*. Wiley-Interscience, NJ.
- Whipple, S.A., 1980. Population dispersion patterns of trees in a Southern Louisiana hardwood forest. *Bulletin of the Torrey Botanical Club* 107, 71–76.
- Wiegand, T., Gunatilleke, S., Gunatilleke, N., 2007. Species associations in a heterogeneous Sri Lankan dipterocarp forest. *The American Naturalist* 170 (4), 77–95.
- Wiegand, T., Moloney, K.A., 2004. Rings, circles and null-models for point pattern analysis in ecology. *Oikos* 104, 209–229.
- Yamada, I., Rogerson, P.A., 2003. An empirical comparison of edge effect correction methods applied to K -function analysis. *Geographical Analysis* 35 (2), 97–109.