Environ Ecol Stat (2010) 17:247–282
DOI 10.1007/s10651-008-0104-x

On the use of nearest neighbor contingency tables for testing spatial segregation

Elvan Ceyhan

Received: 16 April 2008 / Revised: 5 September 2008 / Published online: 1 October 2008

Abstract For two or more classes (or types) of points, nearest neighbor contingency tables (NNCTs) are constructed using nearest neighbor (NN) frequencies and are used in testing spatial segregation of the classes. Pielou's test of independence, Dixon's cell-specific, class-specific, and overall tests are the tests based on NNCTs (i.e., they are NNCT-tests). These tests are designed and intended for use under the null pattern of random labeling (RL) of completely mapped data. However, it has been shown that Pielou's test is not appropriate for testing segregation against the RL pattern while Dixon's tests are. In this article, we compare Pielou's and Dixon's NNCT-tests; introduce the one-sided versions of Pielou's test; extend the use of NNCT-tests for testing complete spatial randomness (CSR) of points from two or more classes (which is called *CSR independence*, henceforth). We assess the finite sample performance of the tests by an extensive Monte Carlo simulation study and demonstrate that Dixon's tests are also appropriate for testing CSR independence; but Pielou's test and the corresponding one-sided versions are liberal for testing CSR independence or RL. Furthermore, we show that Pielou's tests are only appropriate when the NNCT is based on a random sample of (base, NN) pairs. We also prove the consistency of the tests under their appropriate null hypotheses. Moreover, we investigate the edge (or boundary) effects on the NNCT-tests and compare the buffer zone and toroidal edge correction methods for these tests. We illustrate the tests on a real life and an artificial data set.

Keywords Association · Clustering · Completely mapped data · Complete spatial randomness · Edge correction · Random labeling · Spatial point pattern

E. Ceyhan (✉)
Department of Mathematics, Koç University, 34450 Sarıyer, Istanbul, Turkey
e-mail: elceyhan@ku.edu.tr

1 Introduction

The analysis of spatial point patterns in natural populations has been extensively studied in various fields. In particular, spatial patterns in epidemiology, population biology, and ecology have important implications. A spatial point pattern includes the locations of some measurements, such as the coordinates of trees in a region of interest. These locations are referred to as *events* by some authors, in order to distinguish them from arbitrary points in the region of interest (Diggle 2003). However in this article such a distinction is not necessary, as we only consider the locations of events. Hence *points* will refer to the locations of events, henceforth. Most point patterns also include other types of measurements for each point, such as a categorical label (e.g., species label) or size (e.g., height of pine saplings). Such labelled data are *marked point patterns* generated by marked point processes, which define the distributions of the "marks" or "labels" to the locations of the points and perhaps are the most common spatial point patterns. For a general discussion of marked point processes, see Diggle (2003), Gavrikov and Stoyan (1995), Penttinen et al. (1992), and Schlather et al. (2004). For convenience and generality, we call the different types of points as "classes". From the early days on, the related research has mostly been on only one class at a time; i.e., on spatial pattern of each class (e.g., density, clumpiness, etc.). These patterns in a one-class framework fall under the pattern category called *spatial aggregation* (Coomes et al. 1999) or *clustering*. However, it is also of practical interest to investigate the patterns of one class with respect to the other classes (Pielou 1961). The spatial relationships between two or more classes have important consequences especially for plant species. See, for example, Pielou (1961) and Dixon (1994, 2002a), for more detail. Although we refer to types of points as "classes", the "class" can be replaced by any characteristic of an observation at a particular location. For example, the pattern of spatial segregation has been investigated for species (Diggle 2003), age classes of plants (Hamill and Wright 1986) and sexes of dioecious plants (Nanami et al. 1999). We also note that many of the epidemiological applications are for a two-class system of case/control labels (Waller and Gotway 2004).

In various fields, there are many tests available for spatial point patterns. An extensive survey is provided by Kulldorff who enumerates more than 100 such tests, most of which need adjustment for some sort of inhomogeneity (Kulldorff 2006). He also provides a general framework to classify these tests. The most widely used tests include Pielou's test of segregation for two classes (Pielou 1961) due to its ease of computation and interpretation and Ripley's K or L-functions (Ripley 2004). The abundance of tests results because (i) the tests for which Monte Carlo critical values are the only criteria receive wide acceptance in various fields; (ii) there are many different types of segregation patterns and some tests are designed to detect only certain types of segregation patterns; and (iii) the lack of cross-fertilization between different scientific fields so that new tests are proposed unbeknownst to the developers of similar tests.

Nearest neighbor (NN) methods for spatial patterns include at least six different groups (see, e.g., Dixon 2002b). The methods utilize some measure of (dis)similarity between a point and its NN; such as the distance between the points or the class types of the points. The latter type of similarity is used in the NN methods considered in this article. Nearest neighbor contingency tables (NNCTs) are constructed using the NN frequencies of classes and are used in testing spatial patterns. Pielou (1961) proposed tests (for segregation, symmetry, niche specificity, etc.) based on NNCTs under the RL of locations in the study region and Dixon devised cell-specific, class-specific, and overall tests based on NNCTs for the two-class case (Dixon 1994) and extended his methodology to the multi-class case (Dixon 2002a) under RL. Pielou's tests have been used for the two-class case only. However it has been demonstrated that Pielou's test is not appropriate for the NNCTs constructed under the RL of points (Meagher and Burdick 1980).

In this article, we discuss the tests of spatial segregation based on NNCTs. We describe the necessary assumptions and hypotheses, the tests, and the underlying sampling frameworks for Pielou's and Dixon's tests. We propose one-sided versions of Pielou's test to detect the direction of deviation from the RL pattern; then extend the use of Pielou's and Dixon's tests for the CSR of points from two or more classes in the region of interest (i.e., for CSR independence). However, we demonstrate that under CSR independence, Dixon's tests are conditional tests, and propose a method to remove this conditional nature of Dixon's test. We also compare the empirical sizes of the NNCT-tests by an extensive Monte Carlo simulation study, where we demonstrate that Pielou's test and the corresponding one-sided versions are liberal in rejecting RL or CSR independence, while Dixon's tests are about the desired nominal level. We also prove the consistency of the tests under their appropriate null hypotheses; show that Pielou's test is only appropriate when the NNCT is based on a random sample of (base, NN) pairs (which is not realistic in practical situations). We also investigate the edge (or boundary) effects on the NNCT-tests under CSR independence only since edge effects is not a concern under RL.

We describe the spatial point patterns of RL and CSR independence in Sect. 2; describe the NNCT-tests in Sect. 3, in particular we describe the construction of the NNCTs in Sect. 3.1, Pielou's test in Sect. 3.2, Dixon's NNCT-tests in Sect. 3.3, extend Dixon's test for the CSR independence pattern in Sect. 3.4. We prove the consistency of the NNCT-tests in Sect. 4 (and defer the proofs to the Appendix Section); present our extensive Monte Carlo simulation analysis in Sect. 5, in particular we compare the empirical significance levels of the tests under RL in Sect. 5.1, under CSR independence in Sect. 5.2, under the independence of rows and cell counts in NNCTs in Sec. 5.3. We also consider the edge correction methods under the CSR independence pattern in Sect. 6; illustrate our methods on two example data sets in Sect. 7. We provide our discussions and conclusions as well as guidelines for using the tests in Sect. 8.

2 Spatial point patterns

For simplicity, we describe the spatial point patterns for two-class populations; the extension to the multi-class case is straightforward.

In the univariate (i.e., one-class) spatial point pattern analysis, the null hypothesis is usually *complete spatial randomness* (*CSR*) (Diggle 2003). Given a spatial point pattern $\mathcal{P} = \{X_i \cdot \mathbf{I}(X_i \in D), i = 1, \ldots, n : D \subset \mathbb{R}^2\}$ where X_i stands for the location of event i (i.e., point i) and $\mathbf{I}(X_i \in D)$ is the indicator function which denotes the Bernoulli random variable denoting the event that point i is in region D. The pattern

$\mathcal{P}$ exhibits CSR if given n events (ie., locations of the points) in domain D, the events are an independent random sample from the uniform distribution on D. Note that this condition also implies that there is no spatial interaction; i.e., the locations of these points have no influence on one another. Furthermore, when the reference region D is large, the number of points in any planar region with area $A(D)$ follows (approximately) a Poisson distribution with intensity (i.e., number of points per unit area) denoted by λ and mean $\lambda \cdot A(D)$.

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

In CSR independence, points from each of the two classes satisfy the CSR pattern in the region of interest. On the other hand, RL is the pattern in which, given a fixed set of locations in a region, class labels are assigned to these fixed locations randomly so that the labels are independent of the locations. So, RL is less restrictive than CSR independence, in the sense that RL does not impose any restrictions on the distribution of the locations of the events, but CSR independence is a process defining the spatial distribution of event locations.The RL or CSR independence patterns imply a more refined null hypothesis for the NNCT-tests, namely, H_o : randomness in the NN structure.When the points from each class are assumed to be uniformly distributed over the region of interest, then randomness in the NN structure is implied by the CSR independence pattern, which is also referred to as (a type of) "population independence" by some authors (Goreaud and Pélissier 2003). Note that this type of CSR is equivalent to the case where the RL procedure is applied to a given set of points from a CSR pattern in the sense that after points are generated uniformly in the region, the class labels are assigned randomly. When only the labeling of a set of fixed points (the allocation of the points could be regular, aggregated, or clustered, or of lattice type) is random, the randomness in the NN structure is implied by RL pattern.

The distinction between the RL and CSR independence is very important when defining the appropriate null model which depends on the particular ecological context. Goreaud and Pélissier (2003) discuss the differences between independence and RL patterns and show that the incorrect specification of the null pattern may result in incorrect results, e.g., for Ripley's K or L-functions. They also propose some guidelines to determine which null hypothesis is appropriate for a given situation. For the null case of CSR independence (just independence in Goreaud and Pélissier 2003) the locations of the points from two classes are *a priori* the result of (perhaps) different processes (e.g., individuals of different species or age cohorts), whereas for the null case of RL some processes affect *a posteriori* the individuals of a single population (e.g., diseased versus non-diseased individuals of a single species). Notice that although CSR independence and RL are not same, they lead to the same null model (i.e., randomness in NN structure) for tests using NNCT, which does not require spatially-explicit information.

Deviations from the null patterns (RL or CSR independence) were first called (positive or negative) *segregation*. In Pielou's approach, two classes can be described as "unsegregated" if the NN of an individual is as likely to be of the same class as the other class; that is, neither class has a tendency to occur in one-class clumps or clusters. *Negative segregation* occurs if the NN of a point is more likely to be from a different class than the class of the point. *Positive segregation* occurs if the NN of a point is more likely to be of the same class as the class of the point; i.e., the members of the same class tend to be clumped or clustered (see, e.g., Pielou 1961). The concept of "negative segregation" as described above, is more commonly referred to as *association*, whereas "positive segregation" is merely called *segregation*. See, for example, Cressie (1993) and Coomes et al. (1999) for more detail. Two classes may exhibit many different forms of segregation (Pielou 1961). Although it is not possible to list all segregation types, existence of segregation can be tested by using NNCTs. In the statistical literature, association in contingency tables generally refers to *categorical association*. To avoid confusion between this general association and the spatial pattern of association, we call the former as "categorical association" and the latter as "spatial association". No such confusion occurs for segregation.

3 Tests based on nearest neighbor contingency tables

In this section, we present the construction of NNCTs and then Pielou's and Dixon's tests based on NNCTs.

3.1 Construction of nearest neighbor contingency tables

Consider two classes labelled as $\{1, 2\}$. NNCTs are constructed using NN frequencies for each class. Let n_i be the number of points from class i for $i \in \{1, 2\}$ and $n = n_1 + n_2$. If we record the class of each point and its NN, the NN relationships fall into 4 categories: $(1, 1)$, $(1, 2)$; $(2, 1)$, $(2, 2)$, where in cell (i, j), class i is the base class, while class j is the class of its NN. That is, a (base, NN) pair is categorized according to its label. Denoting N_{ij} as the observed frequency of cell (i, j) for $i, j \in \{1, 2\}$, we obtain the NNCT in Table 1 where C_j is the sum of column j; i.e., number of times class j points serve as nearest neighbors for $j \in \{1, 2\}$. Note also that $n = \sum_{i,j} N_{ij}$, $n_i = \sum_{j=1}^{2} N_{ij}$, and $C_j = \sum_{i=1}^{2} N_{ij}$. We adopt the convention that capital letters stand for random quantities, while lower case letters stand for fixed quantities.

Table 1 The NNCT for two classes

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

Let $\pi_{ij} = P(U = i, V = j)$ be the probability that the pair of points (U, V) falls in cell (i, j); i.e., the point V is from class j and is a NN of the point U which is from class i. Furthermore, let $\nu_i = \pi_{i1} + \pi_{i2}$ for $i = 1, 2$, that is, the probability of a base point to be of class i. Similarly, let $\kappa_j = \pi_{1j} + \pi_{2j}$ for $j = 1, 2$, that is, the probability of a NN point to be of class j. The sample versions of these probabilities are $\widehat{\pi}_{ij} = N_{ij}/n$ for π_{ij}, $\widehat{\nu}_i = n_i/n$ for ν_i, and $\widehat{\kappa}_j = C_j/n$ for κ_j.

A (base, NN) pair can be categorized as reflexive or non-reflexive, regardless of the classes of the members of the pair. For a (base, NN) pair, (X, Y), (i.e., Y is a NN of X), if X is a NN of Y (i.e., (Y, X) is also a (base, NN) pair), then the pair (X, Y) is called *reflexive*. If a (base, NN) pair is not reflexive, then it is a *non-reflexive* pair. Moreover, a point can serve as NN to none or several other points. That is, a point can be a shared NN to k ($k < 6$) other points in $\mathbb{R}^2$ (Clark and Evans 1955).

3.2 Pielou's test of segregation

Pielou constructed NNCTs based on NN frequencies which yield tests that are independent of quadrat size (Pielou 1961; Krebs 1972). In the two-class case, Pielou used the usual Pearson's χ^2 test of independence (with 1 df) to test for presence or lack of segregation (Pielou 1961). Due to the ease in computation and interpretation, her test of segregation is widely used in ecology (Meagher and Burdick 1980) for both completely mapped or sparsely sampled data. In particular, Pielou has described and used her test of segregation for *completely mapped data*, although her test is not appropriate for such data (see Meagher and Burdick (1980) and Dixon (1994)). A data set is *completely mapped*, if the points (i.e., locations of all events) in a defined space are observed. Alternatively, *sparse sampling* might be suitable for the use of Pielou's test as suggested by Dixon (1994). Although sparse sampling is not clearly defined in the literature, it can be classified into two types. The sparse sampling schemes depend on the events (members of a class occupying a location) or arbitrary points in the region of interest. The most well known sparse sampling method is the *quadrat sampling*, in which the number of events falling into each of several (preferably random) small subregions (quadrats) is recorded. However construction of NNCTs may not even be possible in such a scheme, since the NN information may be lost when only the number of events are recorded for each quadrat. The second sparse sampling scheme is the *distance sampling*, in which the basic unit is an arbitrary point (not necessarily from the events) and the information based on the distance to the nearest event is recorded (Solow 1989). Notice that this type of distance sampling scheme does not yield sufficient information for the construction of NNCTs either.

Pearson's χ^2 test of independence for a 2×2 contingency table, in general, can be assumed to develop from one of the following frameworks: *Poisson, row-wise binomial*, or *overall multinomial sampling frameworks*. Below we briefly describe these frameworks for 2×2 contingency tables. Let $\widetilde{\pi}_{ij}$ be the probability of a point to fall in cell (i, j) in the contingency table, and $\widetilde{\nu}_i$ and $\widetilde{\kappa}_j$ be the probabilities that the point is of row category i and of column category j, respectively.

The test statistic for Pielou's test (which is same as Pearson's test) is given by

$$\mathcal{X}_P^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(N_{ij} - \mathbf{E}[N_{ij}])^2}{\mathbf{E}[N_{ij}]}. \tag{1}$$

Poisson sampling framework: Each category count in the contingency table is assumed to be an independent Poisson variate. Another key feature is the independence of cell counts, which would imply that $\widetilde{\pi}_{ij} = \widetilde{\nu}_i \, \widetilde{\kappa}_j$ for all $(i, j) \in \{1, 2\}$. This independence can be tested by Pearson's χ^2 test for large samples and Fisher's exact test or the exact version of the Pearson's test for small samples (Agresti 1992). The null hypothesis in this framework is $H_o : \widetilde{\pi}_{ij} = \widetilde{\nu}_i \, \widetilde{\kappa}_j$, and $\mathbf{E}[N_{ij}]$ of Eq. 1 is $\frac{N_i \, C_j}{n}$. Among the alternatives, $H_a : \widetilde{\pi}_{ii} > \widetilde{\nu}_i \, \widetilde{\kappa}_i$ is suggestive of positive categorical association for classes $i = 1, 2$, while $H_a : \widetilde{\pi}_{12} < \widetilde{\nu}_1 \, \widetilde{\kappa}_2$ or $H_a : \widetilde{\pi}_{21} < \widetilde{\nu}_2 \, \widetilde{\kappa}_1$ is suggestive of negative categorical association between classes 1 and 2.

However, in the case of NNCTs, cell counts, e.g., N_{12} and N_{21} are not independent under RL or CSR independence. Because, a (base, NN) pair is more likely to be a reflexive pair, rather than a non-reflexive pair under RL or CSR independence (Meagher and Burdick 1980). Thus under Poisson sampling framework, Pielou's test would be inappropriate for testing RL or CSR independence.

Row-wise binomial sampling framework: In this framework, we assume that $N_i = n_i$ are given and $N_{ij} \sim \text{BIN}(n_i, \widetilde{\pi}_{ij})$, the binomial distribution with n_i trials and probability of success being $\widetilde{\pi}_{ij}$. Notice that for more than two classes, this will be *row-wise multinomial framework*.

Then the null hypothesis for this test is $H_o : \widetilde{\pi}_{11} = \widetilde{\pi}_{21}$ which also implies $\widetilde{\pi}_{12} = \widetilde{\pi}_{22}$. The alternative $H_a : \widetilde{\pi}_{11} > \widetilde{\pi}_{21}$ would correspond to positive categorical association which would also imply $\widetilde{\pi}_{22} > \widetilde{\pi}_{12}$. Similarly, the alternative $H_a : \widetilde{\pi}_{11} < \widetilde{\pi}_{21}$ would correspond to negative categorical association which would also imply $\widetilde{\pi}_{22} < \widetilde{\pi}_{12}$.

Under H_o, we can parametrize the null model as $N_{ij} \sim \text{BIN}(n_i, \widetilde{\kappa}_j)$ where $\widetilde{\kappa}_j$ can be estimated as C_j/n and is assumed to equal the expectation $\mathbf{E}[C_j/n]$. Then $H_o : \widetilde{\pi}_{11} = \widetilde{\pi}_{21} = \widetilde{\kappa}_1$ is equivalent to $H_o : \mathbf{E}[N_{11}/n_1] = \mathbf{E}[N_{21}/n_2] = \widetilde{\kappa}_1$ which is equivalent to $H_o : \mathbf{E}[N_{11}] = n_1 \, \widetilde{\kappa}_1$ and $\mathbf{E}[N_{21}] = n_2 \, \widetilde{\kappa}_1$ which, for large n, n_1, and n_2, is equivalent to $H_o : \mathbf{E}[N_{11}/n] = n_1 \, \widetilde{\kappa}_1/n = \widetilde{\nu}_1 \, \widetilde{\kappa}_1$ and $\mathbf{E}[N_{21}/n] = n_2 \, \widetilde{\kappa}_1/n = \widetilde{\nu}_2 \, \widetilde{\kappa}_2$.Under H_o, if $\widetilde{\kappa}_j$ are known, $\mathcal{X}_P^2$ is approximately distributed as χ_2^2 (i.e., χ^2 distribution with 2 degrees of freedom) for large n_i; if $\widetilde{\kappa}_j$ are not known, but estimated as C_j/n, then $\mathcal{X}_P^2$ is approximately distributed as χ_1^2 for large n_i. In most practical situations, the latter case will occur, so χ_1^2 distribution is used for this test.

In the two-class case, (N_{11}, N_{12}) and (N_{21}, N_{22}) are assumed to be independent and so are the individual trials, namely, (base, NN) pairs. Under RL or CSR independence, this assumption is invalid for completely mapped data. Because the trials that constitute N_{ii} for $i = 1, 2$ are not independent due to reflexivity and shared NN structure; likewise, N_{12} and N_{21} are not independent. Hence, Pielou's test is not appropriate for RL or CSR independence in this framework either.

Overall multinomial sampling framework: An alternative sampling framework for contingency tables, in general, is that the cell counts are assumed to be from independent multinomial trials. That is, for the two-class case,

$$\mathbf{N} = (N_{11}, N_{12}, N_{21}, N_{22}) \sim \mathscr{M}(n, \widetilde{\pi}_{11}, \widetilde{\pi}_{12}, \widetilde{\pi}_{21}, \widetilde{\pi}_{22})$$

hence the name *overall multinomial framework*. The null hypothesis in this framework is $H_o : (\widetilde{\pi}_{11}, \widetilde{\pi}_{12}) = (\widetilde{\pi}_{21}, \widetilde{\pi}_{22})$ and $\mathbf{E}[N_{ij}]$ in Eq. 1 is $N_i\, C_j/n$. The multinomial counts are not independent (since they are negatively correlated) when conditioned on their total. This dependence alleviates as the sample size increases, but might confound the small sample results. In addition to this mild dependence, the NNCT cell counts are not independent due to, e.g., reflexivity. Hence the overall multinomial framework is not appropriate for NNCTs based on RL or CSR independence either.

Note that conditional on $N_i = n_i$, the overall multinomial framework reduces to the row-wise multinomial framework. Furthermore, when the parameters are not known but estimated from the marginal sums, all frameworks yield tests that are approximately distributed as χ^2_1 for large n.

3.2.1 *One-sided versions of Pielou's test of segregation*

Pielou's test is a general two-sided test, hence it does not indicate the direction of the deviation (e.g., positive or negative categorical association) from the null case. To determine the direction, one needs to check the NNCT. Since $\mathcal{X}^2_P \overset{approx}{\sim} \chi^2_1$, for large n, we can write $\mathcal{X}^2_P = Z^2_n$ where $Z_n \overset{approx}{\sim} N(0, 1)$, the standard normal distribution. By some algebraic manipulations, among other possibilities, for the row-wise multinomial framework, Z_n can be written as

$$Z_n = \left(\frac{N_{11}}{n_1} - \frac{N_{21}}{n_2}\right)\sqrt{\frac{n_1\, n_2\, n}{C_1\, C_2}}. \tag{2}$$

See Bickel and Doksum (1977) for the sketch of the derivation. Positive values of Z_n indicate positive categorical association, while negative values indicate negative categorical association. When cell counts are independent, a reasonable α-level test is rejecting H_o if $Z_n > z_{1-\alpha}$ for positive categorical association or if $Z_n < z_\alpha$ for negative categorical association. The α-level χ^2 test in which we reject for $\mathcal{X}^2_P > \chi^2_1(1-\alpha)$ is equivalent to the two-sided α-level test based on Z_n.

The corresponding test statistic for the overall multinomial framework can be written as

$$\widetilde{Z}_n = \left(N_{11} - \frac{n_1\, c_1}{n}\right)\sqrt{\frac{n^3}{n_1\, n_2\, c_1\, c_2}}. \tag{3}$$

Once again, we point out that these one-sided tests are not appropriate for testing RL or CSR independence, due to inherent dependence of cell counts in NNCTs based on such patterns.

Remark 3.1 Appropriate null case for Pielou's tests: In Pielou's test, each of the Poisson and row-wise binomial sampling frameworks for cell counts assumes that the trials (i.e., the cross-categorization of base-NN pairs) are independent and in the overall multinomial framework, there is mild dependence between the cell counts. The independence of rows and individual trials (i.e., cells) would follow if NNCT were based on a random sample of (base label, NN label) pairs. But unfortunately, this usually is not realistic in practice, although it might have theoretical appeal. When we have a random sample of (base label, NN label) pairs (which are also called the (base, NN) pairs), the null hypothesis for Pielou's test is equivalent to the case that the vector of probabilities for the cell frequencies for each row are identical. Hence if the NNCT is based on a random sample of (base, NN) pairs, then any of the sampling frameworks would be appropriate which in turn implies the appropriateness of Pearson's test of independence for the NNCT. The null hypothesis in each of the sampling frameworks will imply independence between the patterns of the two classes. On the other hand the alternatives of positive categorical association will correspond to segregation of the classes, while negative categorical association will correspond to spatial association of the classes. □

3.3 Dixon's NNCT-tests

Dixon proposed a series of tests for segregation based on NNCTs, namely, cell- and class-specific tests, and overall test of segregation under RL (Dixon 1994).

3.3.1 Dixon's cell-specific tests

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

When the null hypothesis is RL, we have

$$\mathbf{E}[N_{ij}] = \begin{cases} n_i(n_i-1)/(n-1) & \text{if } i = j, \\ n_i\, n_j/(n-1) & \text{if } i \neq j, \end{cases} \tag{4}$$

or equivalently

$$\pi_{ij} = \frac{n_i\,(n_i-1)}{n(n-1)}\mathbf{I}(i=j) + \frac{n_i\, n_j}{n(n-1)}\mathbf{I}(i \neq j).$$

The test statistic suggested by Dixon is given by

$$Z^D_{ij} = \frac{N_{ij} - \mathbf{E}[N_{ij}]}{\sqrt{\mathbf{Var}[N_{ij}]}} \tag{5}$$

where

$$\mathbf{Var}[N_{ij}] = \begin{cases} (n+R)p_{ii} + (2n-2R+Q)p_{iii} + (n^2-3n-Q+R)p_{iiii} - (np_{ii})^2 & \text{if } i = j, \\ n\,p_{ij} + Q\,p_{iij} + (n^2-3\,n-Q+R)\,p_{iijj} - (n\,p_{ij})^2 & \text{if } i \neq j, \end{cases} \quad (6)$$

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

$$\begin{aligned} p_{ii} &= \frac{n_i\,(n_{i-1})}{n\,(n-1)}, & p_{ij} &= \frac{n_i\,n_j}{n\,(n-1)}, \\ p_{iii} &= \frac{n_i\,(n_i-1)\,(n_i-2)}{n\,(n-1)\,(n-2)}, & p_{iij} &= \frac{n_i\,(n_i-1)\,n_j}{n\,(n-1)\,(n-2)}, \\ p_{iijj} &= \frac{n_i\,(n_i-1)\,n_j\,(n_j-1)}{n\,(n-1)\,(n-2)\,(n-3)}, & p_{iiii} &= \frac{n_i\,(n_i-1)\,(n_i-2)\,(n_i-3)}{n\,(n-1)\,(n-2)\,(n-3)}. \end{aligned} \quad (7)$$

Furthermore, Q is the number of points with shared NNs, which occurs when two or more points share a NN and R is twice the number of reflexive pairs. Then $Q = 2\,(Q_2 + 3\,Q_3 + 6\,Q_4 + 10\,Q_5 + 15\,Q_6)$ where Q_k is the number of points that serve as a NN to other points k times.

One-sided and two-sided tests are possible for each cell (i, j) using the asymptotic normal approximation of Z^D_{ij} given in Eq. 5 (Dixon 1994). In Dixon's framework, N_{ij} are random quantities; and the quantities in the expectations, hypotheses, and variances are conditional on $N = n$ and $N_i = n_i$ for $i \in \{1, 2\}$. The column sums are irrelevant for Dixon's tests.

We describe the setting in a broader context. Let ν_i be the probability of an arbitrary point being from class i. Then under RL, $\pi_{ij} = \nu_i\,\nu_j$ and the expression $\frac{n_i\,(n_i-1)}{n(n-1)}\mathbf{I}(i = j) + \frac{n_i\,n_j}{n(n-1)}\mathbf{I}(i \neq j)$ can be viewed as an estimate of π_{ij} and denoted as $\widehat{\pi}_{ij}$. Furthermore, given large $N = n$, under the null hypothesis of RL the expected values given in Eq. 4 implies

$$H_o : \pi_{ij} = \nu_i\,\nu_j$$

and the test statistic $(N_{ij}/n - \pi_{ij})\big/\sqrt{\mathbf{Var}[N_{ij}]}$ is approximately equivalent to Z^D_{ij} in Eq. 5. In Dixon's framework, for large n and n_i, the row marginals satisfy $\mathbf{E}[N_i/n] = \nu_i$ and the column marginals satisfy $\mathbf{E}[C_j/n] = \kappa_j = \sum_{i=1}^{2} \nu_i\,\nu_j = \nu_j$.

3.3.2 Dixon's overall test of segregation

Dixon's overall test of segregation tests the hypothesis that expected values of the cell counts in the NNCT are equal to the ones given in Eq. 4. In the two-class case, he calculates $Z_{ii} = (N_{ii} - \mathbf{E}[N_{ii}])/\sqrt{\mathbf{Var}[N_{ii}]}$ for both $i \in \{1, 2\}$ and then combines these test statistics into a statistic that is asymptotically distributed as χ^2_2 (Dixon 1994).

The suggested test statistic is given by

$$\mathcal{X}_D^2 = \mathbf{Y}'\Sigma^{-1}\mathbf{Y} = \begin{bmatrix} N_{11} - \mathbf{E}[N_{11}] \\ N_{22} - \mathbf{E}[N_{22}] \end{bmatrix}' \times \begin{bmatrix} \mathbf{Var}[N_{11}] & \mathbf{Cov}[N_{11}, N_{22}] \\ \mathbf{Cov}[N_{11}, N_{22}] & \mathbf{Var}[N_{22}] \end{bmatrix}^{-1} \times \begin{bmatrix} N_{11} - \mathbf{E}[N_{11}] \\ N_{22} - \mathbf{E}[N_{22}] \end{bmatrix} \tag{8}$$

where $\mathbf{E}[N_{ii}]$ are as in Eq. 4, $\mathbf{Var}[N_{ii}]$ are as in Eq. 6, and

$$\mathbf{Cov}[N_{11},\ N_{22}] = \left(n^2 - 3n - Q + R\right) p_{1122} - n^2\, p_{11}\, p_{22}.$$

Under H_o : $\mathbf{E}[N_{ii}/n] = \nu_i^2$ for $i = 1, 2$, and $\mathbf{E}[\mathcal{X}_D^2] = 2$ and $\mathbf{Var}[\mathcal{X}_D^2] = 4$; i.e., the non-centrality parameter $\lambda = 0$. If we parametrize the segregation alternative as H_a^S : $\mathbf{E}[N_{11}/n] = (\nu_1 + \varepsilon_1)^2$ and $\mathbf{E}[N_{22}/n] = (1 - \nu_1 + \varepsilon_2)^2$ for some $\varepsilon_1, \varepsilon_2 > 0$. Then under H_a^S, the non-centrality parameter satisfies $\lambda = \lambda(\varepsilon_1, \varepsilon_2) > 0$ since $\lambda(\varepsilon_1, \varepsilon_2) = \mathbf{E}_S[\mathbf{Y}]'\, \Sigma_S^{-1}\mathbf{E}_S[\mathbf{Y}]$ where

$$\frac{1}{n}\mathbf{E}_S[\mathbf{Y}]' = [(\nu_1 + \varepsilon_1), (1 - \nu_1 + \varepsilon_2)]$$

and Σ_S is the (positive definite) variance-covariance matrix of the cell counts under H_a^S. If the association alternative is parametrized as above with $\varepsilon_1, \varepsilon_2 < 0$ then we obtain the same non-centrality parameter $\lambda(\varepsilon_1, \varepsilon_2)$.

Dixon (2002a) extends his test for multi-class case (i.e., for the case with three or more classes). He also partitions the overall test statistic $\mathcal{X}_D^2$ into class-specific test statistics each of which are dependent but approximately follow a χ^2 distribution.

3.4 Dixon's tests under CSR independence

The expected values of the NNCT cell counts given in Eq. 4 are derived under RL or CSR independence by Dixon (1994). However, the variances and covariances of the cell counts used in Sects. 3.3.1 and 3.3.2 are derived under the RL pattern only (Dixon 1994, 2002a).

When the null hypothesis is CSR independence, the expressions for the variances and covariances of the NNCT cell counts are as in RL case, except they are conditional on Q and R. The quantities Q and R are fixed under RL, but random under CSR independence. Hence under CSR independence Dixon's cell-specific test given in Eq. 5 asymptotically has $N(0, 1)$ distribution and overall test given in Eq. 8 asymptotically has χ_2^2, conditional on Q and R. Under the CSR independence pattern, the unconditional variances and covariances (hence the unconditional asymptotic distributions) can be obtained by replacing Q and R with their expectations.

Unfortunately, given the difficulty of calculating the expectations of Q and R under CSR independence, it seems reasonable and convenient to use test statistics employing the unconditional variances and covariances even when assessing their behavior under the CSR independence pattern. Alternatively, one can estimate the values of Q and R empirically, and substitute these estimates in the variance and covariance expressions. For example, for homogeneous planar Poisson pattern, we have $\mathbf{E}[Q/n] \approx .632786$ and $\mathbf{E}[R/n] \approx 0.621120$ (estimated empirically by 1000000 Monte Carlo simulations for various values of $n = n_1 + n_2$ on the unit square).

To assess the influence of conditioning on the performance of Dixon's tests for the two-class case, we consider both the conditional version of these tests, as well as the unconditional version, in which the terms Q and R are replaced by $0.63\,n$ and $0.62\,n$, respectively. We call the latter type of correction as *QR-adjustment* and the transformed tests as *QR-adjusted* tests, henceforth. QR-adjusted version of Dixon's cell-specific test statistic for cell (i, j) is denoted by $Z_{ij}^{D,qr}$ and of the overall test statistic is denoted by $\mathcal{X}^2_{D,qr}$.

Remark 3.2 Extension of NNCT-tests to multi-class case: Dixon has extended his tests into multi-class situation with three or more classes (Dixon 2002a). For q classes with $q > 2$, the corresponding NNCT is of dimension $q \times q$. It is possible to define q^2 cell-specific tests as in Eq. 5, and one can combine the tests into one overall test similar to the one given in Eq. 8, which will have $\chi^2_{q(q-1)}$, asymptotically. On the other hand, Pielou's test is defined and has only been used for the two-class spatial patterns. Its inappropriateness discourages its immediate extension to multi-class patterns. □

3.5 Comparison of Pielou's and Dixon's tests

Dixon points out two problems with Pielou's test of independence: (i) it fails to identify certain types of segregation (e.g., mother-daughter processes) and (ii) the sampling distribution of NNCT cell counts is not appropriate (see Dixon 1994). In a mother-daughter process, mothers are distributed randomly in the region of interest, while the daughters are randomly displaced within close vicinity of their mothers. In such a process, it is possible to obtain a NNCT in which the cell counts are very similar to the ones expected under the Pielou null hypothesis, while in reality the pattern exhibits the segregation of the daughters. For more detail on mother-daughter processes and examples for which Pielou's test giving misleading results, see Dixon (1994). Problem (ii) was first noted by Meagher and Burdick (1980) who identify the main source of it to be reflexivity of (base, NN) pairs. As an alternative, they suggest using Monte Carlo simulations for Pielou's test. Dixon shows that Pielou's test is not appropriate for completely mapped data, but suggests that it might be appropriate for sparsely sampled data (Dixon 1994).

In Pielou's test, each of the sampling frameworks requires that the cell counts are independent. However, when a trial is label categorization of a (base, NN) pair, the assumption of independence between trials is violated due to reflexivity and shared NN structure. Thus Pielou's test measures deviations not only from the null pattern of RL or CSR independence but also from the independence of trials. This also suggests that Pielou's test would be liberal in rejecting the null hypothesis. The reflexivity and shared NN structure are not merely finite sample patterns, as they follow a certain non-degenerate distribution even when $n \to \infty$ (Clark and Evans 1955).

By construction, Pielou's test is used to test independence of the class labels of the (base, NN) pairs, but ignores the spatial information (hence ignores the spatial dependence, e.g., reflexivity of NNs). On the other hand, Dixon's tests are used for the null hypotheses of RL or CSR independence and uses more of the spatial information. For Dixon's tests, the underlying sampling framework for cell counts is different from Poisson, row-wise binomial, or overall multinomial sampling models of the contingency tables. In his framework, the probability of class j point serving as a NN of a class i point depends only on the class sizes (i.e., row sums), but not the total number of times class j serves as a NN (i.e., column sums). On the other hand, Pielou's test depends on both row and column sums. In fact, Pielou starts her arguments with NN probabilities depending on class sizes (row sums) in (Pielou 1961, pp. 257–258). Then she leaves this track of development because of dependence due to shared NN structure (i.e., the distribution of Q_k (Clark and Evans 1955)). For testing RL or CSR independence, Dixon's framework is more appropriate as a sampling distribution for NNCT cell counts, as it accounts for the inherent spatial dependence between observations.

4 Consistency of the NNCT-tests

The null hypotheses are different for Pielou's and Dixon's framework of testing spatial patterns, and so are the alternative hypotheses. Hence the acceptance regions are different (see, e.g., Dixon 1994), and no test is uniformly superior to the other, since both $A_D \setminus A_P$ and $A_P \setminus A_D$ are non-empty, where A_D is the acceptance region for Dixon's test and A_P is the acceptance region for Pielou's test. That is, there are situations in which Pielou's test yields a significant result, while Dixon's test finds no significant segregation, and vice versa. For example, a pattern resulting from a mother/daugter process can fall in $A_P \setminus A_D$ (see Sect. 3.5). On the other hand, a process in which row and column sums in a NNCT are close but the cell counts are different than expected under Pielou null hypothesis, and similar to the ones expected under Dixon null hypothesis might yield a pattern that falls in $A_D \setminus A_P$. Therefore the comparison of the tests (even for large samples) is inappropriate. But any reasonable test should have more power as the sample size increases. So, we prove that the tests under consideration are consistent, although they have appropriate size under different null hypotheses. The proofs of lemmas and theorems in this section are all deferred to the Appendix.

In the following theorems we use the consistency of tests based on statistics that have $N(0, 1)$ or χ^2_ν distributions asymptotically. First, we prove the consistency of Pielou's test of segregation and the one-sided versions. Let $z(\alpha)$ be the $100(1-\alpha)^{th}$ percentile for the standard normal distribution and $\chi^2_\nu(\alpha)$ be the $100(1-\alpha)^{th}$ percentile for χ^2 distribution with ν df.

Theorem 4.1 *(I) Suppose the NNCT is constructed based on a random sample of (base, NN) pairs. The test for segregation* $H_a^S : \widetilde{\pi}_{11} > \widetilde{\pi}_{22}$ *(spatial association* $H_a^A : \widetilde{\pi}_{11} < \widetilde{\pi}_{22}$*) which rejects* $H_o : \widetilde{\pi}_{11} = \widetilde{\pi}_{22}$ *for* $Z_n > z(1-\alpha)$ *(for* $Z_n < z(\alpha)$*) with* Z_n *given in Eq. 2 has size* α *and is consistent. Likewise, the test against segregation*

H_a^S : $\mathbf{E}[N_{11}/n] > \widetilde{\nu}_1 \widetilde{\kappa}_1$ which rejects H_o : $\mathbf{E}[N_{11}/n] = \widetilde{\nu}_1 \widetilde{\kappa}_1$ for $\widetilde{Z}_n > z(1-\alpha)$ with $\widetilde{Z}_n$ given in Eq. 3 has size α and is consistent.
(II) Under RL or CSR independence, the size of the above one-sided tests are larger than α (i.e., the tests are liberal in rejecting H_o) but consistent (in the sense that the power goes to 1 as marginal sums tend to ∞ under the alternatives).

Theorem 4.2 *(I) Suppose the NNCT is based on a random sample of (base, NN) pairs. The test for H_a : $\widetilde{\pi}_{11} \neq \widetilde{\pi}_{21}$ which rejects H_o : $\widetilde{\pi}_{11} = \widetilde{\pi}_{21}$ for $\mathcal{X}_P^2 > \chi_1^2(1-\alpha)$ with $\mathcal{X}_P^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(N_{ij} - \mathbf{E}[N_{ij}])^2}{\mathbf{E}[N_{ij}]}$ is consistent.*
(II) Under RL or CSR independence, the level of the test using $\mathcal{X}_P^2$ is larger than α (i.e., it is liberal in rejecting these null patterns) but is consistent (in the sense of part (II) of Theorem 4.1).

Next, we prove the consistency of Dixon's cell-specific and overall tests of segregation.

Theorem 4.3 *Under RL, Dixon's cell-specific test for cell (i, j) in a NNCT denoted by Z_{ij}^D; i.e., the test rejecting H_o : $\pi_{ij} = \frac{n_i\,(n_i-1)}{n(n-1)} \mathbf{I}(i=j) + \frac{n_i\,n_j}{n(n-1)} \mathbf{I}(i \neq j)$ (i.e., RL) against the two-sided (and one-sided alternatives) for $|Z_{ij}^D| > z(1-\alpha/2)$ (and $Z_{ij}^D > z(1-\alpha)$ or $Z_{ij}^D < z(\alpha)$) with $Z_{ij}^D = \frac{N_{ij} - \mathbf{E}[N_{ij}]}{\sqrt{\mathbf{Var}[N_{ij}]}}$ is of size α and is consistent. Under CSR independence, Z_{ij}^D is consistent conditional on Q and R.*

Theorem 4.4 *Under RL, Dixon's overall test of segregation; i.e., the test rejecting H_o : $\pi_{ij} = \frac{n_i\,(n_i-1)}{n(n-1)} \mathbf{I}(i=j) + \frac{n_i\,n_j}{n(n-1)} \mathbf{I}(i \neq j)$ for all $i, j \in \{1, 2\}$ (i.e., RL) against the alternative H_a : $\pi_{ij} \neq \frac{n_i\,(n_i-1)}{n(n-1)} \mathbf{I}(i=j) + \frac{n_i\,n_j}{n(n-1)} \mathbf{I}(i \neq j)$ for some $i, j \in \{1, 2\}$ for $\mathcal{X}_D^2 > \chi_2^2(1-\alpha)$ with $\mathcal{X}_D^2 = (\mathbf{N} - \mathbf{E}[\mathbf{N}])' \Sigma^- (\mathbf{N} - \mathbf{E}[\mathbf{N}])$ is of size α and is consistent. Under CSR independence, $\mathcal{X}_D^2$ is consistent conditional on Q and R.*

5 Monte Carlo simulation analysis

Pielou's test of independence and Dixon's overall test of segregation are not testing the same null pattern, so we can not compare the power of the tests under either segregation or association alternatives and we only implement Monte Carlo simulations to evaluate the finite sample performance of the tests in terms of empirical size. For the null case, we simulate the RL, CSR independence patterns, and independence of the rows in the NNCTs, with two classes labelled as X and Y with sizes n_1 and n_2, respectively.

5.1 Empirical significance levels of the NNCT-tests under RL

Under RL, we consider four cases. In RL Case (1) we use the locations of the trees in the swamp tree data (see Fig. 2 and Dixon 1994) as the fixed points, and randomly assign $n_1 = 182$ points as X and $n_2 = 91$ points as Y points. In each of the other RL cases, we first determine the fixed locations of points for which class labels are to be assigned randomly. Then we apply the RL procedure to these points for respective sample size combinations as follows.

RL Case (2) First, we generate $n = n_1 + n_2$ points iid $\mathcal{U}((0, 1) \times (0, 1))$, the uniform distribution on the unit square, for some combinations of $n_1, n_2 \in \{10, 30, 50, 100\}$. In each (n_1, n_2) combination, the locations of these points are taken to be the fixed locations for which we assign the class labels randomly. For each sample size combination (n_1, n_2), we randomly choose n_1 points (without replacement) and label them as X and the remaining n_2 points as Y points, and repeat the RL procedure $N_{mc} = 10000$ times. At each Monte Carlo replication, we compute the NNCT-tests. Out of these 10000 samples the number of significant outcomes by each test is recorded. The nominal significance level used in all these tests is $\alpha = .05$. The empirical sizes are calculated as the ratio of number of significant results to the number of Monte Carlo replications, N_{mc}. That is, for example, empirical size for Dixon's overall test for (10, 10), denoted by $\widehat{\alpha}_D$, is calculated as $\widehat{\alpha}_D := \sum_{i=1}^{N_{mc}} \mathbf{I}(\mathcal{X}^2_{D,i} \geq \chi^2_2(.05))$ where $\mathcal{X}^2_{D,i}$ is the value of Dixon's overall test statistic for iteration i, $\chi^2_2(.05)$ is the 95^{th} percentile of χ^2_2 distribution, and $\mathbf{I}(\cdot)$ is the indicator function.

RL Case (3) We generate n_1 points iid $\mathcal{U}((0, 2/3) \times (0, 2/3))$ and n_2 points iid $\mathcal{U}((1/3, 1) \times (1/3, 1))$ for some combinations of $n_1, n_2 \in \{10, 30, 50, 100\}$. The locations of these points are taken to be the fixed locations for which we assign the class labels randomly. The RL procedure is applied to these fixed points $N_{mc} = 10000$ times for each sample size combination and the empirical sizes for the tests are calculated similarly as in RL Case (2).

RL Case (4) We generate n_1 points iid $\mathcal{U}((0, 1) \times (0, 1))$ and n_2 points iid $\mathcal{U}((2, 3) \times (0, 1))$ for some combinations of $n_1, n_2 \in \{10, 30, 50, 100\}$. The RL procedure is applied and the empirical sizes for the tests are calculated as in the previous RL Cases.

The locations for which the RL procedure is applied in RL Cases (2–4) are plotted in Fig. 1 for $n_1 = n_2 = 100$. Although there are many possibilities for the allocation of points to which RL procedure can be applied, we only chose the locations of trees in a real life data set and three generic cases. Observe that in RL Case (2), the allocation of the points are a realization of a homogeneous Poisson process in the unit square; in RL Case (3) the points are a realization of two overlapping clusters; in RL Case (4) the points are a realization of two disjoint clusters.

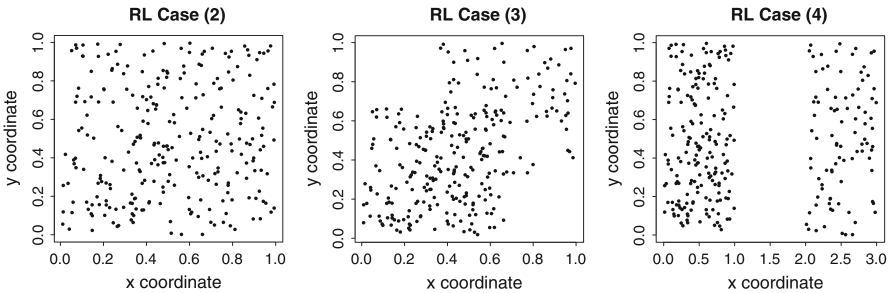


Fig. 1 The fixed locations of points for which RL procedure is applied for RL Cases (2–4) with $n = 200$ (for the case with $n_1 = n_2 = 100$) in the two-class case. Notice that x-axis for RL Case (4) is differently scaled than others

Table 2 The empirical significance levels for the tests under RL cases (1–4) at $\alpha = .05$

	Cell-specific		One-sided		Overall		
RL Case (1)							
(n_1, n_2)	$\widehat{\alpha}^D_{1,1}$	$\widehat{\alpha}^D_{2,2}$	$\widehat{\alpha}_R$	$\widehat{\alpha}_L$	$\widehat{\alpha}_P$	$\widehat{\alpha}_{PY}$	$\widehat{\alpha}_D$
(182, 91)	.0420[a]	.0487	.1563[b]	.1874[b]	.1228[b]	.0972[b]	.0486
RL Case (2)							
(10, 10)	.0604[b]	.0557[b]	.0800[b]	.1479[b]	.1219[b]	.0586[b]	.0349[a]
(10, 30)	.0311[a]	.0699[b]	.0824[b]	.1456[b]	.1522[b]	.0618[b]	.0466
(10, 50)	.0264[a]	.0472	.0953[b]	.0517	.0640[b]	.0307[a]	.0507
(30, 30)	.0579[b]	.0547[b]	.0749[b]	.1065[b]	.1249[b]	.0805[b]	.0497
(30, 50)	.0621[b]	.0608[b]	.0896[b]	.1203[b]	.1338[b]	.0826[b]	.0444[a]
(50, 50)	.0512	.0524	.0794[b]	.1058[b]	.1383[b]	.1025[b]	.0497
(50, 100)	.0625[b]	.0512	.0905[b]	.1060[b]	.1199[b]	.0926[b]	.0482
(100, 100)	.0538[b]	.0534	.0895[b]	.1101[b]	.1321[b]	.1052[b]	.0525
RL Case (3)							
(10, 10)	.0624[b]	.0657[b]	.0806[b]	.1492[b]	.1220[b]	.0571[b]	.0446[a]
(10, 30)	.0297[a]	.0341[a]	.0803[b]	.1454[b]	.1382[b]	.0517	.0327[a]
(10, 50)	.0251[a]	.0384[a]	.0882[b]	.0463[a]	.0591[b]	.0287[a]	.0508
(30, 30)	.0513	.0523	.0839[b]	.1179[b]	.1402[b]	.0933[b]	.0469
(30, 50)	.0626[b]	.0594[b]	.0934[b]	.1174[b]	.1367[b]	.0846[b]	.0411[a]
(50, 50)	.0509	.0511	.0800[b]	.1113[b]	.1414[b]	.1047[b]	.0501
(50, 100)	.0566[b]	.0421[a]	.0906[b]	.1019[b]	.1182[b]	.0906[b]	.0460[a]
(100, 100)	.0439[a]	.0453[a]	.0942[b]	.1127[b]	.1361[b]	.1098[b]	.0505
RL Case (4)							
(10, 10)	.0656[b]	.0640[b]	.0798[b]	.1481[b]	.1236[b]	.0536	.0432[a]
(10, 30)	.0281[a]	.0447[a]	.0798[b]	.1525[b]	.1639[b]	.0521	.0324[a]
(10, 50)	.0260[a]	.0404[a]	.0892[b]	.0506	.0618[b]	.0290[a]	.0500
(30, 30)	.0549[b]	.0553[b]	.0858[b]	.1183[b]	.1459[b]	.0984[b]	.0484
(30, 50)	.0677[b]	.0685[b]	.0936[b]	.1156[b]	.1359[b]	.0861[b]	.0445[a]
(50, 50)	.0504	.0506	.0769[b]	.1094[b]	.1372[b]	.0991[b]	.0488
(50, 100)	.0590[b]	.0484	.0882[b]	.1006[b]	.1179[b]	.0887[b]	.0479
(100, 100)	.0495	.0476	.0941[b]	.1134[b]	.1406[b]	.1137[b]	.0534

Here $\widehat{\alpha}^D_{i,i}$ is the empiricals significance level for Dixon's cell-specific test for cell (i, i) for $i \in \{1, 2\}$, $\widehat{\alpha}_R$ is for the right-sided version of Pielou's test, $\widehat{\alpha}_L$ is for left-sided version of Pielou's test, $\widehat{\alpha}_P$ and $\widehat{\alpha}_{PY}$ are for Pielou's overall test of segregation without and with Yates' correction, respectively, $\widehat{\alpha}_D$ is for Dixon's overall test of segregation

[a] The empirical size is significantly smaller than .05; i.e., the test is conservative

[b] The empirical size is significantly larger than .05; i.e., the test is liberal

The empirical significance levels are presented in Table 2, where $\widehat{\alpha}^D_{i,i}$ is the empirical significance level for cell (i, i) with $i \in \{1, 2\}$, $\widehat{\alpha}_R$ and $\widehat{\alpha}_L$ are the estimated empirical significance levels for the right- and left-sided versions of Pielou's test, respectively

Table 3 The empirical significance levels for the tests under H_o : CSR independence with $N_{mc} = 10000$, $n_1, n_2 \in \{10, 30, 50, 100\}$ at $\alpha = .05$ for uniform class X and Y points in the unit square

(n_1, n_2)	Cell-specific				One-sided		Overall			
	$\widehat{\alpha}^D_{1,1}$	$\widehat{\alpha}^{D,qr}_{1,1}$	$\widehat{\alpha}^D_{2,2}$	$\widehat{\alpha}^{D,qr}_{2,2}$	$\widehat{\alpha}_R$	$\widehat{\alpha}_L$	$\widehat{\alpha}_P$	$\widehat{\alpha}_{PY}$	$\widehat{\alpha}_D$	$\widehat{\alpha}_{D,qr}$
(10, 10)	.0454[a]	.0360[a]	.0465	.0383[a]	.0844[b]	.1574[b]	.1280[b]	.0608[b]	.0432[a]	.0470
(10, 30)	.0306[a]	.0306[a]	.0485	.0427[a]	.0846[b]	.1399[b]	.1429[b]	.0542[b]	.0440[a]	.0411[a]
(10, 50)	.0270[a]	.0270[a]	.0464	.0323[a]	.0947[b]	.0574[b]	.0664[b]	.0318[a]	.0482	.0497
(30, 10)	.0479	.0415[a]	.0275[a]	.0275[a]	.0760[b]	.1406[b]	.1383[b]	.0506	.0390[a]	.0402[a]
(30, 30)	.0507	.0577[b]	.0505	.0578[b]	.0803[b]	.1115[b]	.1339[b]	.0836[b]	.0464	.0492
(30, 50)	.0590[b]	.0591[b]	.0522	.0549[b]	.0821[b]	.1211[b]	.1319[b]	.0834[b]	.0454[a]	.0411[a]
(50, 10)	.0524	.0346[a]	.0263[a]	.0263[a]	.0955[b]	.0544[b]	.0654[b]	.0310[a]	.0529	.0510
(50, 30)	.0535	.0554[b]	.0597[b]	.0597[b]	.0829[b]	.1173[b]	.1275[b]	.0805[b]	.0429[a]	.0405[a]
(50, 50)	.0465	.0456[a]	.0469	.0459[a]	.0804[b]	.1041[b]	.1397[b]	.0999[b]	.0508	.0528
(50, 100)	.0601[b]	.0652[b]	.0533	.0535	.0921[b]	.1090[b]	.1223[b]	.0938[b]	.0560[b]	.0556[b]
(100, 50)	.0490	.0493	.0571[b]	.0620[b]	.0909[b]	.1063[b]	.1190[b]	.0904[b]	.0483	.0495
(100, 100)	.0493	.0491	.0463[a]	.0455	.0927[b]	.1092[b]	.1324[b]	.1076[b]	.0504	.0513

Here $\widehat{\alpha}^{D,qr}_{i,i}$ is the empirical significance level for the QR-adjusted cell-specific test for (i, i) for $i \in \{1, 2\}$, and $\widehat{\alpha}_{D,qr}$ is for QR-adjusted Dixon's overall test of segregation
Other empirical size estimate notation and superscript labeling are as in Table 2

(see Eq. 2), $\widehat{\alpha}_P$ and $\widehat{\alpha}_{PY}$ are for Pielou's overall test of segregation without and with Yates' correction, respectively, $\widehat{\alpha}_D$ is for Dixon's overall test of segregation with $n_1, n_2 \in \{10, 30, 50, 100\}$ and $N_{mc} = 10000$. Notice that among the cell-specific tests only $\widehat{\alpha}^D_{1,1}$ and $\widehat{\alpha}^D_{2,2}$ are presented in Table 3, since $N_{12} = n_1 - N_{11}$ and $N_{21} = n_2 - N_{22}$ which implies $\widehat{\alpha}^D_{1,1} = \widehat{\alpha}^D_{1,2}$ and $\widehat{\alpha}^D_{2,1} = \widehat{\alpha}^D_{2,2}$ in the two-class case. The empirical sizes significantly smaller (larger) than .05 are marked with a(b) which indicate that the corresponding test is conservative (liberal). The asymptotic normal approximation to proportions is used in determining the significance of the deviations of the empirical size estimates from the nominal level of .05. For these proportion tests, we also use $\alpha = .05$ to test against empirical size being equal to .05. With $N_{mc} = 10000$, empirical sizes less (greater) than .0464 (.0536) are deemed conservative (liberal) at $\alpha = .05$ level. Observe that Dixon's cell-specific tests are slightly liberal or conservative or about the desired significance levels in rejecting H_o : RL when $n_1, n_2 \geq 30$. When $n_i \leq 10$ for $i = 1$ or 2, then Dixon's cell-specific tests tend to be conservative if $n_1 \neq n_2$ and liberal otherwise. Notice also that when $n_i \leq 10$ for $i = 1$ or 2 and $n_1 \neq n_2$ Dixon's cell-specific test is more conservative for cell (1, 1) which corresponds to the class with smaller size (i.e., class X) compared to class Y. On the other hand, Pielou's overall test and the right sided version of Pielou's test are extremely liberal for all sample size combinations; left sided version of Pielou's test is extremely liberal for all sample size combinations except for (10, 50). Furthermore, Pielou's test with Yates' correction is liberal when $\min(n_1, n_2) \geq 30$, conservative for (10, 50) and liberal or about the nominal level for other sample size combinations. Notice also that, $\widehat{\alpha}_{PY}$ values are significantly smaller (based on the tests of equality of the proportions for two populations) compared to $\widehat{\alpha}_P$ values. Dixon's overall test of segregation tends to be conservative for (10, 10) and (10, 30) and is about the desired nominal level for most of the other sample size combinations. These results suggest that under RL, Dixon's tests (especially the overall test) are appropriate, but Pielou's tests are not.

5.2 Empirical significance levels of the NNCT-tests under CSR independence

Under CSR independence, at each of $N_{mc} = 10000$ replicates, we generate points iid from $\mathcal{U}((0, 1) \times (0, 1))$, for some combinations of $n_1, n_2 \in \{10, 30, 50, 100\}$. Let $X = \{X_1, \ldots, X_{n_1}\}$ be the set of class 1 points and $Y = \{Y_1, \ldots, Y_{n_2}\}$ be the set of class 2 points.

We present the empirical significance levels for Pielou's tests and Dixon's tests in Table 3, where $\widehat{\alpha}^{D,qr}_{i,i}$ is Dixon's cell-specific test for cell (i, i) and $\widehat{\alpha}_{D,qr}$ is Dixon's overall test with Q and R are replaced with their (empirical) expected values and notations for the other tests are as in Sect. 5.1. The empirical sizes are calculated for some combinations of $n_1, n_2 \in \{10, 30, 50, 100\}$ and $N_{mc} = 10000$. Observe that Dixon's cell-specific tests tend to be slightly liberal or conservative or about the desired significance level for most sample size combinations. In particular, they have about the desired nominal level for $n_1 = n_2 \geq 30$; are extremely conservative for the smaller class when $n_i = 10$ for one of $i = 1, 2$ (see, e.g., $\widehat{\alpha}^D_{1,1}$ for $(n_1, n_2) = (10, 50)$ and $\widehat{\alpha}^D_{2,2}$ for $(n_1, n_2) = (50, 10)$). The QR-adjusted versions of cell-specific tests tend to have different sizes than the uncorrected versions when $n_i \leq 30$ for $i = 1$ or 2. For larger samples, QR-adjustment does not improve the sizes compared to the uncorrected ones. Pielou's overall test and one-sided versions are extremely liberal for all sample sizes (however, notice that Pielou's overall and left-sided tests are least liberal for $(n_1, n_2) = (10, 50)$ and $(n_1, n_2) = (50, 10)$). Pielou's test with Yates' correction is at the nominal level for $(n_1, n_2) = (30, 10)$, conservative for $(n_1, n_2) = (10, 50)$ and $(n_1, n_2) = (50, 10)$, and liberal for other sample size combinations. Notice also that $\widehat{\alpha}_{PY}$ values are significantly smaller than $\widehat{\alpha}_P$ values (i.e., Yates' correction significantly reduces the empirical size for Pielou's overall test). As for Dixon's overall test, it tends to be conservative for small sample size combinations, and is about the desired level for most large sample size combinations. As in the cell-specific tests, QR-adjustment does not improve on the uncorrected versions. For more detail on QR-adjustment for NNCT-tests, see (Ceyhan 2008). Hence *in the following sections, we only provide the uncorrected versions of Dixon's tests*.

Remark 5.1 Proportion of agreement between Pielou's and Dixon's overall tests: At each sample size combination under the RL Cases (2–4) and CSR independence, we also record the number of times both Pielou's and Dixon's overall tests simultaneously yield significant results at $\alpha = .05$. The ratio of number of significant results by both tests to the number of Monte Carlo replications, N_{mc}, is the proportion of agreement between the tests in rejecting the particular null pattern. That is, for example the proportion of agreement between Pielou's and Dixon's overall tests denoted by $\widehat{\alpha}_{P,D}$ for $(n_1, n_2) = (10, 10)$ under RL Case (2) is calculated as $\widehat{\alpha}_{P,D} := \sum_{i=1}^{N_{mc}} \mathbf{I}(\mathcal{X}^2_{P,i} \geq \chi^2_1(0.95))\mathbf{I}(\mathcal{X}^2_{D,i} \geq \chi^2_2(0.95))$ where $\mathcal{X}^2_{P,i}$ is the value of Pielou's

Table 4 The proportion of agreement between Pielou's and Dixon's overall segregation tests $\widehat{\alpha}_{P,D}$ for rejecting the RL Cases (2–4) and CSR independence with $N_{mc} = 10000$, $n_1, n_2 \in \{10, 30, 50, 100\}$ at $\alpha = .05$

(n_1, n_2)	(10, 10)	(10, 30)	(10, 50)	(30, 30)	(30, 50)	(50, 50)	(50, 100)	(100, 100)
RL Case (2)	.0255	.0245	.0344	.0317	.0275	.0321	.0302	.0336
RL Case (3)	.0226	.0189	.0321	.0323	.0275	.0331	.0289	.0320
RL Case (4)	.0208	.0192	.0335	.0329	.0250	.0292	.0298	.0355
CSR independence	.0277	.0242	.0340	.0296	.0270	.0321	.0342	.0314

overall test statistic for iteration i. The estimates of the proportion of agreement values are presented in Table 4. Observe that $\widehat{\alpha}_{P,D}$ values are significantly smaller than $\min(\widehat{\alpha}_P, \widehat{\alpha}_D) = \widehat{\alpha}_D$ at each sample size combination under each null case. This supports the discussion in the first paragraph of Sect. 4; that is, the rejection regions (hence acceptance regions) for both tests are significantly different and neither one is uniformly superior to the other. □

5.3 Empirical significance levels of the tests under indenpendence of rows and cell counts in NNCTs

For the independence of rows and cell counts in the NNCTs, we consider two cases: overall multinomial and row-wise binomial frameworks. In the overall multinomial case, we generate all four cell counts using multinomial distribution, $\mathscr{M}(n, \widetilde{\pi}_{11}, \widetilde{\pi}_{12}, \widetilde{\pi}_{21}, \widetilde{\pi}_{22})$, with $\widetilde{\pi}_{11} = \widetilde{\pi}_{21} = \frac{n_1}{2(n_1+n_2)}$ and $\widetilde{\pi}_{12} = \widetilde{\pi}_{22} = \frac{n_2}{2(n_1+n_2)}$. The NNCT constructed in such a way is (approximately) equivalent to one based on a random sample of (base, NN) pairs. In the row-wise binomial case, we generate the two cell counts in each row using binomial distribution, $N_{11} \sim \text{BIN}(n_1, n_1/(n_1 + n_2))$, and $N_{21} \sim \text{BIN}(n_2, n_1/(n_1 + n_2))$. The NNCT constructed in this way is also equivalent to one based on a random sample of (base, NN) pairs.

For such NNCTs, we can only compute Pielou's test and the one-sided versions, but not Dixon's tests, since Dixon's tests require more information on the NN structure in the spatial distribution of the points, e.g., the quantities such as Q and R, which are not available these cases. That is, in any spatial allocation of points, the NN relations of each point is dependent on the relations of neighboring points, which in turn implies that it is not realistic to have a random sample of (base, NN) pairs in practice.

In Table 5, we present the empirical significance levels for all tests except Dixon's tests under the independence of cells and rows case. Observe that Pielou's test with Yates' correction is extremely conservative under both of the frameworks. On the other hand Pielou's one-sided tests and Pielou's test without Yates' correction have about the desired nominal level.

Remark 5.2 Main result of Monte Carlo simulations for empirical sizes: Based on the simulation results under CSR independence of the points, we recommend the disuse of Pielou's test in practice, as it is extremely liberal, hence it might give false alarms when the pattern is actually not significantly different from RL or CSR independence. Moreover Yates' correction does not seem to fix the problems with Pielou's test, since

Table 5 The empirical significance levels for the tests under independence of cells and rows with $N_{mc} = 10000, n_1, n_2 \in \{10, 30, 50, 100\}$ at $\alpha = .05$ for contingency tables based on the overall multinomial and row-wise binomial frameworks

(n_1, n_2)	Overall multinomial				Row-wise binomial			
	One-sided		Overall		One-sided		Overall	
	$\widehat{\alpha}_R$	$\widehat{\alpha}_L$	$\widehat{\alpha}_P$	$\widehat{\alpha}_{PY}$	$\widehat{\alpha}_R$	$\widehat{\alpha}_L$	$\widehat{\alpha}_P$	$\widehat{\alpha}_{PY}$
(10, 10)	.0542[b]	.0521	.0415[a]	.0102[a]	.0603[b]	.0612[b]	.0426[a]	.0124[a]
(10, 30)	.0535	.0560[b]	.0525	.0213[a]	.0575[b]	.0495	.0510	.0174[a]
(10, 50)	.0489	.0534	.0556[b]	.0261[a]	.0864[b]	.0000[a]	.0468	.0152[a]
(30, 10)	.0559[b]	.0511	.0500	.0187[a]	.0575[b]	.0453[a]	.0504	.0186[a]
(30, 30)	.0470	.0445[a]	.0500	.0268[a]	.0487	.0491	.0518	.0269[a]
(30, 50)	.0493	.0504	.0492	.0272[a]	.0537[b]	.0541	.0554[b]	.0263[a]
(50, 10)	.0579[b]	.0440[a]	.0551[b]	.0254[a]	.0707[b]	.0205[a]	.0410[a]	.0132[a]
(50, 30)	.0504	.0493	.0492	.0272[a]	.0539[b]	.0531	.0526	.0274[a]
(50, 50)	.0483	.0487	.0534	.0344[a]	.0451[a]	.0466	.0592[b]	.0357[a]
(50, 100)	.0495	.0493	.0503	.0336[a]	.0509	.0487	.0488	.0333[a]
(100, 50)	.0493	.0495	.0503	.0336[a]	.0512	.0460[a]	.0510	.0350[a]
(100, 100)	.0499	.0500	.0537[a]	.0379[a]	.0528	.0490	.0538[b]	.0378[a]

The empirical size notation and superscript labeling are as in Table 2

the problems are not caused by the discrete nature of the cell counts. However, Yates' correction seems to improve the performance of Pielou's test, in the sense that empirical size of Pielou's test with Yates' correction gets closer to the nominal level compared to the uncorrected one. Even Dixon's tests fail to have the desired level when at least one sample size is small so that the cell count(s) in the corresponding NNCT have a high probability of being ≤ 5. This usually corresponds to the case that at least one sample size is ≤ 10 or the sample sizes (i.e., relative abundances) are very different in our simulation study. When sample sizes are small (hence the corresponding NNCT cell counts are ≤ 5), the asymptotic approximation of Dixon's tests is not appropriate. So Dixon (1994) recommends Monte Carlo randomization for his test when some NNCT cell count(s) are ≤ 5 under RL. We concur with the same recommendation for the RL pattern and extend this recommendation for CSR independence. In fact, this recommendation is also partly consistent with the inapplicability of asymptotic results for contingency tables in general (not just for NNCTs) when cell counts are too small. In general contingency tables, the chi-squared approximation seems to be valid in most cases if all expected cell counts are larger than 0.5 and at least half are greater than 1.0 (Conover 1999). On the other hand, Cochran (1952) states that the approximation may be poor if any expected cell count is less than 1 or if more than 20% of the expected cell counts are less than 5. □

6 Edge correction for the CSR independence pattern

In this section, we investigate the edge or boundary effects on the NNCT-tests used under CSR independence. Edge effects arise because CSR independence assumes an unbounded region, which is not the case in practical situations. Edge effects on spatial pattern analysis and various correction methods are discussed extensively in spatial pattern literature (Clark and Evans 1954; Cressie 1993). However, the effectiveness of edge correction depends on the type of the statistic used (see, e.g., Yamada and Rogersen 2003). For example, when the study region is rectangular, the edge effects can be minimized by including a *buffer zone* or *area* around the rectangle, or alternatively, the rectangular region is transformed into a *torus* (Dixon 2002b). In literature, buffer area is also referred to as *guard area* (Yamada and Rogersen 2003). The general idea is the same for buffer zone and toroidal edge corrections, but they are implemented in different ways. In buffer zone correction we assume the properties of the process are the same, even if we continue into the buffer area. In toroidal correction, the process is *exactly* the same outside of the study area.

Without any edge correction the cell counts in a NNCT can be written as

$$N_{ij} = \sum_{k=1}^{n} \sum_{l=1}^{n} w_{kl} S_{kl} \mathbf{I}(l \neq k) = \sum_{k,l=1}^{n} w_{kl} S_{kl} \mathbf{I}(l \neq k)$$

where S_{kl} is 1 if point l is of class j and point k is of class i, and 0 otherwise; w_{kl} is 1 if point l is the NN of point k, and 0 otherwise.

The quantities Q and R can be written as

$$Q = \sum_{m,k,l=1}^{n} w_{kl} w_{ml} \mathbf{I}(m \neq k \neq l) \text{ and } R = \sum_{k,l=1}^{n} w_{kl} w_{lk} \mathbf{I}(l \neq k),$$

with the understanding that $\mathbf{I}(m \neq k \neq l) = \mathbf{I}(m \neq k)\mathbf{I}(m \neq l)\mathbf{I}(k \neq l)$.

6.1 Buffer zone correction for the CSR independence pattern

In the *buffer zone correction method*, a guard area is selected inside or outside the study region and the points in the guard area are used only as destinations (not the base points) in NN relations. In the NN pair (U, V), point U is the *base point*, and point V is the *destination point*. When the buffer area is sufficiently large, the edge effects can be completely eliminated, but this is a wasteful procedure, because the large buffer area may contain many observations.

Let R_O be the original study area, R_B be the outer buffer area, and R_b be the inner buffer area and let n be the number of points that fall in R_O. In the outer buffer zone correction, let n_B be the number of points that fall in $R_O \cup R_B$, and points with indices $1, 2, \ldots, n$ lie in R_O, and points with indices $(n+1), (n+2), \ldots, n_B$ lie in R_B. With the outer buffer zone correction, the NNCT cell counts are

$$N_{ij} = \sum_{k=1}^{n} \sum_{l=1}^{n_B} w_{kl} S_{kl} \mathbf{I}(k \neq l).$$

Furthermore, the quantities Q and R are also modified as follows:

$$Q = \sum_{m,k=1}^{n_B} \sum_{l=1}^{n} w_{kl} w_{ml} \mathbf{I}(m \neq k \neq l) \text{ and}$$
$$R = \sum_{k,l=1}^{n_B} w_{kl} w_{lk} \mathbf{I}(k \neq l) - \sum_{k,l=(n+1)}^{n_B} w_{kl} w_{lk} \mathbf{I}(k \neq l).$$

In the inner buffer zone correction, let n_b be the number of points in the inner buffer area R_b, and points with indices $1, 2, \ldots, (n - n_b)$ lie in $R_O \setminus R_b$, and points with indices $(n - n_b + 1), (n - n_b + 2), \ldots, n$ lie in R_b. Then, with the inner buffer zone correction, the NNCT cell counts are

$$N_{ij} = \sum_{k=1}^{n-n_b} \sum_{l=1}^{n} w_{kl} S_{kl} \mathbf{I}(k \neq l).$$

Furthermore, the quantities Q and R are

$$Q = \sum_{m,k=1}^{n} \sum_{l=1}^{n-n_b} w_{kl} w_{ml} \mathbf{I}(m \neq k \neq l) \text{ and}$$
$$R = \sum_{k,l=1}^{n} w_{kl} w_{lk} \mathbf{I}(k \neq l) - \sum_{k,l=(n-n_b+1)}^{n_B} w_{kl} w_{lk} \mathbf{I}(k \neq l).$$

In our simulation study, we consider the edge correction by outer buffer zone correction only. For CSR independence, we generate points iid from $\mathcal{U}((-.5, 1.5) \times (-.5, 1.5))$ until there are n_1 class X points and n_2 class Y points in the unit square for some combinations of $n_1, n_2 \in \{10, 30, 50, 100\}$. We repeat this procedure $N_{mc} = 10000$ times for each n_1, n_2 combination. The corresponding empirical significance levels are provided in Table 6. Observe that compared to the uncorrected sizes, with the (outer) buffer zone edge correction, the empirical sizes of the right-sided version of Pielou's test do not significantly change; empirical sizes of the left-sided version of Pielou's test significantly decrease for smaller samples, do not significantly change for other sample size combinations; empirical sizes of Pielou's test with Yates' correction do not significantly change (except for (10, 10)); empirical sizes of Dixon's cell-specific tests do not change for most sample size combinations. On the other hand, for Dixon's test, the empirical sizes significantly increase to become liberal for $n_1 + n_2 \leq 80$. Furthermore Pielou's overall and one-sided tests still tend to be liberal with the (outer) buffer zone correction; Pielou's test with Yates' correction is liberal when $n_1 + n_2 \geq 40$. Dixon's cell-specific tests become about the desired level when $n_i \geq 50$ for both $i = 1, 2$, the sizes do not change or improve in one direction for smaller samples.

Remark 6.1 Inner versus outer buffer zone correction: The main difference between inner and outer buffer zone correction is the time of the selection of the buffer zone.

Table 6 The empirical significance levels for the overall tests under H_o with $N_{mc} = 10000$, $n_1, n_2 \in \{10, 20, 30, 40, 50\}$ at $\alpha = .05$ for uniform class X and Y points in the unit square when edge correction with buffer zone is applied

(n_1, n_2)	Cell-specific		One-sided		Overall		
	$\widehat{\alpha}^D_{1,1}$	$\widehat{\alpha}^D_{2,2}$	$\widehat{\alpha}_R$	$\widehat{\alpha}_L$	$\widehat{\alpha}_P$	$\widehat{\alpha}_{PY}$	$\widehat{\alpha}_D$
(10, 10)	.0464[e]	.0442[a,e]	.0811[b,e]	.1402[b,c]	.1169[b,c]	.0506[c]	.0604[b,d]
(10, 30)	.0317[a,e]	.0602[b,d]	.0830[b,e]	.1296[b,c]	.1283[b,c]	.0530[e]	.0572[b,d]
(10, 50)	.0258[a,e]	.0526[d]	.0946[b,e]	.0502[c]	.0625[b,e]	.0294[a,e]	.0569[b,d]
(30, 10)	.0565[b,d]	.0289[a,e]	.0799 [b,e]	.1273[b,c]	.1235[b,c]	.0485[e]	.0540[b,d]
(30, 30)	.0518[e]	.0522[e]	.0782[b,e]	.1104[b,e]	.1284[b,e]	.0848[b,e]	.0545[b,d]
(30, 50)	.0608[b,e]	.0554[b,e]	.0886[b,e]	.1088[b,c]	.1253[b,e]	.0769[b,e]	.0521[d]
(50, 10)	.0537[b,e]	.0269[a,e]	.0959[b,e]	.0553[b,e]	.0669[b,e]	.0297[b,e]	.0599[b,d]
(50, 30)	.0573[b,e]	.0640[b,e]	.0870[b,e]	.1122[b,e]	.1257[b,e]	.0822[b,e]	.0540[b,d]
(50, 50)	.0463[a,e]	.0471[e]	.0820[b,e]	.0998[b,e]	.1330[b,e]	.0946[b,e]	.0510[e]
(50, 100)	.0513[c]	.0519[e]	.0939[b,e]	.1040[b,e]	.1209[b,e]	.0905[b,e]	.0544[b,e]
(100, 50)	.0527[e]	.0497[c]	.0906[b,e]	.1038[b,e]	.1211[b,e]	.0903[b,e]	.0481[e]
(100, 100)	.0473[e]	.0475[e]	.0922[b,e]	.1069[b,e]	.1314[b,e]	.1039[b,e]	.0493[e]

The empirical size notation is as in Table 2
[a] The empirical size is significantly smaller than .05; i.e., the test is conservative
[b] The empirical size is significantly larger than .05; i.e., the test is liberal
[c] Empirical size significantly smaller than the uncorrected size
[d] Empirical size significantly larger than the uncorrected size
[e] Empirical size not significantly different from the uncorrected size

In the outer buffer zone correction, a larger region than the intended region of interest is selected prior to recording the observations; while in the inner buffer zone correction some part of the original study region is designated as the buffer zone after the data is collected. Hence, theoretically the inner and outer buffer zones behave similarly. Indeed, $R_O \setminus R_b$ in the inner buffer zone correction acts like the original region R_O of the outer buffer zone correction, and likewise R_b in inner buffer zone correction acts like R_B of the outer buffer zone correction. Thus, we only simulate the outer buffer zone correction, as it can also be equivalently viewed as the inner buffer zone correction. That is, the square $(-.5, 1.5) \times (-.5, 1.5)$ can be viewed as R_O and $[(-.5, 1.5) \times (-.5, 1.5)] \setminus [(0, 1) \times (0, 1)]$ can be viewed as R_b. □

6.2 Toroidal edge correction for the CSR independence pattern

In the *toroidal edge correction*, the original area is surrounded by eight copies of the original study area and the points in these additional copies are used only as destination points. For the toroidal edge correction, clusters around the boundaries might cause bias. Moreover, while toroidal correction applies only to rectangular study regions, buffer zone correction applies to any type of study region (see Yamada and Rogersen 2003).

Let R_O be the original study area, R_T be the eight copies appended to R_O so as to obtain NN structure for the points in R_O as if R_O is part of a torus. For the toroidal correction, let n_T be the number of points that fall in $R_O \cup R_T$, and points with indices $1, 2, \ldots, n$ lie in R_O, and points with indices $(n+1), (n+2), \ldots, n_T$ lie in the toroidal area R_T. With the toroidal correction, the NNCT cell counts are

$$N_{ij} = \sum_{k=1}^{n} \sum_{l=1}^{n_T} w_{kl} S_{kl} \mathbf{I}(k \neq l).$$

Furthermore, the quantities Q and R are also modified as follows:

$$Q = \sum_{m,k=1}^{n_T} \sum_{l=1}^{n} w_{kl} w_{ml} \mathbf{I}(m \neq k \neq l), \text{ and}$$
$$R = \sum_{k,l=1}^{n_T} w_{kl} w_{lk} \mathbf{I}(k \neq l) - \sum_{k,l=(n+1)}^{n_T} w_{kl} w_{lk} \mathbf{I}(k \neq l).$$

For toroidal correction, under H_o : CSR independence, we generate n_1 X-points and n_2 Y-points iid from $\mathcal{U}((0, 1) \times (0, 1))$ for some combinations of $n_1, n_2 \in \{10, 30, 50, 100\}$. We repeat this procedure $N_{mc} = 10000$ times for each n_1, n_2 combination. The corresponding empirical significance levels are presented in Table 7. Observe that toroidal edge correction does not significantly affect the empirical sizes of the NNCT-tests.

Remark 6.2 Main result of edge correction analysis: The Monte Carlo analysis in Sect. 6 suggests that the empirical sizes of the NNCT-tests are not affected by the toroidal edge correction because in our Monte Carlo simulations, we have generated the CSR independence pattern on the unit square. Any clusters in a realization of CSR are due to chance and are equally likely to occur anywhere, so the clusters are more likely to occur away from the boundary of the region. However, the (outer) buffer zone edge correction method seems to have stronger influence on the tests compared to toroidal correction. In particular, the empirical sizes of the Dixon's test tend to significantly increase with buffer zone correction. But for all other tests, buffer zone correction does not change the sizes significantly for most sample size combinations.

This is in agreement with the findings of Barot et al. (1999) which says NN methods only require a small buffer area around the study region. A large buffer area does not help too much since one only needs to be able to see far enough away from an event to find its NN. Once the buffer area extends past the likely NN distances (i.e., about the average NN distances), it is not adding much helpful information for NNCTs. Furthermore, since buffer (inner or outer) zone correction methods are wasteful, and strongly depend on the size of the zone, we do not recommend their use for NNCT-tests. On the other hand, one can use toroidal edge correction, but the gain might not be worth the effort. □

Table 7 The empirical significance levels for the overall tests under H_o with $N_{mc} = 10000$, $n_1, n_2 \in \{10, 20, 30, 40, 50\}$ at $\alpha = .05$ for uniform class X and Y points in the unit square when toroidal edge correction is applied

(n_1, n_2)	Cell-specific		One-sided		Overall		
	$\widehat{\alpha}^D_{1,1}$	$\widehat{\alpha}^D_{2,2}$	$\widehat{\alpha}_R$	$\widehat{\alpha}_L$	$\widehat{\alpha}_P$	$\widehat{\alpha}_{PY}$	$\widehat{\alpha}_D$
(10, 10)	$.0414^{a,e}$	$.0430^{a,e}$	$.0782^{b,e}$	$.1531^{b,e}$	$.1285^{b,e}$	$.0620^{b,e}$	$.0413^{a,c}$
(10, 30)	$.0318^{a,e}$	$.0492^{e}$	$.0845^{b,e}$	$.1397^{b,e}$	$.1434^{b,e}$	.0536e	$.0383^{a,e}$
(10, 50)	$.0265^{a,e}$	$.0466^{e}$	$.0958^{b,e}$	$.0561^{b,e}$	$.0670^{b,e}$	$.0323^{a,e}$	.0490e
(30, 10)	$.0453^{a,e}$	$.0285^{a,e}$	$.0777^{b,e}$	$.1412^{b,e}$	$.1412^{b,e}$	$.0508^{e}$	$.0376^{a,e}$
(30, 30)	$.0494^{e}$	$.0474^{e}$	$.0800^{b,e}$	$.1147^{b,e}$	$.1338^{b,e}$	$.0871^{b,e}$	$.0447^{a,e}$
(30, 50)	$.0592^{b,e}$	$.0500^{e}$	$.0832^{b,e}$	$.1127^{b,e}$	$.1233^{b,e}$	$.0783^{b,e}$	$.0437^{a,e}$
(50, 10)	$.0481^{e}$	$.0267^{a,e}$	$.0969^{b,e}$	$.0549^{b,e}$	$.0660^{b,e}$	$.0309^{a,e}$	$.0499^{e}$
(50, 30)	$.0518^{e}$	$.0604^{b,e}$	$.0827^{b,e}$	$.1204^{b,e}$	$.1297^{b,e}$	$.0785^{b,e}$	$.0447^{a,e}$
(50, 50)	$.0457^{a,e}$	$.0444^{a,e}$	$.0804^{b,e}$	$.1026^{b,e}$	$.1401^{b,e}$	$.1009^{b,e}$	$.0454^{a,e}$
(50, 100)	$.0533^{c}$	$.0518^{e}$	$.0923^{b,e}$	$.1114^{b,e}$	$.1260^{b,e}$	$.0958^{b,e}$	.0528e
(100, 50)	$.0503^{e}$	$.0522^{e}$	$.0903^{b,e}$	$.1079^{b,e}$	$.1245^{b,e}$	$.0946^{b,c}$	$.0485^{e}$
(100, 100)	$.0487^{e}$	$.0465^{e}$	$.0954^{b,e}$	$.1069^{b,e}$	$.1344^{b,e}$	$.1077^{b,c}$	$.0467^{e}$

The empirical size notation and superscript labeling are as in Table 6

7 Examples

We illustrate the tests on two example data sets: Dixon's swamp tree data (Dixon 2002b) and an artificial data set. We present the corresponding NNCTs, test statistics, and edge correction results for both examples. Since an outer buffer zone is not provided in these examples, inner buffer zone correction is the only type of buffer zone correction we can apply. Moreover, since the regions are rectangular, we can also apply toroidal edge correction in these examples.

7.1 Swamp tree data

Dixon (2002b) illustrates NN-methods on tree species in a 50 m by 200 m rectangular plot of hardwood swamp in South Carolina, USA. The plot contains trees from 13 different tree species, of which we only consider the live trees from two species, namely, black gums and bald cypresses. If spatial interaction among the less frequent species were important, a more detailed 12×12 NNCT-analysis should be performed. For more detail on the data, see Dixon (2002b). The locations of these trees in the study region are plotted in Fig. 2.

Dixon (2002a) applies his methodology for this data set assuming the null pattern is the RL of tree species to the given locations. But it is more reasonable to assume that the locations of the tree species a priori result from different processes. Hence the more appropriate null hypothesis would be the CSR independence pattern, which implies that NNCT-test results are conditional ones. The question of interest is whether the two

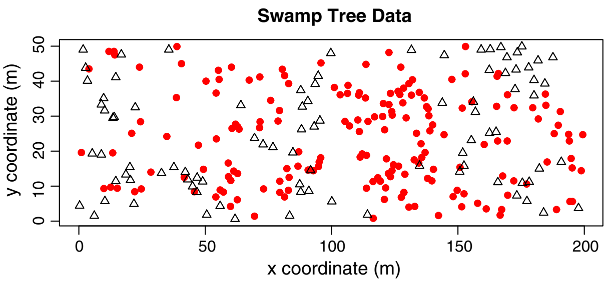


Fig. 2 The scatter plots of the locations of black gum trees (solid circles) and bald cypress trees (triangles)

Table 8 The NNCT for swamp tree data (top left) and the artificial data (bottom left) and the corresponding percentages (right). B.G. = black gums, B.C. = bald cypresses

Swamp tree data		NN			NN		
		B.G.	B.C.	Sum	B.G.	B.C.	Sum
Base	B.G.	149	33	182	82%	18%	67%
	B.C.	43	48	91	47%	53%	23%
	Sum	192	581	273	34%	66%	100%

Artificial data		NN			NN		
		X	Y	Sum	X	Y	Sum
Base	X	30	20	50	60%	40%	50%
	Y	19	31	50	38%	62%	50%
	Sum	49	51	100	49%	51%	100%

tree species are segregated, associated, or do not deviate from the CSR independence pattern. The corresponding NNCT and the percentages are provided in Table 8. The percentages for the cells are based on the sample size of each species. That is, for example 82% of black gums have NNs from black gums, and remaining NNs of black gums are from bald cypresses. The row and column percentages are marginal percentages with respect to the total sample size. The percentage values are also suggestive of segregation, especially for black gum trees.

For the raw data (i.e., data not corrected for edge effects), we find $Q = 178$ and $R = 156$. The test statistics are provided in Table 9, where $\mathcal{X}^2_D$ stands for Dixon's overall segregation test, $\mathcal{X}^2_P$ and $\mathcal{X}^2_{PY}$ for Pielou's test without and with Yates' correction, respectively, Z_n for the directional Z-test. The p-values are for the general alternative of deviation from CSR independence for $\mathcal{X}^2_D$, $\mathcal{X}^2_P$, and $\mathcal{X}^2_{PY}$; and for Z_n, the first p-value in the parenthesis is for the association alternative, while the second is for the segregation alternative. Observe that all two-sided tests are significant, implying significant deviation from CSR independence. The directional (one-sided) tests indicate that black gum trees and bald cypresses are significantly segregated.

Table 9 The values of the NNCT-test statistics and the corresponding p-values (in parenthesis) for the swamp tree data (top) and artificial data set (bottom)

Correction	Cell-specific		One-sided	Overall		
	Z_{11}^D	Z_{22}^D	Z_n	$\mathcal{X}_P^2$	$\mathcal{X}_{PY}^2$	$\mathcal{X}_D^2$
Test statistics and the associated p-values for swamp tree data						
None	4.47	3.54	5.90	34.84	33.20	23.77
	(<.0001)	(.0004)	($\approx$ 1, <.0001)	(<.0001)	(<.0001)	(<.0001)
Toroidal	4.31	3.31	5.62	31.60	30.04	21.29
	(<.0001)	(.0009)	($\approx$ 1, <.0001)	(<.0001)	(<.0001)	(<.0001)
Buffer zone	3.95	4.04	6.08	37.00	35.10	23.39
($k = 0$)	(.0001)	(.0001)	($\approx$ 1, <.0001)	(<.0001)	(<.0001)	(<.0001)
Buffer zone	3.61	4.08	5.90	34.77	32.86	21.92
($k = 1$)	(.0003)	(.0001)	($\approx$ 1, <.0001)	(<.0001)	(<.0001)	(<.0001)
Test statistics and the associated p-values for the artificial data						
None	1.38	1.64	2.20	4.84	4.00	3.36
	(.1670)	(.1000)	(.9861, .0139)	(.0278)	(.0455)	(.1868)
Toroidal	1.38	1.38	2.00	4.00	3.24	2.65
	(.1672)	(.1672)	(.9772, .0228)	(.0455)	(.0719)	(.2660)
Buffer zone	1.07	1.28	1.72	2.95	2.22	2.09
($k = 0$)	(.2843)	(.2004)	(.9572, .0428)	(.0857)	(.1365)	(.3511)
Buffer zone	0.61	0.58	0.81	0.66	0.32	0.54
($k = 1$)	(.5391)	(.5597)	(.7922, .2078)	(.4156)	(.5697)	(.7637)

$\mathcal{X}_D^2$: Dixon's overall segregation test, $\mathcal{X}_P^2$ and $\mathcal{X}_{PY}^2$: Pielou's test without and with Yates' correction, respectively, Z_n: directional Z-test. The p-values are for the general alternative of any deviation from CSR independence except for Z_n, for which the first p-value in the parenthesis is for the association alternative, while the second is for the segregation alternative

Table 9 also contains the p-values when the edge correction methods are applied. The toroidal correction does slightly change the test statistics, but not the conclusions. For the inner buffer zone correction, we first estimate the density of the combined species, namely $\widehat{\lambda} = (n_1 + n_2)/A_r$ where A_r is the area of the study region. Let W be the distance from a randomly chosen event to its NN, then under CSR independence, $\mathbf{E}[W] = 1\Big/\left(2\sqrt{\lambda}\right)$ and $\mathbf{Var}[W] = (4-\pi)/(4\,\pi\,\lambda)$, where λ is the intensity of the point process (Dixon 2002b). So, we move the boundaries inside the rectangular study area by $\mathbf{E}[W] + k\sqrt{\mathbf{Var}[W]} = 1\Big/\left(2\sqrt{\widehat{\lambda}}\right) + k\sqrt{(4-\pi)\Big/\left(4\,\pi\,\widehat{\lambda}\right)}$ where k determines how much of the study region is regarded as the buffer zone. We implement the inner buffer zone correction with $k = 0, 1, 2, 3$, but only present the results for $k = 0, 1$, since each case yields similar test statistics and the same conclusions.

Based on the NNCT-tests, we conclude that the tree species exhibit significant deviation from the CSR independence pattern. Considering Fig. 2 and the corresponding NNCT in Table 8, this deviation is toward the segregation of the tree species. However, the results of NNCT-tests pertain to small scale interaction, i.e., at about the average

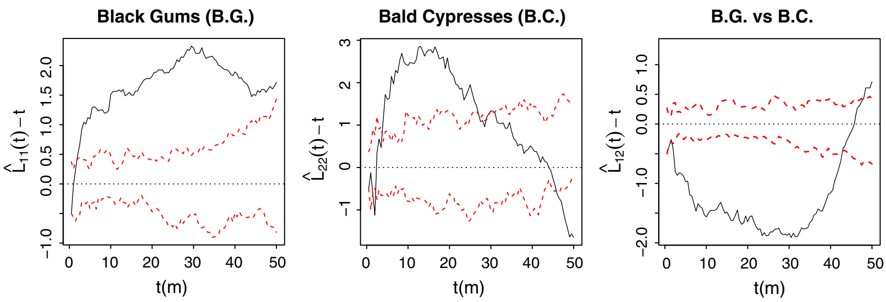


Fig. 3 Second-order properties of swamp tree data. Functions plotted are Ripley's univariate L-functions $\widehat{L}_{ii}(t) - t$ for $i = 1, 2$, and bivariate L-function $\widehat{L}_{12}(t) - t$ where $i = 1$ for black gums and $i = 2$ for bald cypresses. The thick dashed lines around 0 are the upper and lower 95% confidence bounds for the L-functions based on Monte Carlo simulation under the CSR independence pattern. Notice also that vertical axes are differently scaled

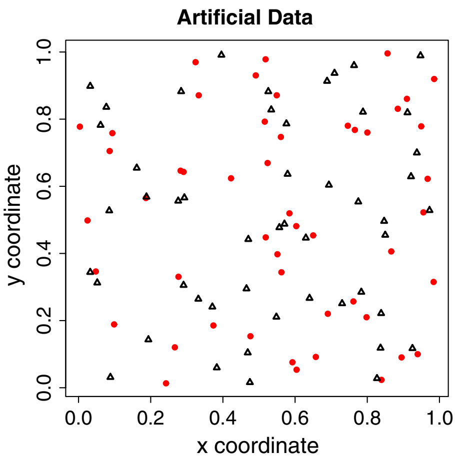
NN distances. To understand (possible) causes of the segregation and the type and level of interaction between the tree species at different scales (i.e., distances between the trees) we also provide Ripley's univariate and bivariate L-functions in Fig. 3, where the spatial interaction is analyzed for distances up to 50 m.

In Fig. 3, we present the plots of Ripley's univariate L-function $\widehat{L}_{ii}(t) - t$ for both species, and bivariate L-function $\widehat{L}_{12}(t)$ for the pair of tree species. Due to the symmetry of $L_{ij}(t)$, we omit $\widehat{L}_{21}(t)$. We also present the upper and lower 95% confidence bounds for each $\widehat{L}_{ii}(t)-t$ and $\widehat{L}_{12}(t)-t$ under CSR independence. Observe that black gums exhibit significant aggregation for distances $t > 1$ m (the $L_{11}(t) - t$ curve is above the upper confidence bound); bald cypresses exhibit no deviation from CSR around $t \leqslant 5$ m, then they exhibit significant spatial aggregation for t up to 30 m, then for larger distances ($t > 45$ m) they exhibit spatial regularity. Observe also that black gums and bald cypresses are significantly segregated for distances up to $t \approx 42$ m ($\widehat{L}_{12}(t) - t$ is below the lower confidence bound), for larger distances their interaction does not deviate significantly from CSR independence. Therefore, the segregation of the species might be due to different levels and types of aggregation of the species in the study region. Note also that average NN distance for swamp tree data is 3.08 ± 1.70 (mean $\pm$ standard deviation) and results of bivariate L-function and NNCT-analysis agree for distances around $t = 3$ m.

7.2 Artificial data set

In the swamp tree example, although the expected NNCT cell counts (not presented) are different for the NNCT-tests and p-values for Dixon's tests are larger than others, we have the same conclusion: there is strong evidence for segregation of tree species. Below, we present an artificial example, a random sample of size 100 (with 50 X-points and 50 Y-points uniformly generated on the unit square). The question of interest is the spatial interaction between X and Y classes. We plot the locations of the points in Fig. 4 and the corresponding NNCT together with percentages are provided in Table 8.

Fig. 4 The scatter plots of the locations of X points (solid circles) and Y points (triangles)



Observe that the percentages are suggestive of mild segregation, with equal degree for both classes.

The data is generated to resemble the CSR independence pattern, so we assume the null pattern is CSR independence, which implies that our inference will be a conditional one. Observe that in Table 9, Pielou's tests are significant while Dixon's test are not, which might be interpreted as evidence of deviation from CSR independence. The graph in Fig. 4 is not suggestive of any such deviation from CSR independence, and the dependence between NNCT cell counts might be confounding the conclusions based on Pielou's tests. With toroidal correction, all p-values increase, but only for Pielou's test with Yates' correction becomes insignificant after the correction. With buffer zone correction with $k = 0$, all p-values get to be insignificant, except for the one-sided test for segregation. Furthermore, with $k = 1$, the changes in p-values are more dramatic. We also have similar changes with $k = 2, 3$ (not presented). Thus, inner buffer zone edge correction, might make a big difference if the pattern is a close call between CSR independence and segregation/association. That is, if a segregation test has a p-value about .05, when a subregion is reserved as the inner buffer zone, either we might have a pattern different from CSR independence in the area for the base points (i.e., $R_O \setminus R_b$) or after the loss of data in the buffer zone the power of the tests might decrease. We also point out that, inner and outer buffer zone correction methods are not recommended in literature for spatial pattern analysis with, e.g., Ripley's K-function (Yamada and Rogersen 2003), and we concur with this suggestion for NNCT-tests.

Since Pielou's and Dixon's tests give different results in terms of significance, we also provide Ripley's univariate and bivariate L-functions in Fig. 5, where the spatial interaction is analyzed for distances up to $t = 0.25$. Observe that X points exhibit significant regularity for distances $.18 < t < .24$ and no deviation from CSR for other distance values. Y points do not deviate significantly from CSR for all distances considered, although they are close to being regular. Observe also that X and Y points are significantly associated for distances around $t \approx 0.06$ and $t \approx 0.10$, for all other distances their interaction does not deviate significantly from CSR independence.

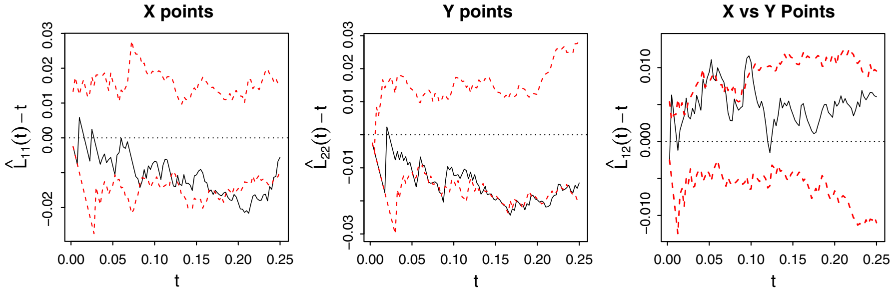


Fig. 5 Second-order properties of the artificial data. Functions plotted are Ripley's univariate L-functions $\widehat{L}_{ii}(t) - t$ for $i = 1, 2$, and bivariate L-functions $\widehat{L}_{12}(t) - t$ where $i = 1$ for X and $i = 2$ for Y points. The thick dashed lines around 0 are the upper and lower 95% confidence bounds for the L-functions based on Monte Carlo simulation under the CSR independence pattern. Notice also that vertical axes are differently scaled

Hence we conclude that the significant segregation implied by Pielou's test seems to be a false alarm, since in fact, the spatial interaction between the points is not significantly different from CSR independence. Note also that average NN distance for the artificial data is $.052 \pm .03$ and results of the bivariate L-function and NNCT-analysis agree around $t = .05$.

8 Discussion and conclusions

In this article we discuss segregation tests based on nearest neighbor contingency tables (NNCTs). These NNCT-tests include Pielou's test (with or without Yates' correction), Dixon's cell-specific and overall tests, and the newly introduced one-sided versions of Pielou's tests. A summary of the test statistics together with the underlying assumptions, the appropriate null hypotheses, and the quantities the statistics are conditional on are presented in Table 10.

Pielou's and Dixon's tests were both defined under the null hypothesis of random labeling (RL) of a fixed set of points (Pielou 1961; Dixon 1994, 2002a). It has been shown that Pielou's test is not appropriate for testing RL, but Dixon's tests are. The main problem with Pielou's NNCT-tests (including the one-sided versions) discussed in this article is the dependence between trials (i.e., categorization of (base, NN) pairs) and between the NNCT cell counts.

In this article, we extend the use of the NNCT-tests for the CSR of points from two classes (i.e., CSR independence). We demonstrate that Pielou's tests are liberal for rejecting RL or CSR independence, while Dixon's tests are appropriate for large samples. For smaller samples (i.e., when some cell count in the NNCT is ≤ 5) we recommend the Monte Carlo randomization version for NNCT-tests. We also show that Pielou's tests are only appropriate for a NNCT based on a random sample of (base, nearest neighbor (NN)) pairs (which is not a realistic situation in practice). We prove the consistency of the tests under appropriate null hypotheses; evaluate the empirical size performance of these tests based on an extensive Monte Carlo simulation study under RL, CSR independence, and independence of cell counts and rows in Sect. 5.

Table 10 The summary of underlying assumptions, appropriate null hypotheses H_o, and the type of conditioning for the tests based on NNCTs

Test statistic	Underlying assumption	Appropriate H_o	Conditional on	Empirical size under	
				RL or CSR	Independence of cell counts
$\mathcal{X}_P^2$ and Z_n under Poisson sampling framework	Cell counts are independent Poisson variates	$\widetilde{\pi}_{ij} = \widetilde{\nu}_i \widetilde{\kappa}_j$ for $i, j = 1, 2$	n	Size $>$ .05 (i.e., liberal)	Size $\approx$ .05 (i.e., appropriate)
$\mathcal{X}_P^2$ and Z_n under row-wise binomial sampling framework	Cell counts in each row have the same binomial distribution	$N_{ii} \sim \text{BIN}(n_i, \widetilde{\pi}_{ii})$ for $i = 1, 2$ with $(\widetilde{\pi}_{11}, \widetilde{\pi}_{12}) = (\widetilde{\pi}_{21}, \widetilde{\pi}_{22})$	n_i and n	Size $>$ .05 (i.e., liberal)	Size $\approx$ .05 (i.e., appropriate)
$\mathcal{X}_P^2$ and $\widetilde{Z}_n$ under overall multinomial sampling framework	Cell counts have a multi-nomial distribution	$(N_{11}, N_{12}, N_{21}, N_{22}) \sim \mathscr{M}(n, \widetilde{\pi}_{11}, \widetilde{\pi}_{12}, \widetilde{\pi}_{21}, \widetilde{\pi}_{22})$ with $(\widetilde{\pi}_{11}, \widetilde{\pi}_{12}) = (\widetilde{\pi}_{21}, \widetilde{\pi}_{22})$	n	Size $>$ .05 (i.e., liberal)	Size $\approx$ .05 (i.e., appropriate)
$\mathcal{X}_D^2$ and Z_{ij}^D under RL	Randomness in the NN structure	$\mathbf{E}[N_{ij}] = \frac{n_i(n_i-1)}{(n-1)}\mathbf{I}(i = j) + \frac{n_i n_j}{(n-1)}\mathbf{I}(i \neq j)$	n_i and n	Size $\approx$ .05 (i.e., appropriate)	NA
$\mathcal{X}_D^2$ and Z_{ij}^D under CSR independence	Randomness in the NN structure	$\mathbf{E}[N_{ij}] = \frac{n_i(n_i-1)}{(n-1)}\mathbf{I}(i = j) + \frac{n_i n_j}{(n-1)}\mathbf{I}(i \neq j)$	n_i, n, Q, and R	Size $\approx$ .05 (i.e., appropriate)	NA

$\mathcal{X}_P^2$ stands for Pielou's overall test, Z_n for one-sided version of Pielou's test, $\mathcal{X}_D^2$ for Dixon's overall test, and Z_{ij}^D for Dixon's cell-specific test. Pielou's overall test with Yates' correction has the same properties as the uncorrected one (hence not presented). NA stands for "not applicable"

Based on the Monte Carlo analysis, for moderate to large sample sizes Dixon's tests are recommended.

Under CSR independence, edge or boundary effects might be a potential concern for the NNCT-tests. Based on Monte Carlo analysis for edge correction methods, with buffer zone correction, we find that the uncorrected and corrected results are not significantly different. In particular, buffer zone correction methods (with larger distances than the average NN distances) are not recommended, as they are wasteful procedures and do not serve the purpose of correcting for the edge effects. The outer buffer zone correction with large buffer areas is redundant hence not worth the effort, while inner buffer zone correction with large buffer areas is wasteful and might cause bias and loss of power in the analysis. With toroidal correction only significant change occurs for Dixon's overall change, but corrected versions are usually liberal. Hence edge effects are not significantly confounding the results of the NNCT-tests (under CSR independence).

In this article, NNCTs are based on NN relations using the usual Euclidean distance. But, this method can be extended to the case that NN relation is based on dissimilarity measures between observations in finite or infinite dimensional space. It is even possible to have situations that are completely non-spatial and yet one can conduct NNCT analysis based on dissimilarity measures. In this general context the NN of object x refers to the object with the minimum dissimilarity to x. The extension of RL pattern is straightforward, but extra care should be taken for such an extension of CSR independence. For example, in either case, the term Q (in Dixon's tests) which is the number of points with shared NNs needs to be revised as $\widetilde{Q} := 2\sum_{k=1}^{N}\binom{k}{2}Q_k$. We conjecture that these tests when applied to other fields for high dimensional data and NN relations based on dissimilarity measures, can be useful.

Acknowledgements I would like to thank Prof. Philip M. Dixon and an anonymous referee for their comments and suggestions on earlier versions of this manuscript. Most of the Monte Carlo simulations presented in this article were executed on the Hattusas cluster of Koç University High Performance Computing Laboratory.

Appendix

Proof of Theorem 4.1 (I) Suppose under H_o, NNCT is based on a random sample of (base, NN) pairs. For the two-class case, we parametrize the segregation alternative as $H_a^S : \widetilde{\pi}_{11} = \widetilde{\pi}_{21} + \varepsilon$ for $\varepsilon \in (0, 1 - \widetilde{\pi}_{21})$. Then the hypotheses become $H_o : \varepsilon = 0$ and $H_a^S : \varepsilon > 0$. The rejection criterion in the theorem is equivalent to $Z_n > z(1-\alpha)$ where Z_n is defined as in Eq. 2. Then $\mathbf{E}[Z_n|H_o] = 0$ and $\mathbf{E}\left[Z_n|H_a^S\right] = \varepsilon > 0$.

In the row-wise binomial framework with $N_{11} \sim \text{BIN}(n_1, \widetilde{\pi}_{11})$ and $N_{21} \sim \text{BIN}(n_2, \widetilde{\pi}_{12})$ being independent, consider $T_n = (N_{11}/n_1 - N_{21}/n_2)$. Under H_o, $\mathbf{E}[T_n|H_o] = 0$ and under H_a^S, the expected value of T_n becomes $\mathbf{E}\left[T_n|H_a^S\right] = \varepsilon > 0$. Under both null and alternative hypotheses the test statistic $(T_n - \mathbf{E}[T_n])/\sqrt{\mathbf{Var}[T_n]}$ has normal distribution asymptotically. By an appropriate application of Slutsky's Theorem, $Z_n = (N_{11}/n_1 - N_{21}/n_2)\sqrt{\frac{n_1\, n_2\, n}{C_1\, C_2}}$ and $(T_n - \mathbf{E}[T_n])/\sqrt{\mathbf{Var}[T_n]}$ have the same asymptotic distribution. Hence the size of the test is α. Furthermore, consistency follows by the asymptotic normality. The consistency for spatial association for the two-class case follows similarly.

In the multinomial framework with $\mathbf{N} = (N_{11}, N_{12}, N_{21}, N_{22}) \sim \mathscr{M}(n, \widetilde{\nu}_1\widetilde{\kappa}_1, \widetilde{\nu}_1\widetilde{\kappa}_2, \widetilde{\nu}_2\widetilde{\kappa}_1, \widetilde{\nu}_2\widetilde{\kappa}_2)$, we parametrize the segregation alternative as $H_a^S : \mathbf{E}[N_{11}/n] = \widetilde{\nu}_1\,\widetilde{\kappa}_1 + \varepsilon$. Consider the test statistic $T_n := N_{11}/n - \widetilde{\nu}_1\,\widetilde{\kappa}_1$. Then expected value of T_n under H_o is $\mathbf{E}[T_n|H_o] = 0$ and $(T_n - \mathbf{E}[T_n])/\sqrt{\mathbf{Var}[T_n]}$ is approximately normal for large n under both null and alternative hypotheses. By an appropriate application of Slutsky's Theorem and some algebraic manipulations, one can see that $(T_n - \mathbf{E}[T_n])/\sqrt{\mathbf{Var}[T_n]}$ and $\widetilde{Z}_n$ given in Eq. 3 are asymptotically equivalent and converge in law to the standard normal distribution under H_o and the same normal distribution under H_a. So, the test has size α and using T_n, consistency of the test for H_a^S follows. The consistency for spatial association for the two-class case follows similarly.

(II) Under RL or CSR independence, $\mathbf{E}[T_n] = \frac{n_1 - 1}{n-1} - \frac{n_1}{n-1} = \frac{-1}{n-1}$ which converges to zero as $n \to \infty$. Let $\mathbf{Var}_{II}[T_n]$ be the variance of T_n under RL or CSR independence and $\mathbf{Var}_I[T_n]$ be the variance of T_n under the case that (base, NN) pairs are independent. Then our claim (which is proved below) that

$$\mathbf{Var}_{II}[T_n] > \mathbf{Var}_I[T_n] \tag{9}$$

for large n. Hence, T_n rejects RL or CSR independence in favor of segregation or spatial association more frequently than it should; i.e., it is liberal in rejecting these null patterns, or equivalently, its nominal significance level is larger than the prespecified level α. However, under the above parametrization of segregation, we have $\mathbf{E}\left[T_n|H_a^S\right] = \varepsilon > 0$ and $\mathbf{Var}_{II}[T_n]$ converges to zero as $n \to \infty$. Hence, consistency for segregation follows. Consistency for the spatial association alternative follows similarly. □

Proof of the Claim in Eq. 9 For large n, the probabilities in Eq. 7 take the form

$$\begin{aligned} p_{ii} &= \nu_i^2,\ p_{ij} = \nu_i\,\nu_j,\ p_{iii} = \nu_i^3,\ p_{iij} = \nu_i^2\,\nu_j, \\ p_{iijj} &= \nu_i^2\,\nu_j^2,\ p_{iiii} = \nu_i^4 \text{ for } i, j \in \{1, 2\}. \end{aligned}$$

Then variance of T_n under RL or CSR independence is

$$\mathbf{Var}_{II}[T_n] = \mathbf{Var}[N_{11}/n_1] + \mathbf{Var}[N_{21}/n_2] - 2\,\mathbf{Cov}[N_{11}/n_1, N_{21}/n_2].$$

Using Eq. 6 in Sect. 3.3.1 and Eq. 6 in (Dixon 2002a), we have,

$$\begin{aligned} \mathbf{Var}[N_{11}/n_1] &= \frac{1}{n^2\,\nu_1^2}\left[(n+R)\,\nu_1^2 + (2n - 2R + Q)\,\nu_1^3 + (n^2 - 3n - Q + R)\,\nu_1^4 - (n\,\nu_1^2)^2\right] \\ &= \frac{1}{n^2}\left[n + R + (2n - 2R + Q)\,\nu_1 + (R - 3n - Q)\,\nu_1^2\right] \end{aligned}$$

and

$$\mathbf{Var}[N_{21}/n_2] = \frac{1}{n^2\,\nu_2^2}\left[n\,\nu_1\,\nu_2 + Q\,\nu_2^2\,\nu_1 + (n^2 - 3\,n - Q + R)\,\nu_1^2\,\nu_2^2 - (n\,\nu_1\,\nu_2)^2\right]$$
$$= \frac{1}{n^2\,\nu_2}\left[n\,\nu_1 + Q\,\nu_1\,\nu_2 + (R - 3\,n - Q)\,\nu_1^2\,\nu_2\right]$$

and

$$\mathbf{Cov}[N_{11}/n_1, N_{21}/n_2]$$
$$= \frac{1}{n^2\,\nu_1\,\nu_2}\left[(n - R - Q)\,\nu_1^2\,\nu_2 + (n^2 - 3\,n - Q + R)\,\nu_1^3\,\nu_2 - n^2\,\nu_1^3\,\nu_2\right]$$
$$= \frac{1}{n^2}\left[(n - R - Q)\,\nu_1 + (R - 3\,n - Q)\,\nu_1^2\right].$$

Hence by algebraic manipulations, we get

$$\mathbf{Var}_{II}[T_n] = \frac{n+R}{n^2} + \frac{\nu_1}{n\,\nu_2}.$$

Furthermore, under RL or CSR independence, $\kappa_i = \nu_i$ for $i = 1, 2$. Then for large n, variance of T_n under RL or CSR independence is

$$\mathbf{Var}_I[T_n] = \frac{\pi_{11}(1-\pi_{11})}{n_1} + \frac{\pi_{21}(1-\pi_{21})}{n_2} = \frac{\nu_1\,\kappa_1(1-\nu_1\,\kappa_1)}{n\,\nu_1} + \frac{\nu_2\,\kappa_1(1-\nu_2\,\kappa_1)}{n\,\nu_2}$$
$$= \frac{\nu_1^2(1-\nu_1^2)}{n\,\nu_1} + \frac{\nu_2\,\nu_1(1-\nu_2\,\nu_1)}{n\,\nu_2} = \frac{\nu_1(1-\nu_1^2)}{n\,\nu_1} + \frac{\nu_1(1-\nu_2\,\nu_1)}{n\,\nu_2}$$
$$= \frac{\nu_1\,(2-\nu_1)}{n}.$$

Then we need to show that, $\frac{n+R}{n} + \frac{\nu_1}{\nu_2} = 1 + \frac{R}{n} + \frac{\nu_1}{\nu_2} > \nu_1\,(2-\nu_1)$ which trivially follows, since $1 > \nu_1\,(2-\nu_1)$ which follows from $(\nu_1 - 1)^2 > 0$. □

Proof of Theorem 4.2 (I) Suppose under H_o, NNCT is based on a random sample of (base, NN) pairs. In the two-class case, deviations from H_o are as in Theorem 4.1. With such a deviation from H_o; i.e., under H_a, for large n, $\mathcal{X}_P^2$ is approximately distributed as a χ^2 distribution with non-centrality parameter $\lambda(\varepsilon)$ and 1 degrees of freedom (d.f.), which is denoted as $\chi_1^2(\lambda(\varepsilon))$. The non-centrality parameter is a quadratic form which can be written as $\boldsymbol{\mu}(\varepsilon)'\,A\,\boldsymbol{\mu}(\varepsilon)$ for some positive definite 2×2 matrix A of rank 1 (see Moser 1996) hence $\lambda(\varepsilon) > 0$ under H_a. Then as $n \to \infty$, the null and alternative hypotheses are equivalent to $H_o : \lambda = 0$ versus $H_a : \lambda = \lambda(\varepsilon) > 0$. Then the size of the test is α and the consistency follows.
(II) Under RL or CSR independence, the dependence in the row sums or column sums, which causes the reduction in d.f. is not the only type of dependence present in the NNCT. In addition to this, the NNCT cell counts are also dependent due to

reflexivity and shared NN structure. Hence, even under RL or CSR independence, the corresponding distribution is still a scaled version of central χ^2 distribution, but has a larger variance than χ^2_1 distribution, hence the nominal level of the test is larger than the prespecified α. On the other hand, as $n \to \infty$, the power of the test goes to 1. □

Proof of Theorem 4.3 Consider the one-sided alternative with $H_a : \pi_{ij} > \nu_{ij}$. Let $T_n = N_{ij}/n - \nu_{ij}$. Then $(T_n - \mathbf{E}[T_n])/\sqrt{\mathbf{Var}[T_n]} = Z^D_{ij}$. Under RL, $\mathbf{E}[T_n] = 0$ and $\mathbf{Var}[T_n]$ is given in Eq. 6. Consider the parametrization of the alternative as H_a : $\pi_{ij} = \nu_{ij} + \varepsilon$ for $\varepsilon \in (0, 1 - \nu_{ij})$. Then under H_a, $\mathbf{E}[T_n|H_a] = \varepsilon > 0$. As $n_i \to \infty$, N_{ii} have asymptotic normal distribution (Cuzick and Edwards 1990). This implies N_{12} is also asymptotically normal, as $n_1 \to \infty$, since $N_{12} = n_1 - N_{11}$. Similarly, N_{21} has asymptotically normal distribution as $n_2 \to \infty$. Thus under both null and alternative hypothesis, $(T_n - \mathbf{E}[T_n])/\sqrt{\mathbf{Var}[T_n]}$ has asymptotic normal distribution. Then the size of the test is α and consistency follows. The consistency for the other types of alternatives follow similarly. □

Proof of Theorem 4.4 Under RL, $\mathbf{Y} = \mathbf{N} - \mathbf{E}[\mathbf{N}]$ is approximately distributed as $N(\mathbf{0}, \Sigma)$ for large n. Let Σ^- be the generalized inverse of Σ whose rank is 2. Then by Theorem 3.1.2 of Moser (1996), under H_o, $\mathbf{Y}'\Sigma^-\mathbf{Y} \sim \chi^2_2(\lambda = 0)$. Hence the test has size α. Consider any deviation from H_o. Then under H_a, $\mathbf{E}[\mathbf{Y}|H_a] = \boldsymbol{\mu}_a$ and $\mathbf{Y}$ have multivariate normal distribution with mean $\boldsymbol{\mu}_a$. Then by Theorem 3.1.2 of Moser (1996), under H_a, $\mathbf{Y}'\Sigma^-\mathbf{Y} \sim \chi^2_2(\lambda = \boldsymbol{\mu}'_a\Sigma^-\boldsymbol{\mu}_a)$. Since Σ^- is positive definite and $\boldsymbol{\mu}_a$ is nonzero, the mean of the quadratic form $\mathcal{X}^2_D$ is $\lambda + 2$ with $\lambda > 0$. So for large N, the null and alternative hypotheses are equivalent to $H_o : \lambda = 0$ and $H_a : \lambda > 0$. Then consistency follows. □

References

Agresti A (1992) A survey of exact inference for contingency tables. Stat Sci 7:131–153

Barot S, Gignoux J, Menaut JC (1999) Demography of a savanna palm tree: predictions from comprehensive spatial pattern analyses. Ecology 80:1987–2005

Bickel PJ, Doksum AK (1977) Mathematical statistics, basic ideas and selected topics. Prentice Hall, Englewood Cliffs

Ceyhan E (2008) QR-adjustment for clustering tests based on nearest neighbor contingency tables. Technical report # KU-EC-08-5. Available online as arXiv:0807.4231v1 [stat.ME]

Clark PJ, Evans FC (1954) Distance to nearest neighbor as a measure of spatial relations in population ecology. Ecology 35(4):445–453

Clark PJ, Evans FC (1955) On some aspects of spatial pattern in biological populations. Science 121:397–398

Cochran WG (1952) The χ^2 test of goodness of fit. Ann Math Stat 23:315–345

Conover WJ (1999) Practical nonparametric statistics, 3rd edn. Wiley, New York

Coomes DA, Rees M, Turnbull L (1999) Identifying aggregation and association in fully mapped spatial data. Ecology 80(2):554–565

Cressie NAC (1993) Statistics for spatial data. Wiley, New York

Cuzick J, Edwards R (1990) Spatial clustering for inhomogeneous populations (with discussion). J Roy Stat Soc Ser B 52:73–104

Diggle PJ (2003) Statistical analysis of spatial point patterns. Hodder Arnold Publishers, London

Dixon PM (1994) Testing spatial segregation using a nearest-neighbor contingency table. Ecology 75(7):1940–1948

Dixon PM (2002a) Nearest-neighbor contingency table analysis of spatial segregation for several species. Ecoscience 9(2):142–151

Dixon PM (2002b) Nearest neighbor methods. In: El-Shaarawi AH, Piegorsch WW (eds) Encyclopedia of environmetrics, vol 3. Wiley, NY, pp 1370–1383
Gavrikov V, Stoyan D (1995) The use of marked point processes in ecological and environmental forest studies. Environ Ecol Stat 2(4):331–344
Goreaud F, Pélissier R (2003) Avoiding misinterpretation of biotic interactions with the intertype K_{12}-function: population independence vs. random labelling hypotheses. J Veg Sci 14(5):681–692
Hamill DM, Wright SJ (1986) Testing the dispersion of juveniles relative to adults: a new analytical method. Ecology 67(2):952–957
Krebs CJ (1972) Ecology: the experimental analysis of distribution and abundance. Harper and Row, New York
Kulldorff M (2006) Tests for spatial randomness adjusted for an inhomogeneity: a general framework. J Am Stat Assoc 101(475):1289–1305
Meagher TR, Burdick DS (1980) The use of nearest neighbor frequency analysis in studies of association. Ecology 61(5):1253–1255
Moran PAP (1948) The interpretation of statistical maps. J Roy Stat Soci Ser B 10:243–251
Moser BK (1996) Linear models: a mean model approach. Academic Press
Nanami SH, Kawaguchi H, Yamakura T (1999) Dioecy-induced spatial patterns of two codominant tree species, *Podocarpus nagi* and *Neolitsea aciculata*. J Ecol 87(4):678–687
Penttinen A, Stoyan D, Henttonen H (1992) Marked point processes in forest statistics. Forest Sci 38(4):806–824
Pielou EC (1961) Segregation and symmetry in two-species populations as studied by nearest-neighbor relationships. J Ecol 49(2):255–269
Ripley BD (2004) Spatial statistics. Wiley-Interscience, New York
Schlather M, Ribeiro P Jr, Diggle P (2004) Detecting dependence between marks and locations of marked point processes. J Roy Stat Soc Ser B (Stati Methodol) 66(1):79–93
Solow AR (1989) Bootstrapping sparsely sampled spatial point patterns. Ecology 70(2):379–382
Waller LA, Gotway CA (2004) Applied spatial statistics for public health ata. Wiley-Interscience, NJ
Yamada I, Rogersen PA (2003) An empirical comparison of edge effect correction methods applied to K-function analysis. Geogr Anal 35(2):97–109

Author Biography

Elvan Ceyhan is an Assistant Professor in the Department of Mathematics at Koç University, Istanbul, Turkey. He had a PhD from Applied Mathematics and Statistics Department at Johns Hopkins University, where he became interested in ecological statistics, especially spatial point patterns. In fact, his dissertation introduced a graph theoretical method which has applications in spatial pattern analysis.