Cell-specific and post-hoc spatial clustering tests based on nearest neighbor contingency tables

CrossMark

Elvan Ceyhan

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

ARTICLE INFO

ABSTRACT

Spatial clustering patterns in a multi-class setting such as segregation and association between classes have important implications in various fields, e.g., in ecology, and can be tested using nearest neighbor contingency tables (NNCTs). A NNCT is constructed based on the types of the nearest neighbor (NN) pairs and their frequencies. We survey the cell-specific (or pairwise) and overall segregation tests based on NNCTs in literature and introduce new ones and determine their asymptotic distributions. We demonstrate that cell-specific tests enjoy asymptotic normality, while overall tests have chi-square distributions asymptotically. Some of the overall tests are confounded by the unstable generalized inverse of the rank-deficient covariance matrix. To overcome this problem, we propose rank-based corrections for the overall tests to stabilize their behavior. We also perform an extensive Monte Carlo simulation study to compare the finite sample performance of the tests in terms of empirical size and power based on the asymptotic and Monte Carlo critical values and determine the tests that have the best size and power performance and are robust to differences in relative abundances (of the classes). In addition to the cell-specific tests, we discuss one(-class)-versus-rest type of tests as post-hoc tests after a significant overall test. We also introduce the concepts of total, strong, and partial segregation/association to differentiate different levels of these patterns. We compare the new tests with the existing NNCT-tests in literature with simulations and illustrate the tests on an ecological data set.

© 2016 The Korean Statistical Society. Published by Elsevier B.V. All rights reserved.

1. Introduction

Spatial clustering of two or more classes with respect to each other can have important consequences in various fields. Multi-class clustering patterns such as segregation and association might result from the interaction between two or more classes (or species). For convenience, categories of the points or units are referred to as "classes", e.g., a class can stand for species or some other characteristic of the unit or subject. *Segregation* is the spatial pattern in which points from the same class are closer to each other, while *association* is the pattern in which points from different classes are closer to each other. These patterns may have important implications in ecology, plant biology, or epidemiology (see, e.g., Diggle, 2003, Hamill & Wright, 1986, and Whipple, 1980). In particular, in ecology, two tree species could be highly dependent on each other (as a result of, say, symbiosis or mutualism), and thus, they coexist in a close vicinity (i.e., they are *associated* with each other), or they could be enjoying the company of conspecifics and thus form one-class clumps or groups (i.e., they are *segregated* from each other). In epidemiology, cases might be clustered compared to controls (i.e., cases and controls are segregated), due to

infectious nature of a disease or proximity of subjects to a disease source such as a source of a pollutant. These patterns can also be of interest in social studies. For example, segregation of residences due to the socioeconomic status or ethnicity can be investigated by generative models (Fossett, 2011). In a social network, segregation of individuals is also modeled via tools from random graph theory (Henry, Pralat, & Zhang, 2011). In veterinary epidemiology, a nonparametric method for detecting spatial segregation according to the genotype and year of occurrence of bovine tuberculosis is employed by Diggle, Zheng, and Durr (2005). Spatial segregation is also used in literature to refer to a one-class pattern of spatial clustering (Robertson & Cushing, 2011), which is also called as aggregation in literature (Coomes, Rees, & Turnbull, 1999).

Many univariate (i.e., one-class) or multivariate (i.e., multi-class) spatial clustering tests have been proposed in literature (see Kulldorff, 2006 for an extensive review). These methods include K-function (Ripley, 2004), or J-function (van Lieshout & Baddeley, 1999), nearest neighbor (NN) methods (Diggle, 2003) and so on. Among NN methods, this article pertains to the ones based on nearest neighbor contingency tables (NNCTs). Pielou (1961) introduced various tests based on NNCTs; Dixon (1994) extended these tests in various directions, and also determined the correct asymptotic distribution of the proposed tests in the two-class setting. Ceyhan (2008, 2009, 2010a) compared NNCT-tests in literature, and also proposed various tests based on NNCTs. All these NNCT-tests are appropriate for completely mapped data.

In this article, we survey these NNCT-tests and also introduce new cell-specific segregation tests (i.e., tests for each cell or entry in the NNCT) and the corresponding overall tests for completely mapped data. Since the generalized inverse is employed on the covariance matrices to obtain the overall tests and these covariance matrices are rank-deficient, the generalized inverse may be highly unstable (in the sense that the generalized inverse may produce a matrix which may not be positive definite). To overcome this problem, we propose a rank-based correction for the computation of the overall tests (which stabilizes and makes them have the correct asymptotic distribution). We demonstrate that cell-specific tests are asymptotically normal, and overall tests tend to chi-square distributions with respective degrees of freedom as class sizes tend to infinity. In practice, cell-specific tests can serve as post-hoc tests to be performed when an overall segregation test yields a significant result. In Ceyhan (2008), cell-specific tests are compared for two types of NNCT-tests. As an alternative post-hoc test after a significant overall test, we discuss one-class-versus-rest (or one-vs-rest) type of NNCT-tests. By extensive Monte Carlo simulations, we compare these tests with the existing NNCT-tests in literature (i.e., with the ones proposed in Ceyhan, 2008, 2010a and Dixon, 1994, 2002a) in terms of empirical size and power based on asymptotic and Monte Carlo critical values, and thus determine which tests perform better for the segregation or association alternative and which ones are more robust to differences in relative abundances (of the classes). There are also methods to test spatial correlation of lattice data, such as Moran's I (Moran, 1950), Ord's statistics (Ord, 1975) and recently developed approximate profile likelihood estimator (Li, Calder, & Cressie, 2011). However, these statistics have been introduced to identify spatial clusters and spatial interactions among areas. But our cell-specific and overall NNCT-tests, though detecting/testing spatial interaction, are not based on area or aggregated data. The term "cell" in our cell-specific test does not refer to areas or quadrats, but to the entries in the NNCT.

We describe the NNCTs and the null and alternative patterns in Section 2, provide the cell-specific tests and the corresponding overall tests in Section 3, empirical size analysis in the two- and three-class cases in Sections 4 and 5, respectively, and empirical power analysis under segregation and association in the two-class and three-class cases in Sections 6 and 7, respectively. We present the empirical size and power analysis for the one-vs-rest type testing in the three-class case in Section 8, the cell-specific and one-vs-rest type tests as post-hoc tests in Section 9, the illustration of the methodology on the example data set in Section 10, and our conclusions and guidelines for using the tests in Section 11.

2. NNCTs and null and alternative spatial patterns

NNCTs are constructed using the NN frequencies of classes. In this article, to find the NN of a point, we employ the usual Euclidean distance. The construction of NNCTs for two classes is described, e.g., in Ceyhan (2010a); here we provide a brief description for $m \geq 2$ classes. Suppose there are m classes labeled as $\{1, 2, \ldots, m\}$. NNCTs are constructed using NN frequencies for each class. Let N_i be the number of points from class i for $i \in \{1, 2, \ldots, m\}$ and $n = \sum_{i=1}^{m} N_i$ and n points be denoted as $\{w_1, w_2, \ldots, w_n\}$. If we record the class of each point and its NN, the NN relationships fall into m^2 categories:

$$(1, 1), \ (1, 2), \ldots, (1, m); \ (2, 1), \ (2, 2), \ldots, (2, m); \ldots, (m, m)$$

where in category or cell (i, j), class i is called the *base class* and class j is called the *NN class*. Denoting N_{ij} as the observed frequency of category (i, j) for $i, j \in \{1, 2, \ldots, m\}$, we obtain the NNCT, $\mathcal{N}$, in Table 1 where C_j is the sum of column j; i.e., number of times class j points serve as NNs for $j \in \{1, 2, \ldots, m\}$. Note also that $n = \sum_{i,j} N_{ij}$, $n_i = \sum_{j=1}^{m} N_{ij}$, $C_j = \sum_{i=1}^{m} N_{ij}$, and

$$N_{ij} = \sum_{l=1}^{n} \sum_{k=1}^{n} \mathbf{I}(w_k \text{ is NN of } w_l) \cdot \mathbf{I}(w_l \text{ is from class } i) \cdot \mathbf{I}(w_k \text{ is from class } j),$$

where $w_k \neq w_l$ and $\mathbf{I}(\cdot)$ denotes the indicator function. Here we adopt the convention that variables denoted by upper case letters are random, while variables denoted by lower case letters are fixed. Thus, in our NNCT-analysis, row sums are assumed to be fixed (i.e., class sizes are given), while column sums are assumed to be random and depend on the NN relationships between the classes.

Table 1
The NNCT, $\mathcal{N}$, for m classes.

		NN class			Total
		class 1	...	class m	
Base class	class 1	N_{11}	...	N_{1m}	n_1
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	class m	N_{m1}	...	N_{mm}	n_m
	Total	C_1	...	C_m	n

We describe the spatial point patterns for two classes only; the extension to multi-class case is straightforward. Our null hypothesis is

H_o : randomness in the NN structure with NN probabilities being proportional to class frequencies

which may result from *random labeling* (RL) or (an appropriate type of) *independence* of points from two classes. Under independence, the two classes result independently from the same stochastic process which ensures that their spatial distribution is identical. In this article, among independence patterns we will consider complete spatial randomness (CSR) of points from two classes. Roughly, under *CSR independence*, two classes are independently uniformly distributed in a region of interest, while RL is the pattern in which, given a fixed set of points in a region, class labels are assigned to these fixed points randomly so that the labels are independent of the locations. For CSR independence, we condition on $N_i = n_i$; i.e., we work with a binomial process. If it is desired to have the class sizes N_i to be random, we may consider a spatial Poisson point process on the region of interest as our null hypothesis. The null model in a NNCT analysis depends on the particular ecological context. Goreaud and Pélissier (2003) state that under CSR independence, the two classes are *a priori* the result of different processes (e.g., individuals of different species or age cohorts). On the other hand, under RL, some processes affect *a posteriori* the individuals of a single population (e.g., diseased vs. non-diseased individuals of a single species).

As alternatives, we consider two major types of deviations from H_o, namely, 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. That is, the probability that this individual having a NN from the same class is larger than the relative frequency of the same class (see, e.g., Pielou, 1961). *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. That is, the probability that this individual having a NN from another class is larger than the relative frequency of the other class in question. These patterns are not symmetric, e.g., for two classes, one class might be more associated with the other class. For example, plant species X could be more dependent on species Y, hence X plants occur in close vicinity of Y plants, while the reverse relation may not be at the same level or type. Also, class X points might exhibit a stronger clustering, compared to class Y points, and thus might be more segregated compared to class Y points. See Ceyhan (2010a) for more details on the null and alternative patterns.

NNCTs are constructed to test spatial patterns in a multi-class framework. Under CSR independence or RL, cell counts, N_{ij}, would be close to their expected values, while under segregation the diagonal counts, N_{ii}, would be larger, while under association the off-diagonal counts, N_{ij}, would be larger than expected. When Pielou (1961) developed NNCT-tests, she used the usual Pearson's χ^2 test of independence for testing segregation, but this approach is not appropriate for completely mapped data due to the (spatial) dependence structure in a NNCT. Dixon (1994) derived the correct asymptotic distribution of the cell counts under RL and hence the appropriate test which also has a χ^2-distribution asymptotically. Pielou's test is shown to be appropriate for a random sample of (base,NN) pairs, while Dixon's test is appropriate for completely mapped data (Ceyhan, 2010b). Dixon (1994) also introduced cell-specific tests based on NNCTs and combined these tests to an overall test of segregation. However, Dixon's cell-specific tests only incorporate cell counts, and to get more information and (thus better tests) from a NNCT, we also consider four cell-specific tests which incorporate row and/or column sums as well. Since overall tests are obtained by combining cell-specific tests, they are expected to inherit the properties of the corresponding cell-specific tests. One main advantage of NNCT-tests is that for $m > 2$ classes, an overall NNCT-test provides a compound measure of any deviation(s) from the null pattern. That is, in analogy with ANOVA F-test for testing equality of multiple group means, an overall NNCT-test is a statistic testing the significance of deviations of cell-statistics from their expected values, or any deviation of a pair of classes from the null pattern. To the author's knowledge, no other such overall test exists in spatial statistics literature. Furthermore, in analogy with the employment of t-tests after a significant ANOVA F-test, cell-specific tests can be employed as post-hoc pairwise tests after a significant overall test, to determine which pair (pairs) deviates (deviate) significantly from the null pattern. The motivation to introduce new cell-specific and overall segregation tests was that Dixon's test is confounded by the differences between the class sizes (i.e., relative abundances), and requires each cell count to be at least 10 for asymptotic approximation to work for the cell-specific tests. So, NNCT-tests which are more robust to differences in relative abundances and asymptotic approximation working for smaller cell counts would have practical importance. Hence other NNCT-tests were introduced in Ceyhan (2008, 2010a) which are more robust to differences in relative abundances and require smaller cell counts when number of classes is more than two. Additionally, we consider Dixon's tests together with four variants of NNCT-tests (three of which are newly introduced) to determine the best performers in terms of robustness to relative abundance differences, and in terms of size and power performance for relatively small as well as large sample sizes.

2.1. Types of segregation and association

When H_o is rejected, if the diagonal entries (i.e., N_{ii} values) tend to be larger than expected, there is *segregation*; if the off-diagonal entries are larger than expected, there is *association*. These types of patterns are easy to detect for $m = 2$ classes, but for $m > 2$, rejecting H_o only indicates that there is some sort of deviation from the null case, but with many possible directions, since rejecting H_o only implies that for some class i, there exist classes that are more likely to serve as NN to class i or less likely to serve as NN to class i than expected under H_o. Let π_{ij} be the probability that a point is from class i and its NN is from class j. For example, for a fixed class i, if $\pi_{ii} \geq \sum_{j \neq i} \pi_{ij}$, then we have *total segregation* of class i from other classes; that is, class i is more likely to have a same class NN than all other classes combined. If $\pi_{ii} \geq \pi_{ij}$ for all $i \neq j$, then we have *strong segregation*, which implies that class i is more likely to have a con-specific NN compared to all other classes one at a time. Notice that total segregation implies strong segregation. The strict inequalities in the above definitions yield strict versions of total and strong segregation patterns.

For fixed classes i and j, with $i \neq j$, if $\pi_{ij} \geq \sum_{k \neq j} \pi_{ik}$, then we have *total association* of class j with class i; that is, class j is more likely to be a NN of class i than all other classes combined. If $\pi_{ij} \geq \pi_{ik}$ for all $k \neq j$, then we have *strong association* of class j with class i, which implies that class j is more likely to be a NN of class i compared to all other classes one at a time. Notice that total association implies strong association. Furthermore, the strict inequalities in the above definitions yield strict versions of total and strong association patterns.

On the other hand, if $\pi_{ii} \geq \pi_{ij}$ for all $j \in S_1 \subsetneq \{1, 2, \ldots, m\} \setminus \{i\}$ and $\pi_{ii} \leq \pi_{ij}$ for all $j \in S_2 = \{1, 2, \ldots, m\} \setminus (S_1 \cup \{i\})$, then we say that class i is *more segregated* from the classes in S_1 and *more associated* with the classes in S_2. We call such cases as *partial segregation* of class i with respect to classes in S_1 and *partial association* of class i with classes in S_2.

3. Cell-specific and overall segregation tests

We describe cell-specific segregation tests of Dixon and introduce new cell-specific tests labeled as *type I–IV* cell-specific tests, henceforth. We also define the overall segregation tests for each type of cell-specific test by following the same mechanism. More specifically, we merge the cell-specific tests to a vector (so that the indices follow the NNCT row-by-row), center it to zero, and then scale by the covariance of the cell-specific tests yielding an overall quadratic test statistic. Among the cell-specific and overall segregation tests considered here, type I, II and IV tests are newly introduced in this article.

3.1. Dixon's cell-specific and overall segregation tests

Dixon's cell-specific tests are used to measure the deviation of observed count in cell (i, j) in a NNCT from its expected value under H_o described in detail in, e.g., Dixon (1994, 2002a). Here, we provide the details for completeness. The test statistic suggested by Dixon for cell (i, j) is given by

$$Z_{ij}^D = \frac{N_{ij} - \mathbf{E}[N_{ij}]}{\sqrt{\mathbf{Var}[N_{ij}]}}, \tag{1}$$

where $\mathbf{E}[N_{ij}]$ is the expected cell count and $\mathbf{Var}[N_{ij}]$ is the variance of cell count N_{ij}.

For $m \geq 2$ classes, under RL or CSR independence, the expected cell count for cell (i, j) is

$$\mathbf{E}[N_{ij}] = \begin{cases} n_i(n_i - 1)/(n - 1) & \text{if } i = j, \\ n_i n_j/(n - 1) & \text{if } i \neq j, \end{cases} \tag{2}$$

where n_i is the fixed sample size for class i for $i = 1, 2, \ldots, m$. Observe that the expected cell counts depend only on the class sizes (i.e., row sums), but not on the column sums. And the variance is

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

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

$$p_{ii} = \frac{n_i\,(n_i - 1)}{n\,(n - 1)}, \qquad p_{ij} = \frac{n_i\, n_j}{n\,(n - 1)},$$

$$p_{iii} = \frac{n_i\,(n_i - 1)\,(n_i - 2)}{n\,(n - 1)\,(n - 2)}, \qquad p_{iij} = \frac{n_i\,(n_i - 1)\, n_j}{n\,(n - 1)\,(n - 2)}, \tag{4}$$

$$p_{iiii} = \frac{n_i\,(n_i - 1)\,(n_i - 2)\,(n_i - 3)}{n\,(n - 1)\,(n - 2)\,(n - 3)}, \qquad p_{iijj} = \frac{n_i\,(n_i - 1)\, n_j\,(n_j - 1)}{n\,(n - 1)\,(n - 2)\,(n - 3)}.$$

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_j is the number of points that serve as a NN to other points j times.

In the multi-class case with m classes, combining the m^2 cell-specific tests in Eq. (1), Dixon (2002a) suggests the following quadratic form to obtain the overall segregation test:

$$\mathcal{X}_D = (\mathbf{N} - \mathbf{E}[\mathbf{N}])' \Sigma_D^- (\mathbf{N} - \mathbf{E}[\mathbf{N}]) \tag{5}$$

where $\mathbf{N}$ is the $m^2 \times 1$ vector of m rows of the NNCT concatenated row-wise, $\mathbf{E}[\mathbf{N}]$ is the vector of $\mathbf{E}[N_{ij}]$ which are as in Eq. (2), Σ_D is the $m^2 \times m^2$ 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 (2002a). Also, Σ_D^- is a generalized inverse of Σ_D (Searle, 2006) and $'$ stands for the transpose of a vector or matrix. Then under RL, $\mathcal{X}_D$ has a $\chi^2_{m(m-1)}$ distribution asymptotically.

3.2. Type I cell-specific and overall segregation tests

In standard cases like multinomial sampling for contingency tables with fixed row totals and conditioning on the column totals, $C_j = c_j$, 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 $N_{ij} - \frac{n_i c_j}{n}$ for cell (i, j). However under RL, $N_i = n_i$ are fixed, but C_j are random quantities and $C_j = \sum_{i=1}^{m} N_{ij}$, hence we suggest as the first type of cell-specific segregation test as

$$T_{ij}^I = N_{ij} - \frac{n_i C_j}{n}.$$

Then under RL,

$$\mathbf{E}\left[T_{ij}^I\right] = \begin{cases} \dfrac{n_i(n_i - 1)}{(n-1)} - \dfrac{n_i}{n}\mathbf{E}[C_i] & \text{if } i = j, \\[2mm] \dfrac{n_i n_j}{(n-1)} - \dfrac{n_i}{n}\mathbf{E}[C_j] & \text{if } i \neq j. \end{cases} \tag{6}$$

For all j, $\mathbf{E}[C_j] = n_j$, since

$$\mathbf{E}[C_j] = \sum_{i=1}^{m}\mathbf{E}[N_{ij}] = \frac{n_j(n_j - 1)}{(n-1)} + \sum_{i \neq j}\frac{n_i n_j}{(n-1)} = \frac{n_j(n_j - 1)}{(n-1)} + \frac{n_j}{(n-1)}\sum_{i \neq j} n_i$$

$$= \frac{n_j(n_j - 1)}{(n-1)} + \frac{n_j}{(n-1)}(n - n_j) = n_j.$$

Therefore,

$$\mathbf{E}\left[T_{ij}^I\right] = \begin{cases} \dfrac{n_i(n_i - n)}{n(n-1)} & \text{if } i = j, \\[2mm] \dfrac{n_i n_j}{n(n-1)} & \text{if } i \neq j. \end{cases} \tag{7}$$

For the variance of T_{ij}^I, we have

$$\mathbf{Var}\left[T_{ij}^I\right] = \mathbf{Var}[N_{ij}] + \left(\frac{n_i^2}{n^2}\right)\mathbf{Var}[C_j] - 2\left(\frac{n_i}{n}\right)\mathbf{Cov}[N_{ij}, C_j] \tag{8}$$

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

As a new cell-specific test, we propose

$$Z_{ij}^I = \frac{T_{ij}^I - \mathbf{E}\left[T_{ij}^I\right]}{\sqrt{\mathbf{Var}\left[T_{ij}^I\right]}}. \tag{9}$$

We can also combine the type I cell-specific tests T_{ij}^I. Let $\mathbf{T_I}$ be the vector of m^2 T_{ij}^I values, i.e.,

$$\mathbf{T_I} = \left(T_{11}^I, T_{12}^I, \ldots, T_{1m}^I, T_{21}^I, T_{22}^I, \ldots, T_{2m}^I, \ldots, T_{mm}^I\right)',$$

and let $\mathbf{E}[\mathbf{T_I}]$ be the vector of $\mathbf{E}\left[T_{ij}^I\right]$ values. Note that

$$\mathbf{E}[\mathbf{T_I}] = \left(\mathbf{E}\left[T_{11}^I\right], \mathbf{E}\left[T_{12}^I\right], \ldots, \mathbf{E}\left[T_{1m}^I\right], \mathbf{E}\left[T_{21}^I\right], \mathbf{E}\left[T_{22}^I\right], \ldots, \mathbf{E}\left[T_{2m}^I\right], \ldots, \mathbf{E}\left[T_{mm}^I\right]\right).$$

Hence to obtain a new overall segregation test, referred to as *type I overall test*, we use the following quadratic form:

$$\mathcal{X}_I = (\mathbf{T_I} - \mathbf{E}[\mathbf{T_I}])' \, \Sigma_I^- \, (\mathbf{T_I} - \mathbf{E}[\mathbf{T_I}]) \tag{10}$$

where Σ_I is the $m^2 \times m^2$ variance–covariance matrix of $\mathbf{T_I}$.

Under RL, the diagonal entries in the variance–covariance matrix Σ_I are $\mathbf{Var}\left[T_{ij}^I\right]$ which are provided in Eq. (8). For the off-diagonal entries in Σ_I, i.e., $\mathbf{Cov}\left[T_{ij}^I, T_{kl}^I\right]$ with $(i, j) \neq (k, l)$, we have

$$\mathbf{Cov}\left[T_{ij}^I, T_{kl}^I\right] = \mathbf{Cov}\left[N_{ij} - \frac{n_i}{n}C_j, \, N_{kl} - \frac{n_k}{n}C_l\right]$$
$$= \mathbf{Cov}[N_{ij}, N_{kl}] - \frac{n_k}{n}\mathbf{Cov}[N_{ij}, C_l] - \frac{n_i}{n}\mathbf{Cov}[N_{kl}, C_j] + \frac{n_i n_k}{n^2}\mathbf{Cov}[C_j, C_l].$$

3.3. Type II cell-specific and overall segregation tests

In Section 3.2, we suggested $N_{ij} - \frac{n_i C_j}{n}$ as the test statistic for cell (i, j). However, under RL, $\mathbf{E}[C_j] = n_j$, so we suggest as the second type of segregation test as

$$T_{ij}^{II} = N_{ij} - \frac{n_i n_j}{n}.$$

Then under RL, $\mathbf{E}\left[T_{ij}^{II}\right] = \mathbf{E}\left[T_{ij}^I\right]$ which is provided in Eq. (7). Moreover, the variance of T_{ij}^{II} is $\mathbf{Var}\left[T_{ij}^{II}\right] = \mathbf{Var}[N_{ij}]$, since n_i, n_j and n are fixed.

As a cell-specific test, we propose

$$Z_{ij}^{II} = \frac{T_{ij}^{II} - \mathbf{E}\left[T_{ij}^{II}\right]}{\sqrt{\mathbf{Var}\left[T_{ij}^{II}\right]}}. \tag{11}$$

We also combine the type II cell-specific tests T_{ij}^{II}. Let $\mathbf{T_{II}}$ be the vector of m^2 T_{ij}^{II} values, i.e.,

$$\mathbf{T_{II}} = \left(T_{11}^{II}, T_{12}^{II}, \ldots, T_{1m}^{II}, T_{21}^{II}, T_{22}^{II}, \ldots, T_{2m}^{II}, \ldots, T_{mm}^{II}\right)',$$

and let $\mathbf{E}[\mathbf{T_{II}}]$ be the vector of $\mathbf{E}\left[T_{ij}^{II}\right]$ values. As the type II overall segregation test, we use the following quadratic form:

$$\mathcal{X}_{II} = (\mathbf{T_{II}} - \mathbf{E}[\mathbf{T_{II}}])' \, \Sigma_{II}^- \, (\mathbf{T_{II}} - \mathbf{E}[\mathbf{T_{II}}]) \tag{12}$$

where Σ_{II} is the $m^2 \times m^2$ variance–covariance matrix of $\mathbf{T_{II}}$.

Under RL, the diagonal entries in the variance–covariance matrix Σ_N are $\mathbf{Var}\left[T_{ij}^{II}\right]$ which are same as $\mathbf{Var}[N_{ij}]$. For the off-diagonal entries in Σ_{II}, i.e., $\mathbf{Cov}\left[T_{ij}^{II}, T_{kl}^{II}\right]$ with $(i, j) \neq (k, l)$, we have $\mathbf{Cov}\left[T_{ij}^{II}, T_{kl}^{II}\right] = \mathbf{Cov}[N_{ij} - \frac{n_i n_j}{n}, N_{kl} - \frac{n_k n_l}{n}] = \mathbf{Cov}[N_{ij}, N_{kl}]$.

3.4. Type III cell-specific and overall segregation tests

In the previous sections, $\mathbf{E}\left[T_{ij}^I\right] = \mathbf{E}\left[T_{ij}^{II}\right] \neq 0$ under RL. Hence, instead of these test statistics, in order to have the expected value of our test statistic to be zero, we suggest the following test statistic:

$$T_{ij}^{III} = \begin{cases} N_{ii} - \dfrac{(n_i - 1)}{(n - 1)}C_i & \text{if } i = j, \\[2mm] N_{ij} - \dfrac{n_i}{(n - 1)}C_j & \text{if } i \neq j. \end{cases} \tag{13}$$

This test statistic is the same as the new cell-specific test introduced in Ceyhan (2010a) and details of this test are provided here for completeness. Then $\mathbf{E}\left[T_{ij}^{III}\right] = 0$, since, for $i = j$,

$$\mathbf{E}\left[T_{ii}^{III}\right] = \mathbf{E}[N_{ii}] - \frac{(n_i - 1)}{(n - 1)}\mathbf{E}[C_i] = \frac{n_i(n_i - 1)}{(n - 1)} - \frac{(n_i - 1)}{(n - 1)}n_i = 0,$$

and for $i \neq j$,

$$\mathbf{E}\left[T_{ij}^{III}\right] = \mathbf{E}[N_{ij}] - \frac{(n_i - 1)}{(n - 1)}\mathbf{E}[C_j] = \frac{n_i n_j}{(n - 1)} - \frac{(n_i - 1)}{(n - 1)}n_j = 0.$$

As for the variance of T_{ij}^{III}, we have

$$
\mathbf{Var}\left[T_{ij}^{III}\right] = \begin{cases} \mathbf{Var}[N_{ii}] + \dfrac{(n_i - 1)^2}{(n-1)^2}\mathbf{Var}[C_i] - 2\dfrac{(n_i - 1)}{(n-1)}\mathbf{Cov}[N_{ii}, C_i] & \text{if } i = j, \\[4mm] \mathbf{Var}[N_{ij}] + \dfrac{n_i^2}{(n-1)^2}\mathbf{Var}[C_j] - 2\dfrac{n_i}{(n-1)}\mathbf{Cov}[N_{ij}, C_j] & \text{if } i \neq j. \end{cases}
\tag{14}
$$

As a new cell-specific test, we propose

$$
Z_{ij}^{III} = \frac{T_{ij}^{III}}{\sqrt{\mathbf{Var}\left[T_{ij}^{III}\right]}}.
\tag{15}
$$

When we combine the type III cell-specific tests T_{ij}^{III}, we obtain type III overall test as follows. Let $\mathbf{T_{III}}$ be the vector of m^2 T_{ij}^{III} values, i.e.,

$$
\mathbf{T_{III}} = \left(T_{11}^{III}, T_{12}^{III}, \ldots, T_{1m}^{III}, T_{21}^{III}, T_{22}^{III}, \ldots, T_{2m}^{III}, \ldots, T_{mm}^{III}\right)',
$$

and let $\mathbf{E}\left[\mathbf{T_{III}}\right]$ be the vector of $\mathbf{E}\left[T_{ij}^{III}\right]$ values. Note that $\mathbf{E}\left[\mathbf{T_{III}}\right] = \mathbf{0}$ where $\mathbf{0}$ stands for a vector of zeros. As the type III overall segregation test, we use the following quadratic form:

$$
\mathcal{X}_{III} = \mathbf{T_{III}'}\,\Sigma_{III}^{-}\,\mathbf{T_{III}}
\tag{16}
$$

where Σ_{III} is the $m^2 \times m^2$ variance–covariance matrix of $\mathbf{T_{III}}$.

Under RL, the diagonal entries in the variance–covariance matrix Σ_{III} are $\mathbf{Var}\left[T_{ij}^{III}\right]$ which are provided in Eq. (14). For the off-diagonal entries in Σ_{III}, i.e., $\mathbf{Cov}\left[T_{ij}^{III}, T_{kl}^{III}\right]$ with $(i, j) \neq (k, l)$, there are four cases to consider:

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

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

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

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

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

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

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

3.5. Type IV cell-specific and overall segregation tests

For T_{ij}^{III}, we introduced a coefficient in front of the second term, i.e., $n_i C_j / n$, to obtain the expected value for our statistic to be zero under RL. In this section, we modify the first term and obtain the following test statistic:

$$
T_{ij}^{IV} = \begin{cases} \dfrac{n_i(n-1)}{n(n_i - 1)}N_{ii} - \dfrac{n_i}{n}C_i = \dfrac{n_i}{n}\left(\dfrac{n-1}{n_i - 1}N_{ii} - C_i\right) & \text{if } i = j, \\[4mm] \dfrac{n-1}{n}N_{ij} - \dfrac{n_i}{n}C_j = \dfrac{1}{n}\left((n-1)\,N_{ij} - n_i\,C_j\right) & \text{if } i \neq j. \end{cases}
\tag{17}
$$

Then $\mathbf{E}\left[T_{ij}^{IV}\right] = 0$, since, for $i = j$,

$$
\mathbf{E}\left[T_{ii}^{IV}\right] = \frac{n_i}{n}\left(\frac{n-1}{n_i - 1}\mathbf{E}[N_{ii}] - \mathbf{E}[C_i]\right) = \frac{n_i}{n}\left(\frac{n-1}{n_i - 1}\frac{n_i(n_i - 1)}{n-1} - n_i\right) = 0,
$$

and for $i \neq j$,

$$\mathbf{E}\left[T_{ij}^{IV}\right] = \frac{1}{n}\left((n-1)\mathbf{E}[N_{ij}] - n_i\mathbf{E}[C_j]\right) = \frac{1}{n}\left((n-1)\frac{n_i\,n_j}{n-1} - n_i\,n_j\right) = 0.$$

As for the variance of T_{ij}^{IV}, we have

$$\mathbf{Var}\left[T_{ij}^{IV}\right] = \begin{cases} \dfrac{n_i^2}{n^2}\left(\dfrac{(n-1)^2}{(n_i-1)^2}\mathbf{Var}[N_{ii}] + \mathbf{Var}[C_i] - 2\dfrac{(n-1)}{(n_i-1)}\mathbf{Cov}[N_{ii}, C_i]\right) & \text{if } i = j, \\ \dfrac{1}{n^2}\left((n-1)^2\mathbf{Var}[N_{ij}] + n_i^2\mathbf{Var}[C_j] - 2(n-1)n_i\mathbf{Cov}[N_{ij}, C_j]\right) & \text{if } i \neq j. \end{cases} \tag{18}$$

As a new cell-specific test, we propose

$$Z_{ij}^{IV} = \frac{T_{ij}^{IV}}{\sqrt{\mathbf{Var}\left[T_{ij}^{IV}\right]}}. \tag{19}$$

When we combine the type IV cell-specific tests T_{ij}^{IV}, we obtain type IV overall test as follows. Let $\mathbf{T_{IV}}$ be the vector of $m^2\ T_{ij}^{IV}$ values, i.e.,

$$\mathbf{T_{IV}} = \left(T_{11}^{IV}, T_{12}^{IV}, \ldots, T_{1m}^{IV}, T_{21}^{IV}, T_{22}^{IV}, \ldots, T_{2m}^{IV}, \ldots, T_{mm}^{IV}\right)',$$

and let $\mathbf{E}\left[\mathbf{T_{IV}}\right]$ be the vector of $\mathbf{E}\left[T_{ij}^{IV}\right]$ values. Note that $\mathbf{E}\left[\mathbf{T_{IV}}\right] = \mathbf{0}$. As the type IV overall segregation test, we use the following quadratic form:

$$\mathcal{X}_{IV} = \mathbf{T_{IV}'}\Sigma_{IV}^{-}\mathbf{T_{IV}} \tag{20}$$

where Σ_{IV} is the $m^2 \times m^2$ variance–covariance matrix of $\mathbf{T_{IV}}$.

Under RL, the diagonal entries in the variance–covariance matrix Σ_{IV} are $\mathbf{Var}\left[T_{ij}^{IV}\right]$ which are provided in Eq. (18). For the off-diagonal entries in Σ_{IV}, i.e., $\mathbf{Cov}\left[T_{ij}^{IV}, T_{kl}^{IV}\right]$ with $(i, j) \neq (k, l)$, there are four cases to consider:

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

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

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

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

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

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

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

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

3.6. Rank-adjustment for the overall NNCT-tests

Asymptotically, under RL, $\mathcal{X}_D$ and $\mathcal{X}_{II}$ have a $\chi^2_{m(m-1)}$ distribution since rank of Σ_D and Σ_{II} is $m(m-1)$, in fact, $\Sigma_D = \Sigma_{II}$. $\mathcal{X}_{III}$ has a $\chi^2_{(m-1)^2}$ distribution asymptotically since rank of Σ_{III} is $(m-1)^2$. Similarly, $\mathcal{X}_I$ and $\mathcal{X}_{IV}$ also have $\chi^2_{(m-1)^2}$ distribution asymptotically.

In the mechanism for obtaining overall tests from the cell-specific tests, we form the vector of m^2 (normalized) cell-specific tests, and then form the quadratic term for the overall tests that involve the generalized inverse of the $m^2 \times m^2$ covariance matrix for the vector of cell-specific tests. Although, in general, generalized inverse of a matrix is not unique, the Moore–Penrose procedure yields a unique generalized inverse under certain constraints (see, e.g., Penrose, 1955). Since the covariance matrix for Dixon's and type II tests are of rank $m(m-1)$ and for type I, III, and IV tests, the rank is $(m-1)^2$, the covariance matrix is rank-deficient. The generalized inverse procedure for rank-deficient matrices is a complicated procedure which may result in an inverse that is not well-behaving (Meyer, 1973). For example, the generalized inverse may yield unstable results for the overall tests in the sense that the quadratic overall test may give negative or extremely large values, even under RL or CSR independence, where getting a negative overall test would be impossible from such a quadratic form. To avoid this problem, we adjust the corresponding cell-specific vectors and the covariance matrices. Along this line, for Dixon's and type II tests, we remove the cell-specific tests with indices corresponding to the last column entries in the NNCT (since it contains redundant information, given the class sizes) in our computations. The removal of the last column is an arbitrary choice, as removal of any other column entries would also yield the same result. Similarly, for type I, III and IV tests, we remove the cell-specific tests with indices corresponding to the last row and column in the NNCT (as given the row and column sums, these are redundant). Similarly, removal of last row and column entries is an arbitrary choice, as removal of any other row and column entries would also work. The covariance matrices formed with the modified NNCT as above are now of full rank, and the generalized inverse is a stable and reliable procedure for such matrices. Without such an adjustment, the tests tend to be unstable with substantially larger variances, and especially type IV overall test yields much different results compared to type III test, although the corresponding cell-specific tests are identical. To see the effect of such a rank-adjustment on the overall tests, we generate n_i points i.i.d. uniformly on the unit square for each of classes $i = 1, 2, \ldots, m$, and determine the rank of the covariance matrices for the full NNCT and also compute the overall tests for the full NNCT (i.e., without rank-adjustment) and also for the reduced NNCT (i.e., with the rank-adjustment). We repeat the above procedure $N_{mc} = 1000$ times, and present the ranks and the proportion of negative overall tests, denoted p_{neg}, in Table 2. Observe that in each replication, we obtain the rank as expected for each overall test (i.e., for Dixon's and type II tests, rank is $m(m-1)$ and for type I, III, and IV tests, the rank is $(m-1)^2$). With no rank adjustment, the proportion of negative overall tests is zero or very close to zero for Dixon's and type II tests, but it is positive for type I, III, and IV tests (and substantially large for some of them). With the rank-adjustment, we obtain no negative overall test and the rank values are as expected (results are not presented) and each test has more stable behavior with a distribution much closer to its corresponding limiting (asymptotic) distribution. We further note that for Dixon's and type II tests, the rank-adjustment generally has a mild influence, and slightly improves their behavior. Nonetheless, in our subsequent analysis, *we implement rank-adjustment for all of the overall tests considered.*

The asymptotic distributions of the overall tests provide a natural classification of these NNCT-tests. More specifically, Dixon's and type II overall tests only use the cell counts and row sums (i.e., class sizes) in the corresponding cell-specific tests and hence asymptotically have χ^2 distribution with $m(m-1)$ df, while type I, III and IV overall tests use the column sums in addition to cell counts and row sums and hence have asymptotic χ^2 distribution with $(m-1)^2$ df. That is, if only the row sums are incorporated, then one df is lost in each row as the sums of the row cells yield the fixed class sizes. On the other hand, if both row and column sums are incorporated, one row and column can be obtained given the row and column sums, hence leaving only $(m-1)^2$ df for the overall tests. It can be easily established that Dixon's and type II tests are identical. Type III and IV cell-specific tests (after standardization) and overall tests are also identical, although the T_{ij} values and the variance–covariance matrices are different. Additionally, type I tests and type III tests are similar (but not identical), and hence give similar results.

3.7. Further remarks on NNCT-tests

(i) **Asymptotic structures for NNCT-tests**: There are two major types of asymptotic structures for spatial data in literature: *infill asymptotics* and *increasing domain asymptotics* (Lahiri, 1996). In infill asymptotics, the region of interest is a fixed bounded region and the number of observed points gets larger in this region. Hence the minimum distance between data points tends to zero as the sample size tends to infinity. In increasing domain asymptotics, any two observations are required to be at least a fixed distance apart. Hence as the number of observations increases, the region on which the process is observed eventually becomes unbounded (Cressie, 1993). The sampling structure in our asymptotic sampling distribution could be either one of these asymptotic structures. Because we only consider the class sizes tending to infinity (i.e., $\min_{i=1}^{m} n_i \to \infty$) and hence the total sample size tending to infinity regardless of the size of the study region.

Table 2
The ranks and the proportion of negative overall tests, p_{neg}, based on the full (i.e., not adjusted for rank) NNCTs for uniform data in the unit square with m classes based on $N_{mc} = 1000$ replicates. For $m = 2$, the class sizes are $n_1 = 100$, $n_2 = 200$; for $m = 3$ the class sizes are $n_1 = 100$, $n_2 = 150$, $n_3 = 200$; and for $m = 4$, the class sizes are $n_1 = 100$, $n_2 = 133$, $n_3 = 166$, $n_4 = 200$.

	Overall tests				
	$\mathcal{X}_D$	$\mathcal{X}_I$	$\mathcal{X}_{II}$	$\mathcal{X}_{III}$	$\mathcal{X}_{IV}$
		$m = 2$ classes			
Rank	2	1	2	1	1
p_{neg}	0	0.117	0	0.088	0.110
		$m = 3$ classes			
Rank	6	4	6	4	4
p_{neg}	0	0.010	0	0.467	0.030
		$m = 4$ classes			
Rank	12	9	12	9	9
p_{neg}	0.002	0.090	0.002	0.056	0.071

(ii) **Asymptotic distribution of cell-specific tests**: Under RL, Dixon's cell-specific tests have $N(0, 1)$ distribution asymptotically for $m = 2$ classes, but for $m > 2$ classes his cell-specific tests for the diagonal cells, Z_{ii}^D, are shown to have $N(0, 1)$ distribution asymptotically, while the asymptotic normality of the off-diagonal cells in NNCTs is not rigorously established yet, although extensive Monte Carlo simulations indicate approximate normality for large samples (Dixon, 2002a). The asymptotic normality of the off-diagonal cells for $m > 2$ classes can be established when classes are from a homogeneous Poisson Process or a binomial process as suggested in Penrose (2003), and for RL, the result of Cuzick and Edwards (1990) could be generalized. Both aspects (i.e., asymptotic normality under CSR independence and RL) are topics of ongoing research, with significant progress achieved in them recently. However, the particular result in Cuzick and Edwards (1990) required more conditions (e.g., symmetry) which do not hold for the off-diagonal cells. The Monte Carlo simulations provide substantial evidence in favor of asymptotic normality even at moderate sample sizes. Therefore, as a safety check, for the tests on the off-diagonal cells and for the overall test for three or more classes, one should employ both the asymptotic approximation version and the Monte Carlo randomized version. The same conclusions can be extended to type I–IV cell-specific tests in a similar fashion. Furthermore, after standardization, Dixon's cell-specific test and type II cell-specific test are identical, and so are type III and IV cell-specific tests. The same holds for the corresponding overall tests, since they are constructed based on the cell-specific tests. However, the cell-specific test statistics are dependent, hence their squares do not sum to the corresponding overall segregation tests.

(iii) **Status of Q and R under CSR independence and RL**: Under CSR independence, the cell-specific and overall tests are as in the RL case. However, under RL, Q and R are fixed quantities, as they depend only on the location of the points, but not the types of NNs, while under CSR independence, they are random. Under CSR independence, the distributions of the test statistics above are similar to the RL case. The only difference is that the new cell-specific tests asymptotically have $N(0, 1)$ distribution conditional on Q and R. Hence, under CSR independence, $\mathbf{Var}[N_{ij}]$, $\mathbf{Cov}[N_{ij}, N_{kl}]$, $\mathbf{Cov}[N_{ij}, C_k]$, $\mathbf{Cov}[C_i, C_j]$, and all other quantities depending on Q and R are conditional on Q and R. The unconditional variances can be obtained by replacing Q and R with their expectations (see Ceyhan, 2010a for more details). Since Q and R are random under CSR independence, the variances of the cell-specific test statistics tend to be larger compared to the ones under RL.

(iv) **Other spatial patterns**: Not all stochastic independence patterns yield random NN structure similar to that under RL or CSR independence. For example, a Matérn cluster process can be generated in two stages (Stoyan & Stoyan, 1994): in the first stage, parent points are generated from a homogeneous Poisson process with intensity λ and in the second stage, daughter points are uniformly and independently distributed within a distance r from each parent event. In the resulting process, the daughter points constitute a realization of a Matérn cluster process. Even if two classes are generated independently from the same Matérn cluster process, they will be very likely to be segregated unless the same parent points are used for both classes, because points from the classes will be clustered around their corresponding parents. The random NN structure similar to that under RL or CSR independence can be obtained if the same parent points are used for both classes.

(v) **Interpretation of NNCT-tests**: Each of the cell-specific tests measures the deviation of the test statistic from its expected value under H_o. Dixon's and type II cell-specific tests depend on N_{ij} (i.e., cell counts) and row sums only, and type I, III, and IV cell-specific tests incorporate column sums as well. For the cell-specific tests, the z-score for cell (i, j) indicates the level and direction of spatial interaction between classes i and j. If the z-score for cell (i, i) is significantly larger (less) than zero, then class i exhibits (lack of) segregation from other classes. If the z-score for cell (i, j) with $i \neq j$ is significantly larger (less) than zero, then class j exhibits (lack of) association with class i. Moreover, for cell (i, j) with $i \neq j$, the cell-specific tests are not symmetric. For example, the cell-specific test for cell (i, j) may exhibit a different level of interaction compared to the cell (j, i). The overall tests combine cell-specific tests in one compound summary statistic. The performance of cell-specific tests is expected to carry over to the overall tests, provided the correct degrees of freedom is used.

Since each cell count N_{ij} has asymptotic normal distribution in the two-class case (Cuzick & Edwards, 1990), the new cell-specific tests $Z_{ij}^I, Z_{ij}^{II}, Z_{ij}^{III}$ and Z_{ij}^{IV} also converge in law to $N(0, 1)$ as $n \to \infty$ (with $n_i \to \infty$ for all i). Moreover, one and two-sided versions of these tests are also possible. For the diagonal cells, the right-sided (left-sided) version of these tests is for (lack of) segregation and for the off-diagonal cells, the right-sided (left-sided) version of these tests is for (lack of) association. In the two-class case, only at most two cells contain all the information provided by the NNCT. In particular, for $i \neq j$, segregation of class i from class j implies lack of association between classes i and j and lack of segregation of class i from class j implies association between classes i and j. For Dixon's cell-specific test, we have $Z_{i1}^D = -Z_{i2}^D$ for $i = 1, 2$. For type I cell-specific test, $Z_{11}^I = Z_{22}^I = -Z_{12}^I = -Z_{21}^I$; and for type II cell-specific test, we have $Z_{ij}^D = Z_{ij}^{II}$; for type III cell-specific test, we have $Z_{1j}^{III} = -Z_{2j}^{III}$ for $j = 1, 2$; and for type IV cell-specific test, we have $Z_{ij}^{III} = Z_{ij}^{IV}$ for $i, j = 1, 2$.

In the multi-class case with $m > 2$, 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. See also Section 2.1.

The cell-specific and overall tests are all consistent under both segregation and association alternatives, which can be proved as in Ceyhan (2010b).

3.8. Post-hoc tests after the overall tests: class-specific, pairwise, and one-vs-rest type tests

In our construction of the NNCT-tests, although we first introduce the cell-specific tests and then develop overall tests based on the cell-specific tests, in practice, it is more natural to conduct the tests in reverse order. That is, first an overall NNCT-test could be performed, and if significant, then one can perform cell-specific tests to determine the types and levels of the spatial interaction patterns between the classes. This procedure is somewhat analogous to ANOVA F-test to compare multiple groups, in the sense that if the F-test yields a significant result, then one performs pairwise tests to determine which pairs are different. However, NNCT-tests can provide more alternatives (compared to ANOVA F-test) as post-hoc tests after a significant overall test. In the multi-class case, when an overall test is rejected; i.e., there is evidence in favor of some sort of deviation from randomness of the spatial pattern, the next natural question is what type of deviation occurs for each class (or species). To this end, one can conduct several post-hoc tests: (i) One type of post-hoc tests is the class-specific tests discussed in Ceyhan (2009) and Dixon (2002a). (ii) For pairwise comparison of the interaction between classes, one can resort to two options: (a) in an $m \times m$ NNCT, one can consider cell-specific tests for each cell (which also provides interaction of the class with itself on the diagonal cells) and (b) one can restrict attention to the pair of classes i, j with $i \neq j$ one at a time and conduct the tests as in the two-class case with a 2×2 NNCT. Here the 2×2 NNCT can be formed as if only classes i and j are present in the study region ignoring the other classes. Notice that in (b), Q and R values should also be updated. We recommend the approach in (a), since it incorporates all the classes in question and provides the types of interaction in the presence of all classes, while the approach in (b) ignores the possible confounding effects of classes different from the pair in question. Furthermore, the approach in (b) might not give the exact picture of the mixed relationships between all the classes in practice. (iii) Alternatively, for class i, we can pool the remaining classes and treat them as the other class in a two-class setting. Then we apply the two-class tests to the resulting NNCT. To emphasize the difference, this version of the class-specific test is called *one-vs-rest type test*. For $m > 2$ classes, recall that $\mathcal{N}$ represents the $m \times m$ NNCT with cell counts being N_{ij} and let $\widetilde{\mathcal{N}}$ be the 2×2 NNCT for the one-versus-rest type procedure with cell counts being $\widetilde{N}_{ij}$. When we are performing a one-versus-rest type testing for class i, without loss of generality, we can reserve the first row in $\widetilde{\mathcal{N}}$ to class i and the second row to the rest (of the classes). Then $\widetilde{N}_{11} = N_{ii}, \widetilde{N}_{12} = \sum_{j \neq i} N_{ij}, \widetilde{N}_{21} = \sum_{j \neq i} N_{ji}$, and $\widetilde{N}_{22} = \sum_{j \neq i, k \neq i} N_{jk}$. Hence in the one-versus-rest type testing, the cell-specific test for cell $(1, 1)$ in $\widetilde{\mathcal{N}}$ would be same as the cell-specific test for cell (i, i) in $\mathcal{N}$. Therefore, to extract information from $\widetilde{\mathcal{N}}$ that is not provided by $\mathcal{N}$, we consider the cell-specific tests for cell $(2, 2)$ of $\widetilde{\mathcal{N}}$. The overall test statistics for $\widetilde{\mathcal{N}}$ are also different than the overall tests for $\mathcal{N}$. Although one-vs-rest type tests were first suggested by Dixon (2002a) in passing, their properties and performance are only discussed and assessed in detail in this article.

In a multi-class case with $m \geq 2$ classes, there are m class-specific and m one-vs-rest types of tests and $\binom{m}{2} = m(m-1)/2$ pairwise tests and m^2 cell-specific tests. As m increases the class-specific tests are less intensive computationally and easier to interpret, whereas the pairwise tests might yield conflicting results.

4. Empirical size analysis in the two-class case

We provide the empirical significance levels for Dixon's and the new cell-specific and overall segregation tests in the two-class case under CSR independence.

4.1. Empirical size analysis under CSR independence of two classes

For the CSR independence pattern in the two-class case, we label the classes as X and Y, or class 1 and class 2, interchangeably. We generate n_1 points from class X and n_2 points from class Y both of which are independent of each other and independently uniformly distributed on the unit square, $(0, 1) \times (0, 1)$. We use the class size combinations
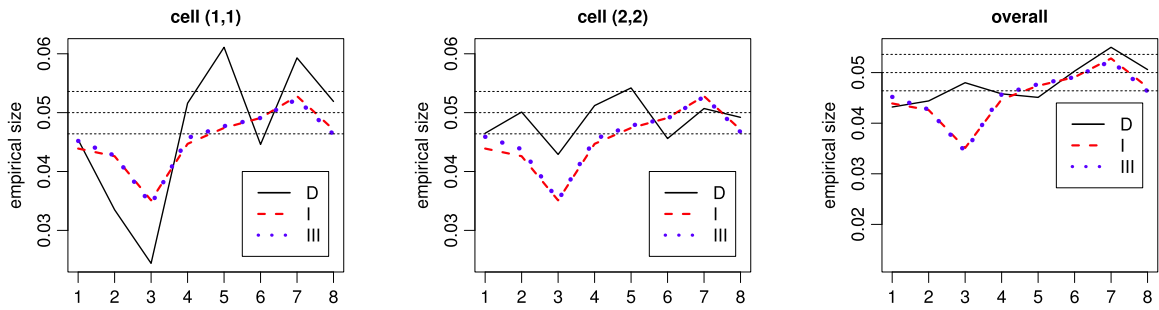
Fig. 1. The empirical size estimates of the cell-specific tests for cells (1, 1) (left), cell (2, 2) (middle), and overall segregation tests (right) under the **CSR independence pattern in the two-class case**. The horizontal lines are located at 0.0464 (upper threshold for conservativeness), 0.05 (nominal level), and 0.0536 (lower threshold for liberalness). 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). The legend labeling: D = Dixon's, I = type I, and III = type III cell-specific or overall tests.

$(n_1, n_2) \in \{(10, 10), (10, 30), (10, 50), (30, 30), (30, 50), (50, 50), (50, 100), (100, 100)\}$ and perform $N_{mc} = 10\,000$ replications. The empirical sizes are calculated as the ratio of number of significant results to the number of Monte Carlo replications, N_{mc}. We use 0.05 as our nominal significance level.

We present the empirical significance levels for the NNCT-tests in Fig. 1. The empirical sizes significantly smaller (larger) than 0.05 are deemed as conservative (liberal). The asymptotic normal approximation to proportions are used in determining the significance of the deviations of the empirical sizes from the nominal level of 0.05. For these proportion tests, we also use $\alpha = 0.05$ to test against empirical size being equal to 0.05. With $N_{mc} = 10\,000$, empirical sizes less than 0.0464 are deemed conservative, greater than 0.0536 are deemed liberal at $\alpha = 0.05$ level. These thresholds are indicated as the dashed horizontal lines in Fig. 1. Note also that the class sizes are arranged in the increasing order for the first and then the second entries. The size values for discrete class size combinations are joined by straight lines for better visualization. Let $\widehat{\alpha}_{i,j}^D, \widehat{\alpha}_{i,j}^I$–$\widehat{\alpha}_{i,j}^{IV}$ be the empirical significance levels of Dixon's and the type I–IV cell-specific tests, respectively, $\widehat{\alpha}_D$ be for Dixon's and $\widehat{\alpha}_I$–$\widehat{\alpha}_{IV}$ be for the type I–IV overall segregation tests. Notice that in the two-class case $\widehat{\alpha}_{1,1}^D = \widehat{\alpha}_{1,2}^D$ and $\widehat{\alpha}_{2,1}^D = \widehat{\alpha}_{2,2}^D$ for the two-sided alternative, since $N_{12} = n_1 - N_{11}$ and $N_{21} = n_2 - N_{22}$. The same holds for $\widehat{\alpha}_{i,j}^{II}$. Furthermore, $\widehat{\alpha}_{1,1}^{III} = \widehat{\alpha}_{2,1}^{III}$ and $\widehat{\alpha}_{1,2}^{III} = \widehat{\alpha}_{2,2}^{III}$ for the two-sided alternative, and the same holds for $\widehat{\alpha}_{i,j}^{IV}$. On the other hand $\widehat{\alpha}_{i,j}^I$ are equal for all i, j for the two-sided alternative. So we only present cell-specific tests for cells (1, 1) and (2, 2) in the two-class settings in the rest of the article. Furthermore, since Dixon's cell specific test and type II cell-specific test are equivalent, and so are type III and IV cell-specific tests, we only present Dixon's, type I and III cell-specific tests. Since the same holds for the overall test, we present only the corresponding tests as well.

For cell (1, 1), Dixon's cell-specific test has empirical size close to the nominal level of 0.05 for balanced class sizes (i.e., for $n_1 \approx n_2$ or when relative abundance of classes are similar), while for unbalanced class sizes, it tends to be liberal or conservative. On the other hand, type I and type III cell-specific tests are less severely affected by the differences in relative abundances of the classes, i.e., they are closer to the nominal level for all class size combinations. For cell (2, 2), Dixon's cell-specific test is much closer to 0.05 for all class size combinations, while type I and III cell-specific tests have similar performance as in cell (1, 1). Thus, Dixon's cell-specific test has much better empirical size performance for the diagonal cell corresponding to the class with larger size, while type I and type III cell-specific tests have better size performance for the diagonal cell corresponding to the class with smaller size.

For the overall tests, Dixon's test has better size performance for classes of smaller size. Type I and III overall tests are conservative for classes of smaller size, while they have the desired level for classes of larger size.

The empirical size performance under RL is similar (see Ceyhan, 2013), hence is not presented.

5. Empirical size analysis in the three-class case

In this section, we provide the empirical significance levels for the overall and cell-specific segregation tests in the three-class case under RL and CSR independence patterns.

5.1. Empirical size analysis under CSR independence of three classes

The symmetry in cell counts for rows in Dixon's cell-specific tests and columns in the new cell-specific tests occurs only in the two-class case. To assess the performance of the cell-specific and overall tests better, we also consider the three-class case. In the three-class case, we label the classes as X, Y, and Z or classes 1, 2, and 3, interchangeably. We generate n_1, n_2, n_3 points distributed independently uniformly on the unit square $(0, 1) \times (0, 1)$ from these classes. We use

$$(n_1, n_2, n_3) \in \{(10, 10, 10), (10, 10, 30), (10, 10, 50), (10, 30, 30), (10, 30, 50), (30, 30, 30), (10, 50, 50),$$
$$(30, 30, 50), (30, 50, 50), (50, 50, 50), (50, 50, 100), (50, 100, 100), (100, 100, 100)\};$$

and $N_{mc} = 10\,000$. The empirical sizes and the significance of their deviation from 0.05 are calculated as in Section 4.1.
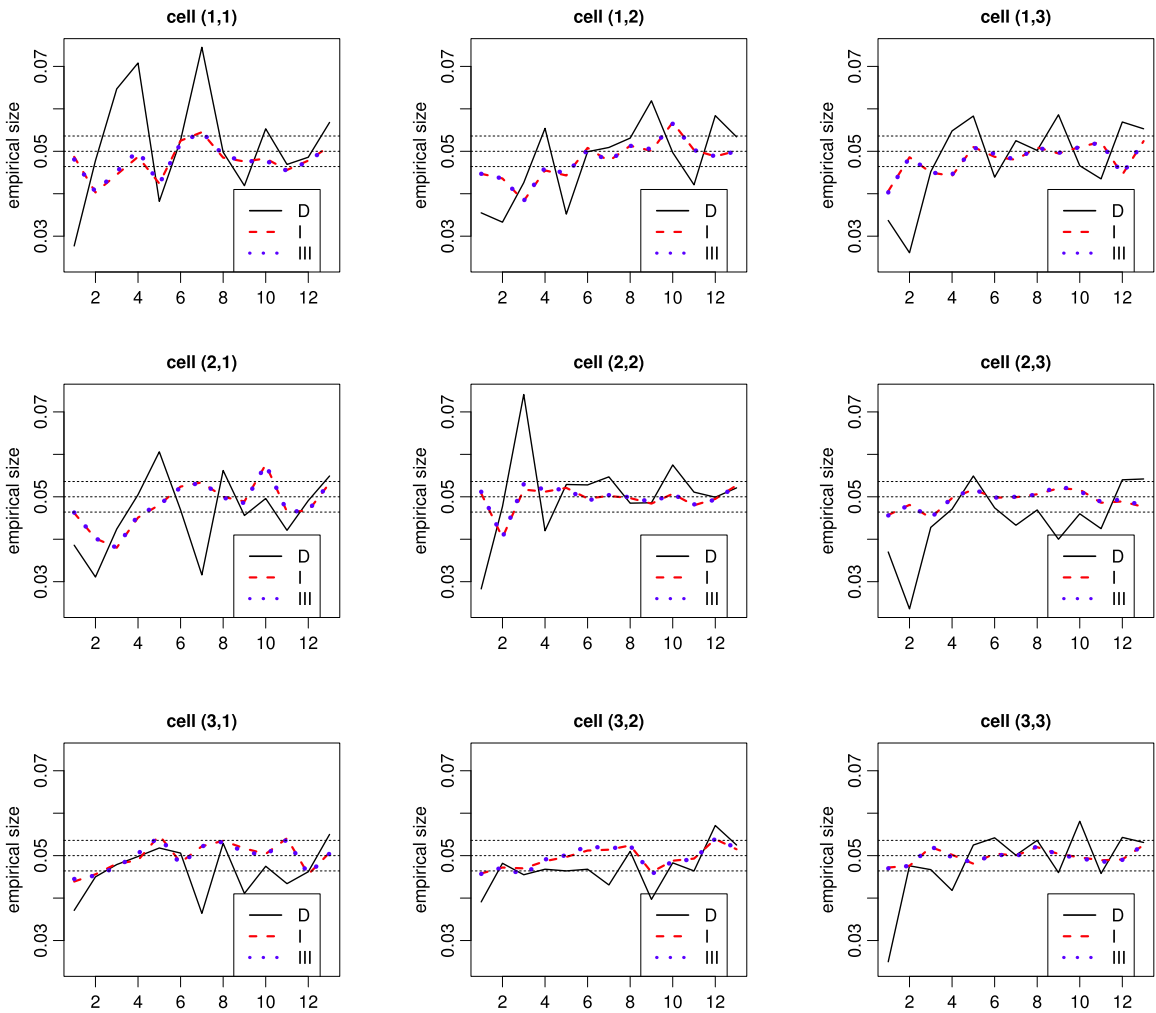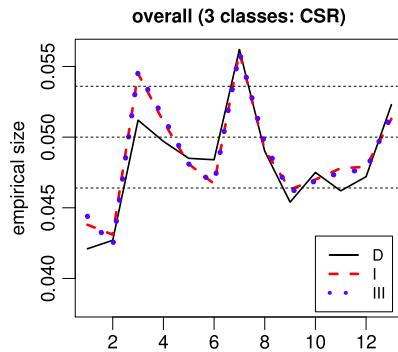
Fig. 2. The empirical size estimates of the cell-specific tests for cells $(1, 1)$–$(3, 3)$ under the **CSR independence pattern in the three-class case**. The horizontal lines and legend labeling are as in Fig. 1. The horizontal axis labels are: $1 = (10, 10, 10)$, $2 = (10, 10, 30)$, $3 = (10, 10, 50)$, $4 = (10, 30, 30)$, $5 = (10, 30, 50)$, $6 = (30, 30, 30)$, $7 = (10, 50, 50)$, $8 = (30, 30, 50)$, $9 = (30, 50, 50)$, $10 = (50, 50, 50)$, $11 = (50, 50, 100)$, $12 = (50, 100, 100)$, $13 = (100, 100, 100)$.

Fig. 3. The empirical size estimates of the overall tests under the **CSR independence pattern in the three-class case**. The horizontal lines, and legend labeling are as in Fig. 1 and axis labels are as in Fig. 2.

We present the empirical significance levels for the cell-specific tests in Fig. 2 and for the overall tests in Fig. 3. For the cell-specific tests, clearly, type I and III tests are closer to the desired level, and are less affected by the differences in class sizes. On the other hand, Dixon's test is extremely liberal or conservative, when class sizes are very different (which may
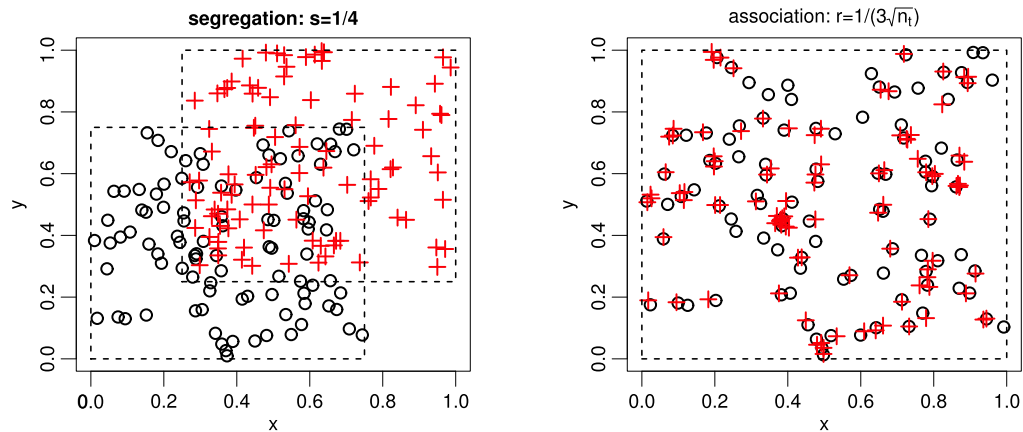
Fig. 4. A sample realization of segregation alternative H_S^{II} (left) and the association alternative H_A^{II} (right) with $n_1 = n_2 = 100$. Class X (or class 1) points are marked with circles (∘), and class Y (or class 2) points are marked with pluses (+). The supports of the classes are indicated by squares bounded by dashed lines.

result in smaller expected cell counts). The overall tests have similar size performance with Dixon's test being slightly better for classes of smaller size, while types I and III slightly better for classes of larger size.

The empirical size performance under RL is similar (see Ceyhan, 2013), hence is not presented.

6. Empirical power analysis in the two-class case

To assess the empirical power performance of the tests, we consider three cases for each of segregation and association alternatives in the two-class case.

6.1. Empirical power analysis under segregation of two classes

Under the segregation alternatives for two classes, we generate $X_i \overset{i.i.d.}{\sim} \mathcal{U}(S_1)$ and $Y_j \overset{i.i.d.}{\sim} \mathcal{U}(S_2)$ where $S_1 = (0, 1-s) \times (0, 1-s)$ and $S_2 = (s, 1) \times (s, 1)$ for $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$ and $s \in (0, 1)$. This is the same segregation setting used in Ceyhan (2008, 2010a). We have opted this very same setup, because we want to compare all the tests under the settings same as before. We consider the following three segregation alternatives:

$$H_S^I : s = 1/6, \qquad H_S^{II} : s = 1/4, \quad \text{and} \quad H_S^{III} : s = 1/3. \tag{21}$$

Notice that, the level of segregation increases as s increases; that is, segregation gets stronger from H_S^I to H_S^{III}. A sample realization of H_S^{II} with $n_1 = n_2 = 100$ is provided in Fig. 4 (left). Class X (or class 1) points are marked with circles (∘), and class Y (or class 2) points are marked with pluses (+). 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 under H_S^I and H_S^{II} are presented in Fig. 5 (the power estimates under H_S^{III} have a similar trend, hence are omitted). We also omit the power estimates of the cell-specific tests for cells (1, 2) and (2, 1), since they would be same as cells (1, 1) and (2, 2) (although cell-specific tests for cells (1, 2) and (2, 1) have high (virtually zero) power for the left-sided (right-sided) alternative, and those for cells (1, 1) and (2, 2) have high (virtually zero) power for the right-sided (left-sided) alternative). As expected, the power estimates increase as segregation gets stronger and also as sample size increases. For the cell-specific and overall tests, type I and III tests have higher power estimates.

We also estimate the power of the tests under the segregation alternatives based on the Monte Carlo critical values. For 10 000 Monte Carlo replicates, the 0.05 level Monte Carlo critical value for the right-sided alternative (i.e., for segregation) is the 950th value out of the calculated test statistics under CSR independence for each class size combination. The power estimates under $H_S^I - H_S^{III}$ have similar trends as in Fig. 5, hence are omitted. When Monte Carlo critical values are used, we get the same power estimates for Dixon's and type II cell-specific tests, and also the same power estimates for type I, III, and IV cell-specific tests. The same holds for the overall tests as well. The power estimates follow a similar trend as those based on asymptotic critical values, hence are not presented. That is, for the cell-specific and the overall tests, type I, III and IV tests have higher power estimates compared to Dixon's and type II tests.

6.2. Empirical power analysis under association of two classes

Under the association alternatives for two classes, we consider three cases also. In each case, we generate $X_i \overset{i.i.d.}{\sim} \mathcal{U}((0, 1) \times (0, 1))$ for $i = 1, 2, \ldots, n_1$. Then we generate Y_j associated with X's for $j = 1, 2, \ldots, n_2$ as follows. For each j,
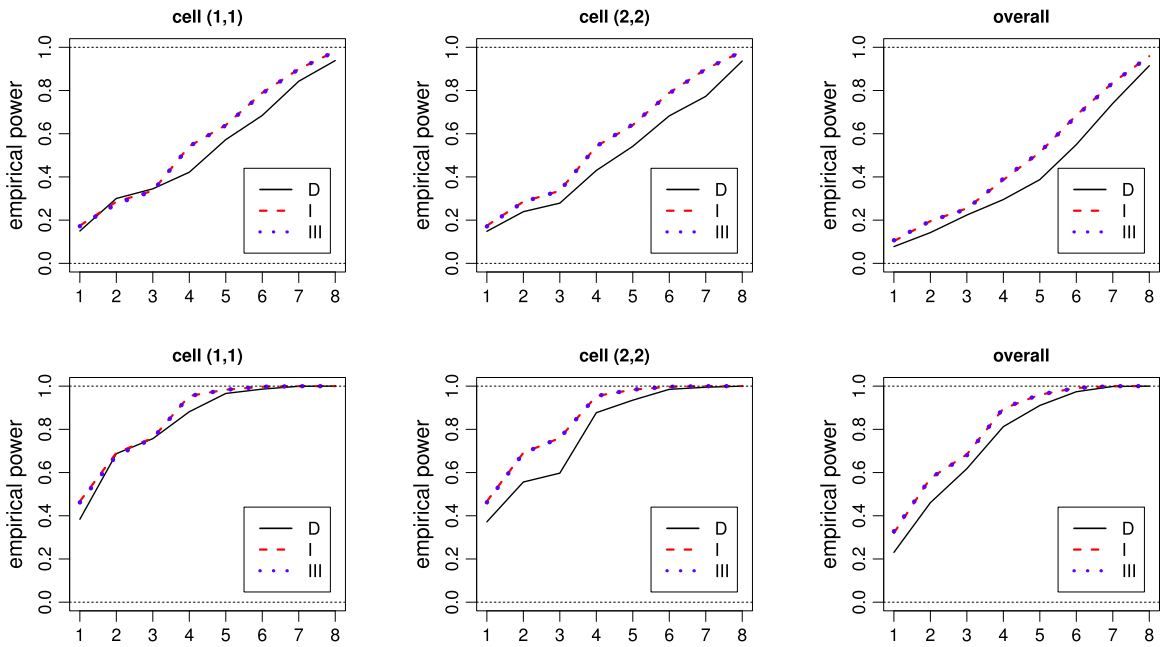
Fig. 5. The empirical power estimates based on the asymptotic critical values for the cell-specific (left and middle columns) and the overall tests (right column) under the segregation alternatives, H_S^I (top row) and H_S^{II} (bottom row), in the two-class case. The horizontal axis labels and legend labeling are as in Fig. 1.

select an i randomly, and set $Y_j = X_i + R_j \left(\cos T_j, \sin T_j\right)'$ where $R_j \overset{i.i.d.}{\sim} \mathcal{U}(0, r)$ with $r \in (0, 1)$ and $T_j \overset{i.i.d.}{\sim} \mathcal{U}(0, 2\pi)$. We consider the following association alternatives (which were also recommended in Ceyhan, 2014):

$$H_A^I : r = 1/(2\sqrt{n_t}), \qquad H_A^{II} : r = 1/(3\sqrt{n_t}), \quad \text{and} \quad H_A^{III} : r = 1/(4\sqrt{n_t}) \tag{22}$$

where $n_t = n_1 + n_2$. Notice that association gets stronger as r decreases; that is, association gets stronger from H_A^I to H_A^{III}. Furthermore, by construction, the association of Y points with X points is stronger, compared to the association of X points with Y points. These association alternatives are motivated from the expected distance between points from homogeneous Poisson Process (HPP). Letting D be the distance from a randomly chosen point to the nearest other point in a HPP with intensity ρ, we have $\mathbf{E}[D] = 1/(2\sqrt{\rho})$ and $\mathbf{Var}[D] = (4 - \pi)/(4\pi\rho)$ (Dixon, 2002b). In our case, under CSR independence, intensity of n_t points would be $\widehat{\rho} = n_t$, since area of the unit square is 1. Hence we have set $r = 1/(2\sqrt{n_t})$, $r = 1/(3\sqrt{n_t})$, and $r = 1/(4\sqrt{n_t})$ for H_A^I to H_A^{III}. For example, under H_A^I, the displacements of Y_j around X_i would be limited by the average distance between n_t points under H_o. A sample realization of H_A^{II} with $n_1 = n_2 = 100$ is provided in Fig. 4 (right).

The association parameterization in Eq. (22) is not considered in Ceyhan (2008, 2010a) where the association parameter r was fixed (and independent of the class sizes). In particular, the alternatives employed were $H_A^I : r = 1/4$, $H_A^{II} : r = 1/7$, and $H_A^{III} : r = 1/10$. The new parameterization is chosen because the previous one was confounded by the differences in the class sizes, while the current parameterization takes the class sizes into account (Ceyhan, 2014).

The empirical power estimates under association alternatives H_A^{II} are presented in Fig. 6 (the power estimates under H_A^I and H_A^{III} have a similar trend, hence omitted). As association gets stronger, the power estimates increase. However, there is a decline in power from $(n_1, n_2) = (10, 10)$ to $(10, 30)$ and $(10, 50)$, and this decline is more dramatic for Dixon's cell $(1, 1)$ test. For balanced class sizes, the power tends to increase as n_t increases. Furthermore, type I and III tests have higher power for all class size combinations for cell $(1, 1)$, and for most class size combinations for cell $(2, 2)$ and the overall test.

We also estimate power based on the Monte Carlo critical values. A 0.05 level Monte Carlo critical value for the left-sided alternative (i.e., for association) is the 50th value out of the 10 000 calculated test statistics under CSR independence for each class size combination. The power estimates under H_A^{II} are presented in Fig. 6 (the power estimates under H_A^I and H_A^{III} have a similar trend, hence omitted). The power estimates follow a similar trend as those based on asymptotic critical values. In particular, for the cell-specific tests, type I, III and IV tests have the higher power estimates compared to Dixon's and type II tests. The same holds for the overall tests for most of the class size combinations.

7. Empirical power analysis in the three-class case

To assess the empirical power performance of the tests, we also consider three cases for each of segregation and association alternatives in the three-class case.
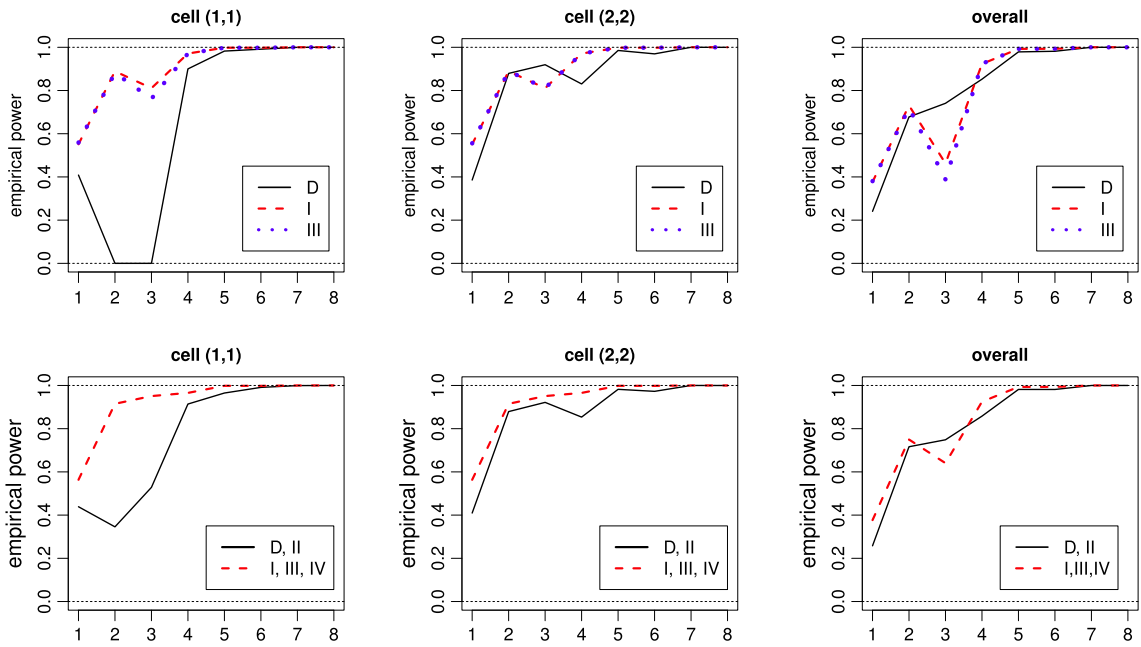
Fig. 6. The empirical power estimates for the NNCT-tests under the association alternatives, H_A^{II}, based on the asymptotic critical values (top row) and Monte Carlo critical values (bottom row), in the two-class case. The horizontal axis labels and legend labeling are as in Fig. 1.
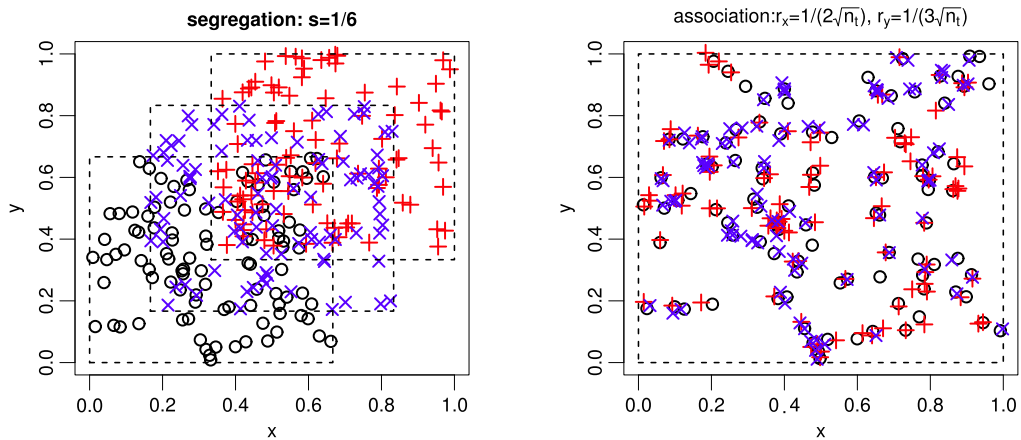

Fig. 7. A sample realization of segregation alternative H_{S_3} (left) and the association alternative H_{A_1} (right) with $n_1 = n_2 = n_3 = 100$. Class X (or class 1) points are marked with circles ($\circ$), class Y (or class 2) points are marked with pluses ($+$), class Z (or class 3) points are marked with crosses ($\times$). The supports of the classes are indicated by squares bounded by dashed lines.

7.1. Empirical power analysis under segregation of three classes

Under the segregation alternatives for three classes, we generate $X_i \overset{i.i.d.}{\sim} \mathcal{U}(S_1)$, $Y_j \overset{i.i.d.}{\sim} \mathcal{U}(S_2)$, and $Z_k \overset{i.i.d.}{\sim} \mathcal{U}(S_3)$ for $i = 1, \ldots, n_1, j = 1, \ldots, n_2$, and $k = 1, \ldots, n_3$ where $S_1 = (0, 1 - 2s) \times (0, 1 - 2s)$, $S_2 = (2s, 1) \times (2s, 1)$, and $S_3 = (s, 1 - s) \times (s, 1 - s)$ with $s \in (0, 1/2)$. We consider the following segregation alternatives:

$$H_{S_1} : s = 1/12, \qquad H_{S_2} : s = 1/8, \quad \text{and} \quad H_{S_3} : s = 1/6. \tag{23}$$

Notice that, as s increases, segregation between the classes gets stronger; that is, segregation gets stronger from H_{S_1} to H_{S_3}. Furthermore, by construction classes X and Y are more segregated compared to Z and X or Z and Y. In fact, the segregation between X and Z and segregation between Y and Z are identical (as a stochastic process). A sample realization of H_{S_3} with $n_1 = n_2 = n_3 = 100$ is provided in Fig. 7 (left).

Empirical power estimates for the two-sided alternatives for the diagonal cells $(1, 1)$, $(2, 2)$, and $(3, 3)$ under segregation alternatives H_{S_2} and H_{S_3} and those for the off-diagonal cells $(1, 2)$, $(1, 3)$, and $(2, 3)$ under H_{S_2} are plotted in Fig. 8 (the power estimates under H_{S_1} have a similar trend, hence omitted). For diagonal cells $(1, 1)$ and $(2, 2)$ type I and III tests have higher
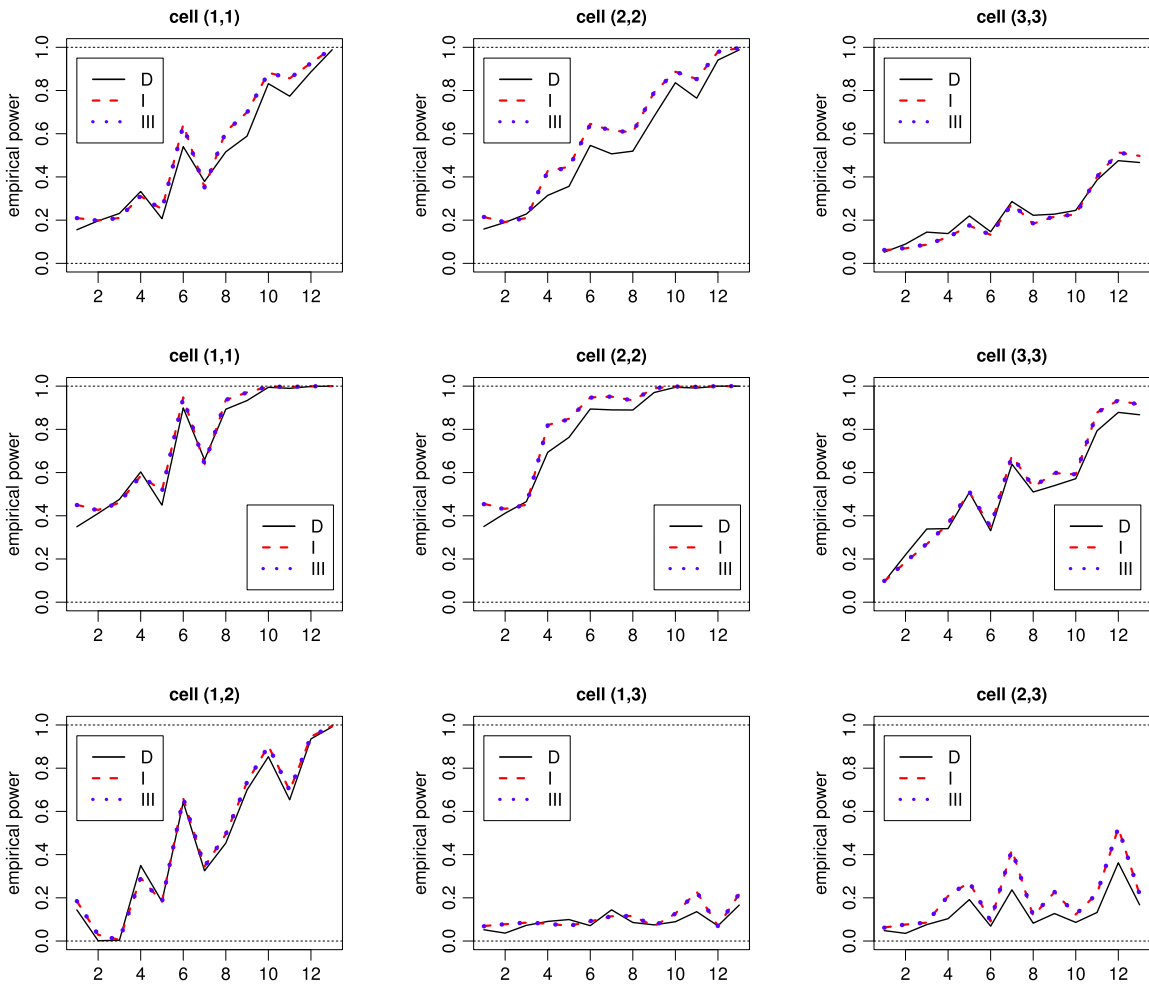
Fig. 8. The empirical power estimates of the cell-specific tests for cells (1, 1), (2, 2), and (3, 3) under the segregation alternatives H_{S_2} (top row) and H_{S_3} (middle row) and empirical power estimates of the cell-specific tests for cells (1, 2), (1, 3), and (2, 3) under the segregation alternative H_{S_2} (bottom row) in the three-class case. The legend labeling is as in Fig. 1 and horizontal axis labels are as in Fig. 2.
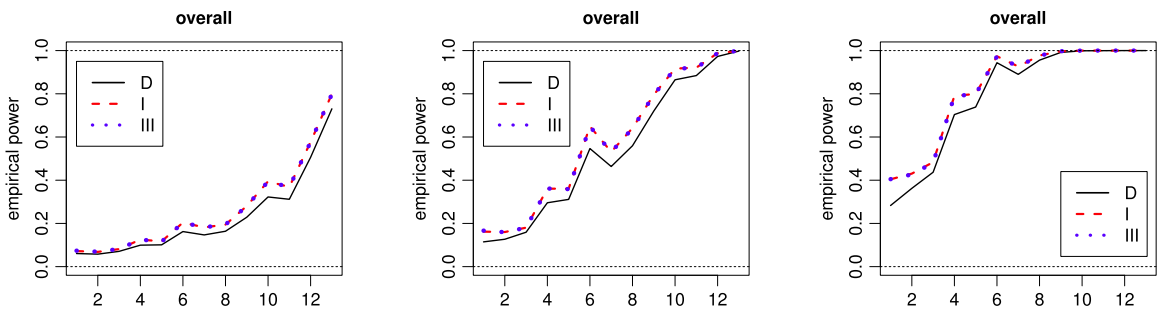

Fig. 9. The empirical power estimates of the overall tests under the segregation alternatives H_{S_1} (left), H_{S_2} (middle), and H_{S_3} (right) in the three-class case. The legend labeling is as in Fig. 1 and horizontal axis labels are as in Fig. 2.

power, while for diagonal cell (3, 3), all tests have similar power estimates. For the off-diagonal cells (1, 2) and (1, 3) all tests have similar power estimates (although type I and III tests have slightly higher power), while for cell (2, 3) type I and III tests have higher power. In line with our simulation setup, power estimates for cells (1, 1) and (2, 2) are higher compared to cell (3, 3), as classes X and Y are more segregated compared to class Z. For the same reason, power estimates for cell (1, 2) is higher compared to cells (1, 3) and (2, 3).

Empirical power estimates for the overall tests are presented in Fig. 9. Type I and III tests have higher power compared to Dixon's test.
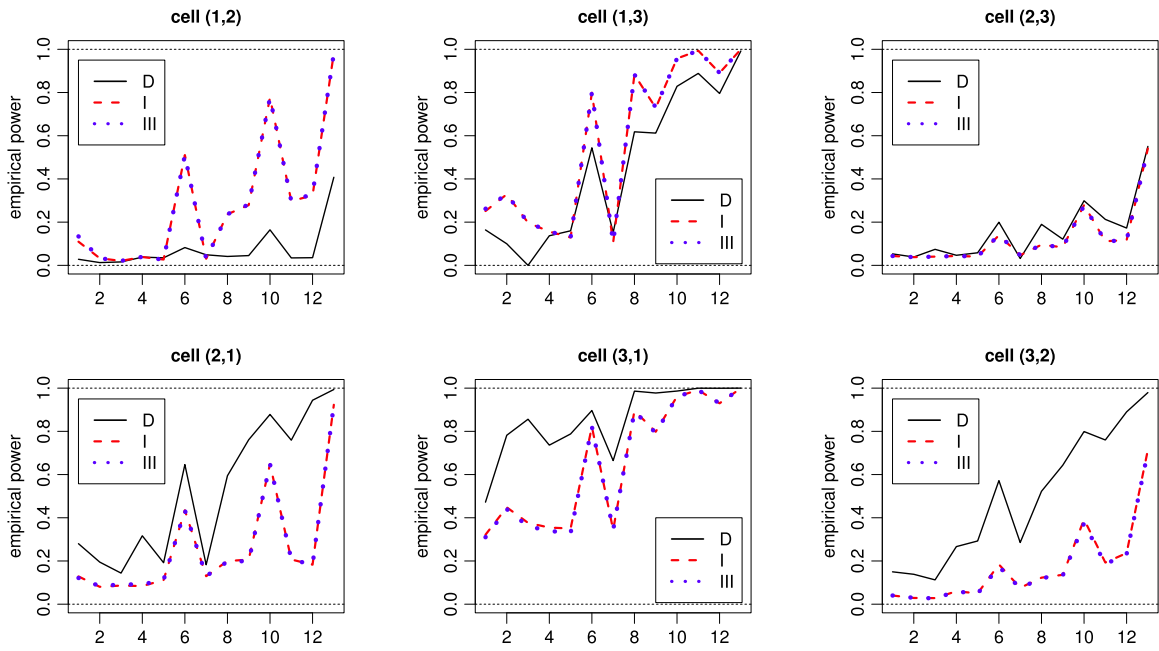
Fig. 10. The empirical power estimates of the cell-specific tests for cells (1, 2), (1, 3) and (2, 3) (top row) and (2, 1), (3, 1) and (3, 2) (bottom row) under the association alternative H_{A_1} in the three-class case. The legend labeling is as in Fig. 1 and horizontal axis labels are as in Fig. 2.
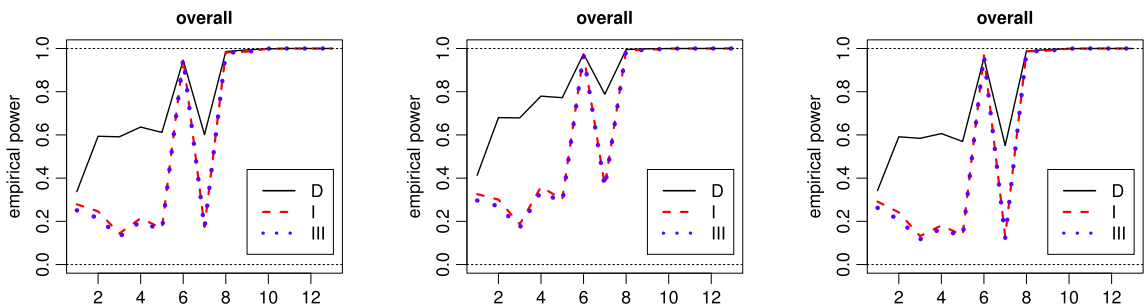

Fig. 11. The empirical power estimates of the overall tests under the association alternatives H_{A_1} (left), H_{A_2} (middle), and H_{A_3} (right) in the three-class case. The legend labeling is as in Fig. 1 and horizontal axis labels are as in Fig. 2.

7.2. Empirical power analysis under association of three classes

Under the association alternatives for three classes, we also consider three cases. We generate $X_i \overset{i.i.d.}{\sim} \mathcal{U}((0, 1) \times (0, 1))$ for $i = 1, 2, \ldots, n_1$. Then we generate Y_j and Z_k for $j = 1, 2, \ldots, n_2$ and $k = 1, 2, \ldots, n_3$ as follows. For each j, select an i randomly, and set $Y_j := X_i + R_j^Y (\cos T_j, \sin T_j)'$ where $R_j^Y \overset{i.i.d.}{\sim} \mathcal{U}(0, r_y)$ with $r_y \in (0, 1)$ and $T_j \overset{i.i.d.}{\sim} \mathcal{U}(0, 2\pi)$. Similarly, for each k, select an i' randomly, and set $Z_k := X_{i'} + R_k^Z (\cos U_\ell, \sin U_\ell)'$ where $R_k^Z \overset{i.i.d.}{\sim} \mathcal{U}(0, r_z)$ with $r_z \in (0, 1)$ and $U_k \overset{i.i.d.}{\sim} \mathcal{U}(0, 2\pi)$. We consider the following association alternatives:

$$H_{A_1} : r_y = 1/(2\sqrt{n_t}), \qquad r_z = 1/(3\sqrt{n_t}), \qquad H_{A_2} : r_y = 1/(2\sqrt{n_t}), \qquad r_z = 1/(4\sqrt{n_t}),$$

$$\text{and} \quad H_{A_3} : r_y = 1/(3\sqrt{n_t}), \ r_z = 1/(4\sqrt{n_t}) \tag{24}$$

where $n_t = n_1 + n_2 + n_3$. As r_y and r_z decrease, the level of association increases. That is, the association between X and Y and association between X and Z get stronger from H_{A_1} to H_{A_3}. By construction, classes Y and Z are associated with class X, while classes Y and Z are not associated, but perhaps mildly segregated for small r_y and r_z. Furthermore, by construction, classes X and Z are more associated compared to classes X and Y. A sample realization of H_{A_1} with $n_1 = n_2 = n_3 = 100$ is provided in Fig. 7 (right).

The empirical power estimates under H_{A_1} for cells (1, 2), (2, 1), (1, 3), (3, 1), (2, 3) and (3, 2) are presented in Fig. 10. For cells (1, 2) and (1, 3), type I and III cell-specific tests have higher power, while for cells (2, 1), (3, 1), (2, 3) and (3, 2), Dixon's cell-specific test has higher power. The power estimates for the overall tests are presented in Fig. 11. For the overall tests, Dixon's test has higher power estimates.
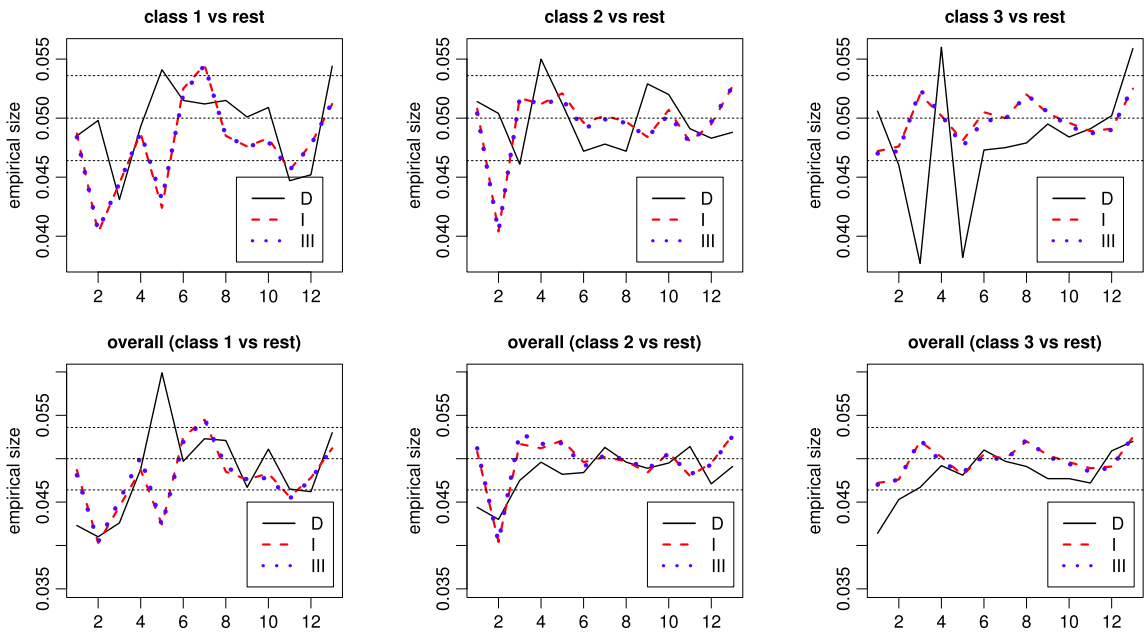
Fig. 12. The empirical size estimates of the cell-specific tests for cell (2, 2) and overall tests under CSR independence with **one-vs-rest type testing** in the three-class case. The legend labeling is as in Fig. 1 and horizontal axis labels are as in Fig. 2.

8. Empirical size and power analysis for the one-vs-rest type tests in the three class case

In one-versus-rest type testing, we implement Monte Carlo simulations as in Section 5.1 to assess the empirical size performance of these tests under CSR independence. We present the empirical size estimates for various class size combinations in Fig. 12 where only cell-specific tests for cell (2, 2) and the overall test are presented, since for cell (1, 1) the cell-specific test is the same as the cell-specific test in the 3×3 NNCT analysis in Sections 5.1 and 7. Among cell-specific tests, type I and type III tests perform better compared to Dixon's test, since they are closer to the nominal level especially for large classes. For the overall tests, the tests are about the nominal level with type I and III tests being slightly closer than Dixon's test.

To evaluate the power performance of these tests, we perform simulations under segregation alternatives as in Section 7.1. The empirical power estimates under the segregation alternative H_{S_2} are presented in Fig. 13. Among the tests, type I and III tests have higher power estimates compared to Dixon's test. One class-vs-rest tests for classes 1 and 2 have higher power estimates compared to that of class 3. This occurs, since by construction, classes 1 and 2 are equally segregated from other classes, and these classes are more segregated compared to class 3.

For the association alternatives, we perform the simulations as in Section 7.2. The corresponding power estimates under the association alternative H_{A_2} are presented in Fig. 14. For the one-vs-rest cell-specific tests, Dixon's test has higher power for class 1-vs-rest and 2-vs-rest tests, and types I and III have higher power for class 3-vs-rest test. For the overall one-vs-rest tests, Dixon's test has higher power for classes 1 and 2, and for class 3, all tests have similar power estimates.

9. Size and power performance of cell-specific and one-vs-rest as post-hoc tests

Although in practice, we recommended the use of cell-specific and one-vs-rest type tests as post-hoc tests (after a significant overall test), we have assessed their size and power performance unconditionally (i.e., not after a significant overall test). If we perform these tests after a significant overall test, then its size and power performance would change.

When we perform the overall test first, and if it is significant, perform the corresponding cell-specific or one-vs-rest type tests, the tests tend to become significantly conservative. For example, the size estimates of the cell-specific tests as post-hoc tests in the two and three class cases for the diagonal cells are provided in Fig. 15. Comparing these size estimates with those in Figs. 1 and 2, we observe that Dixon's test becomes extremely conservative in both two and three class cases when used as a post-hoc test. However, type I and type III tests are not affected in the two class case, but become significantly conservative in the three class case when used as post-hoc tests. In both two and three class cases, type I and type III cell-specific tests are closer to the nominal level compared to Dixon's test.

When conducting one-vs-rest tests as post-hoc tests, there are two types of tests: one is the cell-specific test for cell (2, 2) in the new 2×2 NNCT, and the other is the overall test in the new NNCT. Notice that in one-vs-rest setting, after a significant overall test for the $m \times m$ NNCT, there is also an overall test for the 2×2 NNCT as a post-hoc test. When one-vs-rest tests are used as post-hoc tests, they become extremely conservative as well. For example, the size estimates of the one-vs-rest
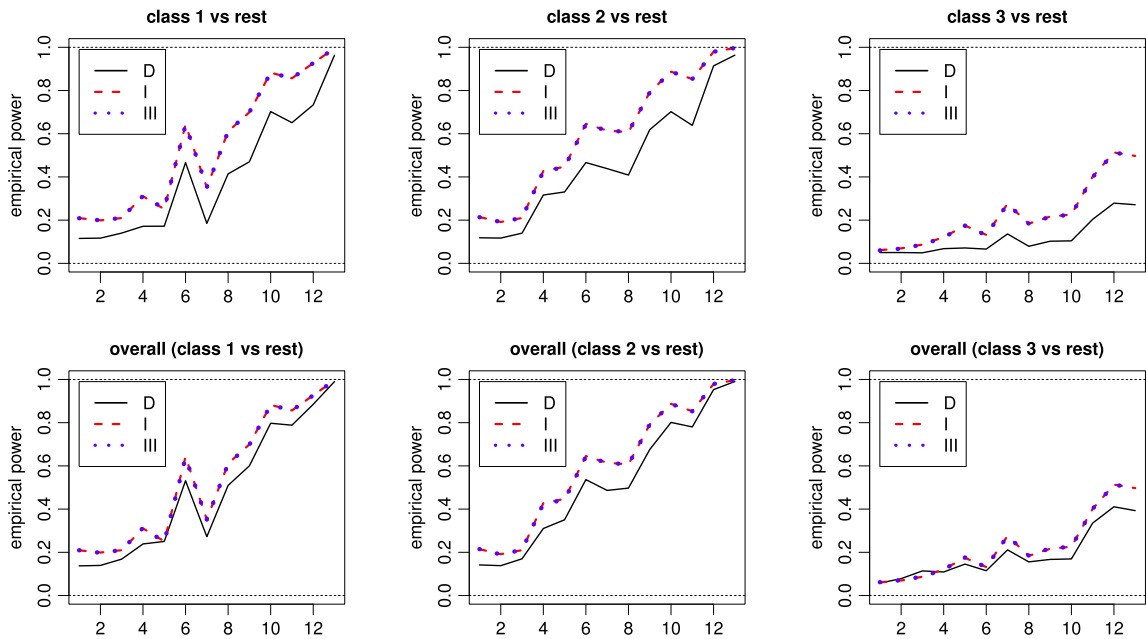
Fig. 13. The empirical power estimates of the cell-specific tests for cell $(2, 2)$ (top), and the overall tests (bottom) under the segregation alternative H_{S_2} in the three-class case with **the one-versus-rest type testing**. The legend labeling is as in Fig. 1 and horizontal axis labels are as in Fig. 2.
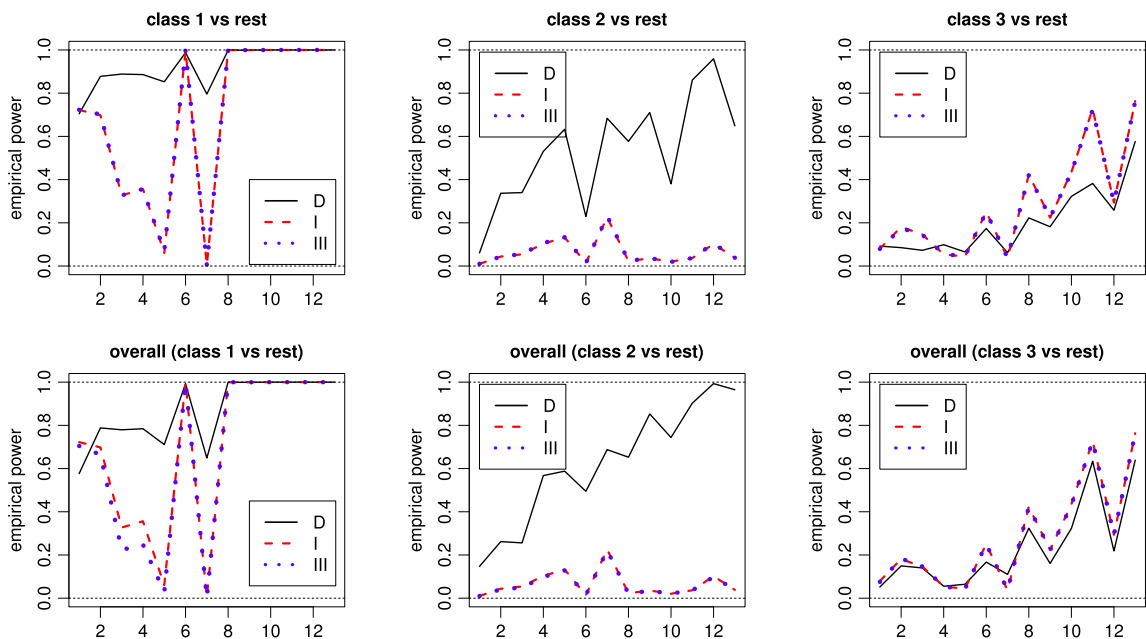

Fig. 14. The empirical power estimates the cell-specific tests for cell $(2, 2)$ (top) and the overall tests (bottom) under the association alternative H_{A_2} in the three-class case with the **one-vs-rest type testing**. The legend labeling is as in Fig. 1 and horizontal axis labels are as in Fig. 2.

tests as post-hoc tests in the three class case are provided in Fig. 16. For the cell-specific tests as class i-vs-rest, type I and type III tests are closer to the nominal level, while for the overall test as post-hoc test, Dixon's test is closer to the nominal level.

Just like the size estimates of the cell-specific and one-vs-rest tests decrease when used as post-hoc tests, the corresponding power estimates under the segregation and association alternatives tend to decrease as well. However, the overall trend and the order of power estimates do not change compared to the use of these tests as stand-alone tests. For example, in a two class setting, we present the power estimates of the cell-specific tests as post-hoc tests for cells $(1, 1)$ and $(2, 2)$ under the segregation alternative H_S^I and under the association alternative H_A^{II} in Fig. 17. Comparing these power
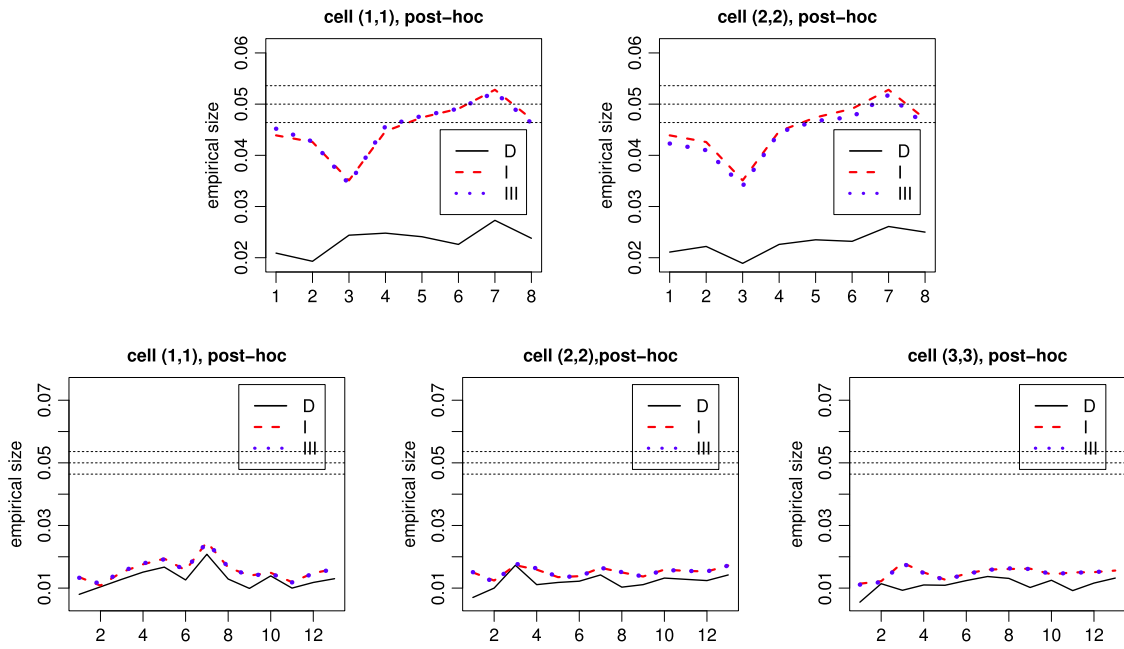
Fig. 15. The empirical size estimates of the cell-specific tests as post hoc tests for the diagonal cells in the two-class case (top row) and in the three class case (bottom row) under the **CSR independence pattern in the two-class case**. The horizontal lines, axis and legend labeling are as in Fig. 1 in the two class case, and are as in Fig. 2 in the three class case.
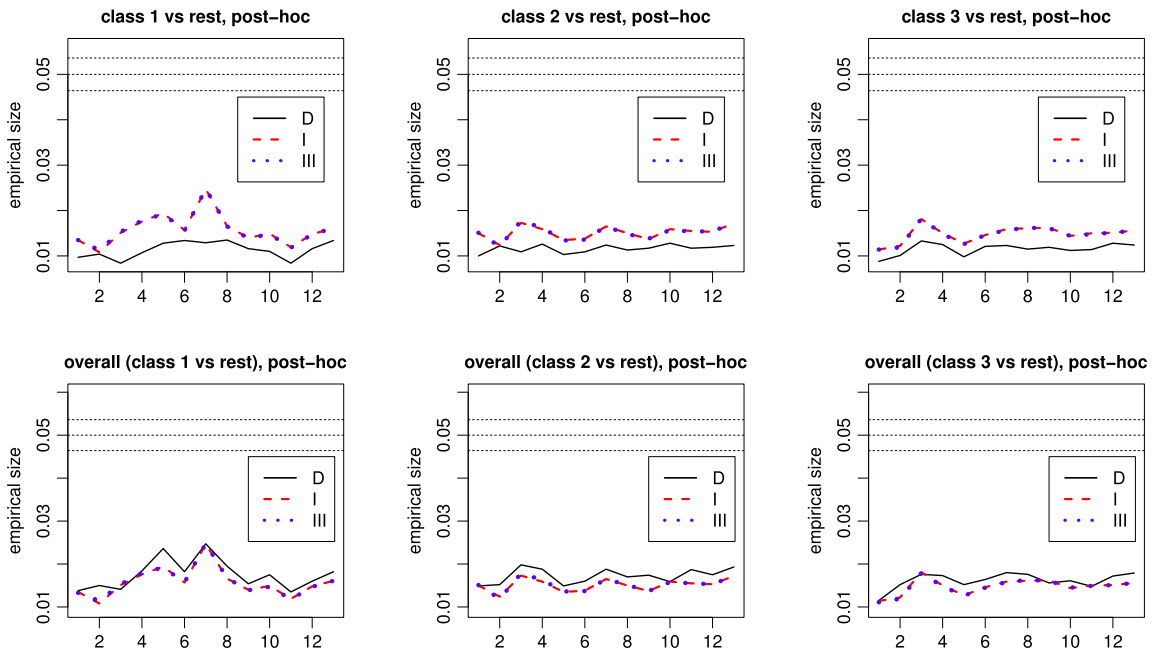

Fig. 16. The empirical size estimates of the cell-specific tests for cell (2, 2) and overall tests under CSR independence with **one-vs-rest type testing** as post-hoc tests in the three-class case. The legend labeling is as in Fig. 1 and horizontal axis labels are as in Fig. 2.

estimates with the ones in Figs. 5 and 6, we see that the power estimates exhibit a downward shift in Fig. 17 when the tests are used in a post-hoc fashion at each sample size combination, preserving the trend and order of the power estimates.

10. Example data: swamp tree data

The NNCT methodology is illustrated on an ecological data set: the swamp tree data of Good and Whipple (1982) which was also analyzed by Dixon (1994, 2002a). Briefly, the plot contains 13 different tree species, of which four species account
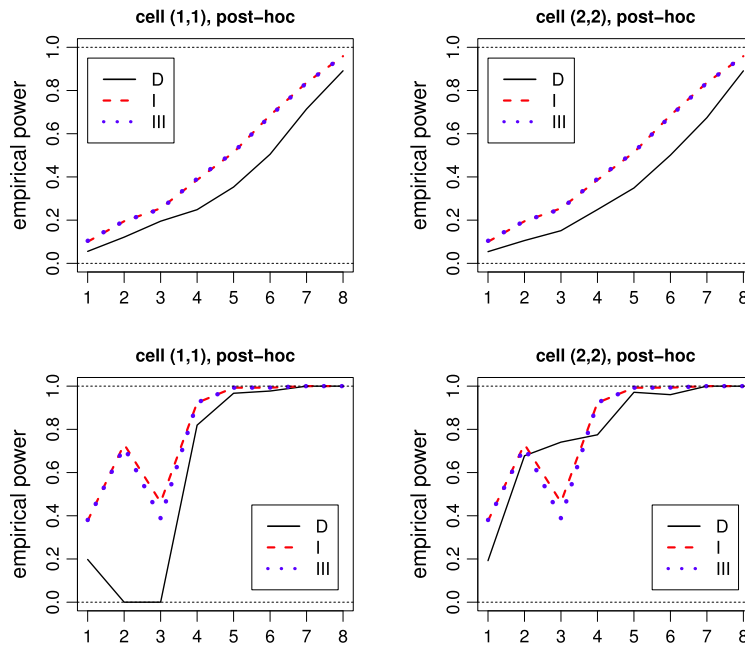
Fig. 17. The empirical power estimates based on the asymptotic critical values for the cell-specific tests as post-hoc tests under the segregation alternative, H_S^I (top row) and under the association alternative H_A^{II} (bottom row), in the two-class case. The horizontal axis labels and legend labeling are as in Fig. 1.
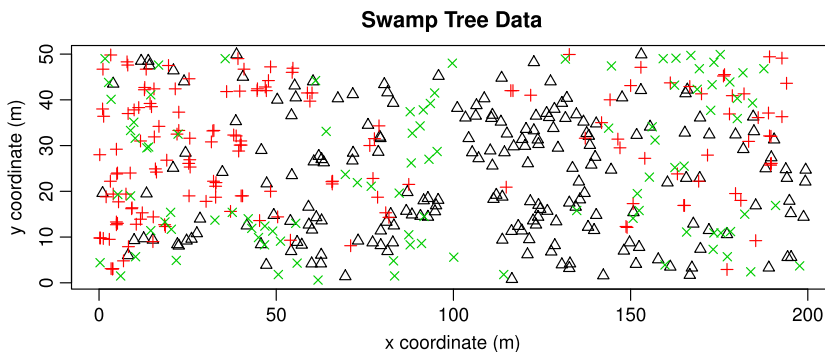

Fig. 18. The scatter plot of the locations of black gum trees (triangles △), Carolina ashes (pluses +), bald cypress trees (crosses ×).

for over 90% of the 734 tree stems. In our analysis, we only consider black gums (*Nyssa sylvatica*), Carolina ashes (*Fraxinus caroliniana*), and bald cypresses (*Taxodium distichum*) as if only these three tree species exist in the area, so we are ignoring the possible effects of other species on the spatial interaction between these species for illustrative purposes. This data was also analyzed before but with other species combinations (Ceyhan, 2008, 2010a). Thus, we perform a 3×3 NNCT-analysis on this data set. See Fig. 18 for the location of the trees in this plot and Table 3 for the associated 3×3 NNCT together with cell percentages based on the base class sizes, and marginal percentages based on the grand sum, n. When, e.g., black gum is the base species and Carolina ash is the NN species, the cell count is 40 which is 20% of the black gums (and Carolina ashes are 34% of all trees). The percentages in Table 3 and Fig. 18 suggest that each tree species is segregated from the other trees as the observed percentages of species in the diagonal cells are much larger than the row percentages (or species percentages).

In the swamp tree data, the locations of the tree species can be viewed a priori resulting from different processes, so the more appropriate null hypothesis would be the CSR independence pattern (Goreaud & Pélissier, 2003). We compute $Q = 282$ and $R = 288$ for this data set and our inference will be conditional on these values. Dixon's and the new overall segregation tests and the associated p-values are presented in Table 4, where p_{asy} stands for the p-value based on the asymptotic approximation, p_{mc} is the p-value based on 10 000 Monte Carlo replications of the CSR independence pattern 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 all significant. The cell-specific test statistics and the associated p-values are presented in Table 5, where p-values are calculated as in Table 4. Again, all three p-values in Table 5 are similar for each cell-specific test.

The overall segregation tests are all highly significant which implies that there is significant deviation from the CSR independence pattern for at least one of the tree species. To determine which species exhibit segregation or association, we

Table 3

The NNCT for swamp tree data and the corresponding percentages and $\widehat{\pi}_{ij} = N_{ij}/n$ values (in parentheses), where the cell percentages are with respect to the size of the base species (i.e., row sums), and marginal percentages are with respect to the total size, n. B.G. = black gums, C.A. = Carolina ashes, and B.C. = bald cypresses.

		NN			sum
		B.G.	C.A.	B.C.	
Base	B.G.	142 (69%, 0.31)	40 (20%, 0.09)	23 (11%, 0.05)	205 (45%)
	C.A.	34 (22%, 0.07)	97 (62%, 0.21)	25 (16%, 0.05)	156 (34%)
	B.C.	38 (39%, 0.08)	32 (33%, 0.07)	28 (29%, 0.06)	98 (21%)
	sum	214 (47%)	169 (37%)	76 (17%)	459 (100%)

Table 4

Test statistics and p-values for the overall tests and the corresponding p-values. 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. $\mathcal{X}_D$ stands for Dixon's overall test, $\mathcal{X}_I$ and $\mathcal{X}_{III}$ are for type I and type III overall tests, respectively.

Overall tests			
	$\mathcal{X}_D$ 75.78	$\mathcal{X}_I$ 65.35	$\mathcal{X}_{III}$ 65.39
p_{asy}	<0.0001	<0.0001	<0.0001
p_{mc}	<0.0001	<0.0001	<0.0001
p_{rand}	<0.0001	<0.0001	<0.0001

Table 5

Test statistics and p-values for the cell-specific tests and the corresponding p-values (in parentheses). The p-values are given in the order of p_{asy}, p_{mc}, and p_{rand}, whose labeling is as in Table 4. B.G. = black gums, C.A. = Carolina ashes, and B.C. = bald cypresses.

	B.G.	C.A.	B.C.
Dixon's cell-specific tests			
B.G.	6.57 (<0.0001, <0.0001, <0.0001,)	−4.46 (<0.0001, <0.0001, <0.0001,)	−3.74 (0.0002, <0.0001, 0.0003)
C.A.	−5.65 (<0.0001, <0.0001, <0.0001)	6.60 (<0.0001, <0.0001, <0.0001)	−1.70 (0.0893, 0.0918, 0.1032)
B.C.	−1.18 (0.2395, 0.2470, 0.2596)	−0.30 (0.7672, 0.7796, 0.8140)	1.51 (0.1320, 0.1345, 0.1445)
Type I cell-specific tests			
B.G.	6.91 (<0.0001, <0.0001, <0.0001)	−6.29 (<0.0001, <0.0001, <0.0001)	−2.37 (0.0177, 0.0170, 0.0176)
C.A.	−6.86 (<0.0001, <0.0001, <0.0001)	6.49 (<0.0001, <0.0001, <0.0001)	−0.21 (0.8352, 0.8439, 0.8408)
B.C.	−1.67 (0.0944, 0.0954, 0.0900)	−0.96 (0.3382, 0.3407, 0.3433)	2.61 (0.0091, 0.0087, 0.0081)
Type III cell-specific tests			
B.G.	6.91 (<0.0001, <0.0001, <0.0001)	−6.29 (<0.0001, <0.0001, <0.0001)	−2.37 (0.0180, 0.0172, 0.0179)
C.A.	−6.86 (<0.0001, <0.0001, <0.0001)	6.49 (<0.0001, <0.0001, <0.0001)	−0.20 (0.8381, 0.8455, 0.8436)
B.C.	−1.67 (0.0943, 0.0953, 0.0898)	−0.96 (0.3375, 0.3401, 0.3426)	2.60 (0.0094, 0.0088, 0.0084)

perform the cell-specific tests and one-vs-rest type tests for post-hoc analysis. At 0.05 level, Dixon's and the new cell-specific tests agree for all cells in term of significance except for (B.C., B.C.) cell, at which Dixon's test is not significant but types I and III are significant. At 0.10 level tests agree for cells except (B.C., B.G) and (C.A., B.C.), at cell (B.C., B.G) Dixon's test is not significant but types I and III are significant, while at cell (C.A., B.C.) Dixon's test is significant but types I and III are not. At 0.01 level tests agree at cells except for cell (B.G., B.C.) at which Dixon's test is significant while types I and III are not. The test statistics are all positive (negative) for the diagonal (off-diagonal) cells which also support the segregation of species.

For a given class i, we estimate probabilities π_{ij} of Section 2.1 as $\widehat{\pi}_{ij} = N_{ij}/n$. The estimated probabilities are presented in parentheses as decimals in Table 3. For example, for (B.G., C.A.) cell, $\widehat{\pi}_{12} = N_{12}/n = 40/459 \approx 0.09$. For black gums, we have $\widehat{\pi}_{11} = 0.31 > \widehat{\pi}_{12} + \widehat{\pi}_{13} = 0.09 + 0.05 = 0.14$, so black gums exhibit total segregation from the other two tree species. Similarly, for Carolina ashes, we have $\widehat{\pi}_{22} = 0.21 > \widehat{\pi}_{21} + \widehat{\pi}_{23} = 0.07 + 0.05 = 0.12$, so Carolina ashes exhibit total segregation from the other two tree species. However, bald cypresses exhibit neither strong nor total segregation, since $\widehat{\pi}_{33} = 0.06 < \widehat{\pi}_{31} = 0.08$ and $\widehat{\pi}_{33} < \widehat{\pi}_{32} = 0.07$. Furthermore, black gums seem to be strongly associated with bald cypresses as $\widehat{\pi}_{31} = 0.08 > \widehat{\pi}_{32} = 0.07$ and $\widehat{\pi}_{31} > \widehat{\pi}_{33} = 0.06$.

We present the one-vs-rest cell-specific and overall tests (see Table 6). For each species, we observe that the other species combined tend to be segregated from the species in consideration, but to a lesser extent for bald cypresses.

The spatial interaction is significant for each species, but at different levels. In particular, black gums exhibit significant segregation from other species (they are significantly segregated from both Carolina ashes and bald cypresses), Carolina ashes exhibit significant segregation from other species (they are significantly segregated from black gums but not from

Table 6

Test statistics and p-values for one-vs-rest cell-specific tests for cell $(2, 2)$ and one-vs-rest overall tests. The corresponding p-values are presented in parentheses. Z_{22}^D stands for Dixon's cell-specific test, and Z_{22}^I and Z_{22}^{III} stand for type I and III cell-specific tests. $\mathcal{X}_D$ stands for Dixon's overall test and $\mathcal{X}_I$ and $\mathcal{X}_{III}$ are for type I and III overall tests.

One-vs-rest cell-specific tests			
	Z_{22}^D	Z_{22}^I	Z_{22}^{III}
B.G.-vs-rest	5.09 (<0.0001)	6.91 (<0.0001)	6.91 (<0.0001)
C.A.-vs-rest	3.86 (0.0001)	6.49 (<0.0001)	6.49 (<0.0001)
B.C.-vs-rest	4.12 (<0.0001)	2.61 (0.0046)	2.61 (0.0045)

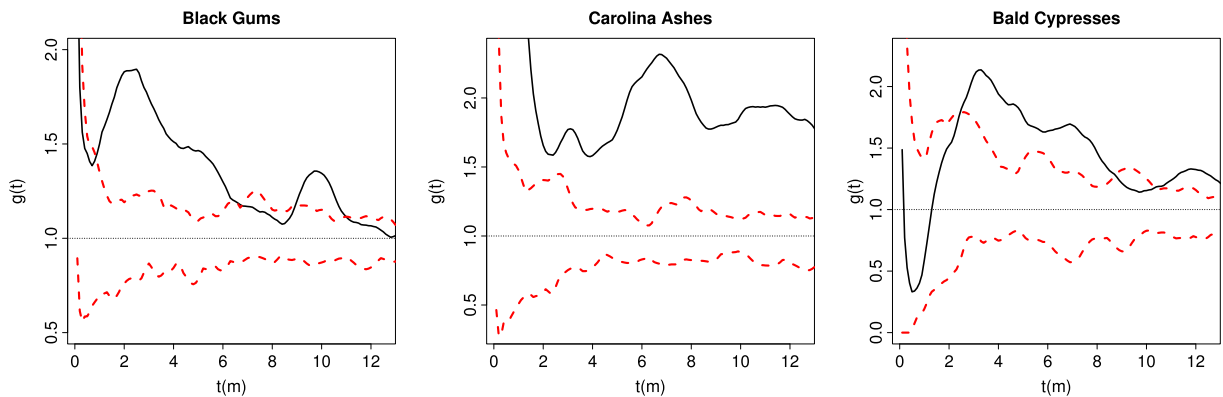| One-vs-rest overall tests			
	$\mathcal{X}_D$	$\mathcal{X}_I$	$\mathcal{X}_{III}$
B.G.-vs-rest	48.86 (<0.0001)	47.70 (<0.0001)	47.72 (<0.0001)
C.A.-vs-rest	44.79 (<0.0001)	42.11 (<0.0001)	42.15 (<0.0001)
B.C.-vs-rest	16.96 (0.0002)	6.79 (0.0091)	6.75 (0.0094)

Fig. 19. Pair correlation functions for each 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 pair correlation functions based on Monte Carlo simulation under the CSR independence pattern.

bald cypresses), and Bald cypresses exhibit significant segregation from other species (they are moderately segregated from black gums only but when the two species of black gums and Carolina ashes are considered together, the (B.C., B.C.) cell is significant).

However, these results pertain to interaction at about the average NN distances. For the swamp tree data average NN distance ($\pm$ standard deviation) is about 2.1 (± 1.35) meters. We might also be interested in the possible causes of the segregation and the type and level of interaction between the tree species at different distances between the trees. Along this line, we also present the second-order analysis of the swamp tree data by the pair correlation function $g(t)$ (Stoyan & Stoyan, 1994). The pair correlation function of a (one-class) stationary point process is defined as $g(t) = \frac{K'(t)}{2\pi t}$ where $K'(t)$ is the derivative of Ripley's $K(t)$ function. For a one-class stationary Poisson process, $g(t) = 1$; values of $g(t) > 1$ suggest clustering (or aggregation) and the values of $g(t) < 1$ suggest inhibition (or regularity) between points. The pair correlation functions for each species are plotted in Fig. 19. Black gums are aggregated for distance values of about 1–6 and 9–11 m; Carolina ashes are aggregated for all the range of the plotted distances; and bald cypresses are aggregated for distance values of about 2–8 and around 11 m. These distance ranges at which species are aggregated include the mean NN distance for our data, hence this aggregation could be the reason of the significant segregation between the species.

The same definition of the pair correlation function can be applied to Ripley's bivariate (i.e., two-class) K or L-functions. Under CSR independence we have $g(t) = 1$; $g(t) > 1$ suggests association of the classes; and $g(t) < 1$ suggests segregation of the classes. The bivariate pair correlation functions for the species in swamp tree data are plotted in Fig. 20. Black gums and Carolina ashes are segregated for about 2–2.5, 3.5–4.5, 7.5–8.5, and 10.5–12 m; black gums and bald cypresses are segregated for about 2.5, 3, and 6 m; and Carolina ashes and bald cypresses are associated for 7 and 9 m.

The pair correlation function estimates have considerably high variability for small t if $g(t) > 0$, hence are not so reliable for small distances (Stoyan & Stoyan, 1996). See for example Figs. 19 and 20 where the confidence bands for small t values are much wider compared to those for larger t values. So pair correlation function analysis is more reliable for larger distances, say, larger than about the average NN distance in the data set. While the pair correlation function provides information on the one-class and bivariate patterns at all distances, NNCT-tests summarize the spatial interaction for distances about the average NN distance in the data set.
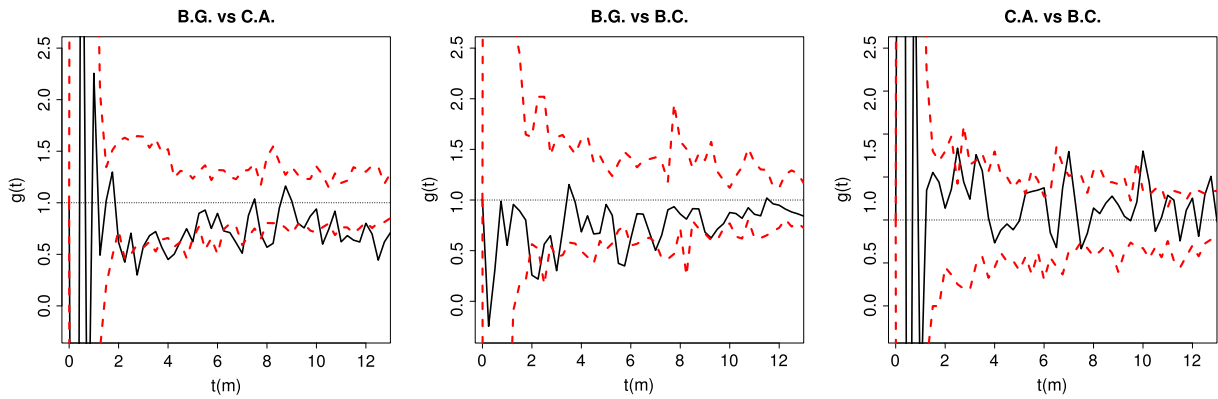
Fig. 20. Pair correlation functions for each pair of species in the swamp tree data. Wide dashed lines around 1 (which is the theoretical value) are the upper and lower (pointwise) 95% confidence bounds for the pair correlation functions based on Monte Carlo simulations under the CSR independence pattern. B.G. = black gums, C.A. = Carolina ashes, and B.C. = bald cypresses.

Table 7
The summary table for the NNCT-test statistics for m classes. The common assumption for all these tests is that the NNCT is constructed based on completely mapped data. Furthermore, all tests are conditional on Q and R under CSR independence, and unconditional under RL.

Test	Asymptotic distribution for the tests				Depends on
	Usual ($m > 1$)		One-vs-rest[a] ($m > 2$)		
	Cell-specific[a,b]	Overall	Cell-specific	Overall	
Dixon's	Normal	$\chi^2_{m(m-1)}$	Normal	χ^2_2	Row sums[c]
Type I	Normal	$\chi^2_{(m-1)^2}$	Normal	χ^2_1	Row & column sums
Type II	Normal	$\chi^2_{m(m-1)}$	Normal	χ^2_2	Row sums
Type III	Normal	$\chi^2_{(m-1)^2}$	Normal	χ^2_1	Row & column sums
Type IV	Normal	$\chi^2_{(m-1)^2}$	Normal	χ^2_1	Row & column sums

[a] The cell-specific tests and one-vs-rest tests can be used as stand-alone tests (unconditionally), or as post-hoc tests after a significant overall test.
[b] Asymptotic normality is proved for all cells in the two-class case, but for $m > 2$, it is proved only for the diagonal cells, and established by extensive Monte Carlo simulations for the off-diagonal cells.
[c] Row sums (of the NNCT) are equal to the class sizes for all tests.

11. Discussion and conclusions

We survey the cell-specific and overall segregation tests based on nearest neighbor contingency tables (NNCTs) that exist in literature, introduce new ones and discuss their properties. NNCT-tests are used in testing randomness in the nearest neighbor (NN) structure between two or more classes with NN probabilities being proportional to the class frequencies. The overall test is used for testing any deviation from the null pattern in all of the NNCT cells combined; cell-specific test for cell (i, j) is used for testing any deviation from the null case in cell (i, j), i.e., the probability of a (base,NN) pair in which base class is i and NN class is j being proportional to the product of frequencies of classes i and j. This statistic tests the segregation or lack of it, if $i = j$; the association or lack of it between classes i and j, if $i \neq j$. A summary of the properties of the NNCT-tests is provided in Table 7. Among many possible patterns, the null pattern is implied by the RL or CSR independence patterns. However, under the CSR independence pattern, NNCT-tests are conditional on Q and R, while under the RL pattern, these tests are unconditional (see Section 3.7 part (iii)).

In the computation of the overall tests, we are employing a generalized inverse procedure on the covariance matrices which are rank-deficient. Hence the resulting overall tests are confounded and may be severely distorted by the poor performance of the generalized inverse procedure on these rank-deficient covariance matrices. To overcome this problem, we also propose a rank-adjustment for the overall tests that makes the generalized inverse procedure yield the correct (generalized) inverse and hence stabilizes the behavior and distribution of the overall tests. Although we consider five types of cell-specific and overall tests, we demonstrate that essentially, these tests yield three distinct types of cell-specific and overall tests. More specifically, Dixon's tests and type II tests are identical, and so are type III and type IV tests. Hence in our empirical size and power analysis (as well as in the analysis of example data), we only use and present Dixon's, type I and type III test statistics. In the two-class case, cell-specific tests are essentially different only for at most two cells, since cells $(1, 1)$ and $(1, 2)$ yield the same test statistic in absolute value for Dixon's cell-specific test, likewise for cells $(2, 1)$ and $(2, 2)$. Similarly, cells $(1, 1)$ and $(2, 1)$ yield the same test statistic in absolute value for type III cell-specific test, likewise for cells $(1, 2)$ and $(2, 2)$. For type I cell-specific test, cells $(1, 1)$ and $(2, 2)$ yield the same test statistic value, and the off-diagonal cells give the negative of this value.

We observe that the cell-specific tests tend to standard normal distribution, as the class sizes get larger. Although this is only proved for all cells in the two-class case and for the diagonal cells for $m > 2$ classes, there is considerable progress for proving the asymptotic normality of off-diagonal cells in author's ongoing research using the tools in Penrose (2003) under CSR independence and those in Cuzick and Edwards (1990) under RL. However, the proof of the asymptotic normality of the off-diagonal cells in the case of three or more classes under RL is still an open problem, although simulations provide substantial evidence in favor of asymptotic normality even at moderate sample sizes. As a result, as a safety check, we recommend the use of asymptotic version and Monte Carlo randomized version of the tests for the off-diagonal cells and the corresponding overall test for three or more classes. See also Remark (ii) in Section 3.7 On the other hand, the overall tests tend to chi-square distribution with the corresponding degrees of freedom with the increasing class sizes. In terms of the asymptotic distribution of the overall tests, we have two groups of tests. For m classes, Dixon's overall test has χ^2 distribution with $m(m-1)$ df, while type I and III tests have χ^2 distribution with $(m-1)^2$ df.

The results obtained from our extensive Monte Carlo simulations are summarized below:

- **Sample size requirement for asymptotic approximation**: 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 tests when all cell counts are at least 5. For NNCTs with smaller cell counts, we recommend the Monte Carlo randomization of the tests when the null hypothesis is RL, and we recommend the use of Monte Carlo critical values (as discussed in Section 6.1) when the null hypothesis is CSR independence.
- **Empirical size performance**:
  - In the two-class case, type I and type III cell-specific tests have better performance for the cell corresponding to the smaller class, while Dixon's cell-specific test has better performance for the cell corresponding to the larger class. For the overall test, the performance of the tests is similar for Dixon's and type I and type III tests.
  - In the three class case, type I and type III cell-specific tests have better performance, and overall tests have similar size estimates. We also observe that type I and III cell-specific tests and type III overall test are more robust to the differences in class sizes (i.e., differences in relative abundances).
- **Empirical power performance under segregation**:
  - In the two-class case, type I and type III cell-specific tests have similar power estimates which are larger than those of Dixon's, and the same holds for the overall tests.
  - In the three class case, type I, III and Dixon's cell-specific tests have similar power estimates, with type I and type III being slightly higher. The same holds for the overall tests.
- **Empirical power performance under association**:
  - In the two-class case, type I and type III cell-specific and overall tests tend to have higher power estimates for most of the class size combinations. The only exception is when the classes are highly unbalanced and the cell-specific test is for the diagonal cell with the larger class. In this case, Dixon's tests have higher power.
  - In the three class case, type I and type III cell-specific tests have higher power estimates for cell (i, j), if n_i is less than n_j, while Dixon's cell-specific tests have higher power estimates if n_i is larger than n_j.
  - For the overall tests, Dixon's overall test has the highest power estimates.
- **Overall recommendations**:
  - When empirical size and power performances are considered together, among cell-specific tests, type I and type III cell-specific tests are recommended against the segregation alternatives, while type I, type III, and Dixon's cell-specific tests are recommended against the association alternatives depending on the class sizes in the off-diagonal cells.
  - Among overall tests, type I and type III overall tests are recommended against the segregation alternatives, while Dixon's overall test is recommended against the association alternatives.
  - We extend this recommendation to one-vs-rest type tests as well. Furthermore, for one-vs-rest type tests, all the tests have similar size performance, but type I and type III are more robust to differences in relative abundances.

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 (Baddeley, Møller, & Waagepetersen, 2000) provide information on the pattern at various scales (i.e., around other distance values). Hence NNCT-tests and pair correlation or K-functions are not comparable but provide complementary information about the pattern in question. However, an advantage of overall NNCT-tests is that they provide the interaction in a multi-class setting in the presence of all classes, while the second order analysis with K or g functions allows a comparison of pairs of classes (one at a time). Furthermore, when an overall NNCT-test is significant, NNCT offers various post-hoc tests as follow-up to find the specifics of the interaction: (i) cell-specific tests, (ii) one-class-vs-rest type tests, and (iii) class-specific tests. In the cell-specific tests for cell (i, j), the interaction between classes i and j is examined in the presence of all other classes, and in the class i-vs-rest testing, the interaction of all the classes other than class i with class i is investigated. The pair correlation function and K-functions can also be adapted for one-vs-rest type analysis, as classes i and the rest of the classes can be treated as the two classes. On the other hand, the bivariate pair correlation function or K-functions are also applicable for classes i and j; however, this analysis is restricted to the classes i and j only in the sense that the influence of the other classes in the region is ignored. To the author's knowledge, the class-specific tests (which are discussed in Ceyhan, 2009) have no counterpart among the K-function type second order methods.

The course of action we recommend depends on which null hypothesis is more appropriate. If CSR independence is the reasonable null pattern, we recommend the overall segregation tests to detect the spatial interaction at small scales at 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, class-specific or one-vs-rest type tests can be performed; we recommend all versions as they provide information on different aspects of the spatial interaction. To detect spatial interaction at larger distances, pair correlation function is recommended (Stoyan & Penttinen, 2000), 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, we recommend the NNCT-tests to detect the interaction at about the mean NN distance, and Diggle's D-function (Diggle, 2003) or the modified version of Ripley's K function (Baddeley et al., 2000) to detect the interaction at higher distances.

Acknowledgments

References

Baddeley, A., Møller, J., & Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, *54*(3), 329–350.
Ceyhan, E. (2008). Overall and pairwise segregation tests based on nearest neighbor contingency tables. *Computational Statistics & Data Analysis*, *53*(8), 2786–2808.
Ceyhan, E. (2009). Class-specific tests of segregation based on nearest neighbor contingency tables. *Statistica Neerlandica*, *63*(2), 149–182.
Ceyhan, E. (2010a). New tests of spatial segregation based on nearest neighbor contingency tables. *Scandinavian Journal of Statistics*, *37*, 147–165.
Ceyhan, E. (2010b). On the use of nearest neighbor contingency tables for testing spatial segregation. *Environmental and Ecological Statistics*, *17*(3), 247–282.
Ceyhan, E. (2013). *New cell-specific and overall tests of spatial interaction based on nearest neighbor contingency tables. Technical Report # KU-EC-13-1.* Istanbul, Turkey: Koç University, arXiv:1206.1850v2 [stat.ME].
Ceyhan, E. (2014). Simulation and characterization of multi-class spatial patterns from stochastic point processes of randomness, clustering and regularity. *Stochastic Environmental Research and Risk Assessment*, *38*(5), 1277–1306.
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.* New York: Wiley.
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* (second ed.). London: Hodder Arnold Publishers.
Diggle, P., Zheng, P., & Durr, P. (2005). Nonparametric estimation of spatial segregation in a multivariate point process: Bovine tuberculosis in Cornwall, UK. *Proceedings of National Academy Sciences*, *54*(3), 645–658.
Dixon, P. M. (1994). Testing spatial segregation using a nearest-neighbor contingency table. *Ecology*, *75*(7), 1940–1948.
Dixon, P. M. (2002a). Nearest-neighbor contingency table analysis of spatial segregation for several species. *Ecoscience*, *9*(2), 142–151.
Dixon, P. M. (2002b). Nearest neighbor methods. In Abdel H. El-Shaarawi, & Walter W. Piegorsch (Eds.), *Encyclopedia of environmetrics, Vol 3* (pp. 1370–1383). NY: John Wiley & Sons Ltd..
Fossett, M. (2011). Generative models of segregation: Investigating model-generated patterns of residential segregation by ethnicity and socioeconomic status. *The Journal of Mathematical Sociology*, *35*(1–3), 114–145.
Good, B. J., & Whipple, S. A. (1982). Tree spatial patterns: South Carolina bottomland and swamp forests. *Bulletin of the Torrey Botanical Club*, *109*(4), 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.
Hamill, D. M., & Wright, S. J. (1986). Testing the dispersion of juveniles relative to adults: A new analytical method. *Ecology*, *67*(2), 952–957.
Henry, A. D., Pralat, P., & Zhang, C. (2011). Emergence of segregation in evolving social networks. *Proceedings of National Academy Sciences*, *108*(21), 8605–8610.
Kulldorff, M. (2006). Tests for spatial randomness adjusted for an inhomogeneity: A general framework. *Journal of the American Statistical Association*, *101*(475), 1289–1305.
Lahiri, S. N. (1996). On consistency of estimators based on spatial data under infill asymptotics. *Sankhya: The Indian Journal of Statistics, Series A*, *58*(3), 403–417.
Li, H., Calder, C., & Cressie, N. (2011). One-step estimation of spatial dependence parameters: Properties and extensions of the APLE statistic. *Journal of Multivariate Analysis*, *105*, 68–84.
Meyer, C. D., Jr. (1973). Generalized inverses and ranks of block matrices. *SIAM Journal on Applied Mathematics*, *25*, 597–602.
Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, *37*, 17–23.
Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, *70*, 120–126.
Penrose, R. A. (1955). A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society*, *51*, 406–413.
Penrose, M. (2003). *Oxford studies in probability*: *Vol. 5. Random geometric graphs.* Oxford, UK: Oxford University Press.
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* (second ed.). New York: Wiley-Interscience.
Robertson, S. L., & Cushing, J. M. (2011). Spatial segregation in stage-structured populations with an application to Tribolium. *Journal of Biological Dynamics*, *5*(5), 398–409.
Searle, S. R. (2006). *Matrix algebra useful for statistics.* New York: Wiley-Interscience.
Stoyan, D., & Penttinen, A. (2000). Recent applications of point process methods in forestry statistics. *Statistical Science*, *15*(1), 61–78.
Stoyan, D., & Stoyan, H. (1994). *Fractals, random shapes and point fields: methods of geometrical statistics.* New York: John Wiley and Sons.
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*(4), 511–532.
Whipple, S. A. (1980). Population dispersion patterns of trees in a Southern Louisiana hardwood forest. *Bulletin of the Torrey Botanical Club*, *107*(1), 71–76.