Scandinavian Journal of Statistics, Vol. 37: 147–165, 2010
doi: 10.1111/j.1467-9469.2009.00667.x
© 2009 Board of the Foundation of the Scandinavian Journal of Statistics. Published by Blackwell Publishing Ltd.

New Tests of Spatial Segregation Based on Nearest Neighbour Contingency Tables

ELVAN CEYHAN

Department of Mathematics, Koç University

ABSTRACT. The spatial clustering of points from two or more classes (or species) has important implications in many fields and may cause segregation or association, which are two major types of spatial patterns between the classes. These patterns can be studied using a nearest neighbour contingency table (NNCT) which is constructed using the frequencies of nearest neighbour types. Three new multivariate clustering tests are proposed based on NNCTs using the appropriate sampling distribution of the cell counts in a NNCT. The null patterns considered are random labelling (RL) and complete spatial randomness (CSR) of points from two or more classes. The finite sample performance of these tests are compared with other tests in terms of empirical size and power. It is demonstrated that the newly proposed NNCT tests perform relatively well compared with their competitors and the tests are illustrated using two example data sets.

Key words: association, complete spatial randomness, independence, nearest neighbour methods, random labelling, second-order analysis, spatial clustering, spatial point pattern

1. Introduction

The spatial clustering of multiple types of points have received considerable attention in the statistical literature. For convenience and generality, the different types of points are referred to as 'classes', but a *class* can stand for any characteristic of an observation or a point at a particular location. Many methods to analyse the spatial clustering of points from one class or multiple classes have been developed in the literature (Diggle, 2003; Kulldorff, 2006). These include the K- or L-function of Ripley (2004), the D-function of Diggle (2003, p. 131), which is a modified version of Ripley's K-function, the pair correlation function of Stoyan & Stoyan (1994), and the univariate and multivariate J-functions of van Lieshout & Baddeley (1996, 1999). Furthermore, there are nearest neighbour (NN) methods such as comparison of NN distances (Cuzick & Edwards, 1990; Diggle, 2003) and analysis of nearest neighbour contingency tables (NNCTs) which are constructed using the NN frequencies of classes (Pielou, 1961; Meagher & Burdick, 1980). Pielou (1961) proposed various tests and Dixon (1994) introduced an overall test of segregation, cell- and class-specific tests based on NNCTs for the two-class case and extended his tests to the multi-class case (Dixon, 2002a). There are also spatial pattern tests that adjust for an inhomogeneity and are mostly used for testing the clustering of cases in epidemiology such as the k-NN tests of Cuzick & Edwards (1990), the spatial scan statistic of Kulldorff (1997), Whittemore's test, Tango's MEET, Besag–Newel's R and Moran's I (Song & Kulldorff, 2003). An extensive survey of such tests is provided by Kulldorff (2006).

The null hypothesis for the NNCT tests is H_o: *randomness in the NN structure* which usually results from one of the two (random) pattern types: *random labelling* (RL) of a set of fixed points with two (or more) classes or *complete spatial randomness* (CSR) of points from two (or more) classes which is called *CSR independence*, henceforth. That is, if the points from each class are assumed to be uniformly distributed over a region of interest, then H_o is implied by CSR independence which is also referred to as (a type of) 'population independence' in the literature (Goreaud & Pélissier, 2003). If only the labelling of a set of fixed points is random, then H_o is implied by RL. The distinction between CSR independence and RL is very important and the null model depends on the particular context. Goreaud & Pélissier (2003) state that under CSR (independence) the (locations of the points from) two classes are *a priori* the result of different processes (e.g. individuals of different species or age cohorts), whereas; under RL, some processes affect *a posteriori* the individuals of a single population (e.g. diseased versus non-diseased individuals of a single plant species). Notice also that although CSR independence and RL are not the same, they lead to the same null model for tests using NNCT, which does not require spatially explicit information.

Two major types of (bivariate) spatial clustering patterns, namely, *association* and *segregation* are considered as alternative patterns. *Association* occurs if the NN of an individual is more likely to be from another class. For example, in plant biology, the two classes might be two mutualistic plant species, so that the species depend on each other to survive. As another example, one class of points might be a parasitic plant species exploiting another plant species which constitutes the other class. *Segregation* occurs if the NN of an individual is more likely to be of the same class as the individual; i.e. the members of the same class tend to be clumped or clustered (see, e.g. Pielou, 1961). For instance, one type of plant might not grow well around another type of plant and vice versa. In plant biology, points from one class might represent the coordinates of trees from a species with large canopy, so that other plants (whose coordinates are the points from the other class) that need light cannot grow around these trees. See, for instance, Dixon (1994) and Coomes *et al.* (1999) for more details. The segregation and association patterns are not symmetric in the sense that, when two classes are segregated, they do not necessarily exhibit the same degree of segregation or when two classes are associated, one class could be more associated with the other. Although it is not possible to list all of the many different types of segregation, its existence can be tested by an analysis of the NN relationships between the classes (Pielou, 1961).

Pielou (1961) used the usual Pearson's chi-squared test of independence for detecting the segregation of the two classes. Because of the ease in computation and interpretation, Pielou's test of segregation has been frequently used for both completely mapped and sparsely sampled data (Meagher & Burdick, 1980). However, it is shown that Pielou's test is not appropriate to test RL (Meagher & Burdick, 1980; Dixon, 1994). For the two-class case, Ceyhan (2008b) demonstrated that Pielou's test is liberal under CSR independence and RL and is only appropriate for a random sample of (base, NN) pairs. Here, if Y is a NN of point X, then X is the *base point* and Y is the NN *point*.

In this article, new overall segregation tests based on NNCTs are introduced and their performance is compared with various other tests in literature with extensive Monte Carlo simulations. It is shown that one of the new tests has better power and size performance compared to other NNCT tests. Most of the methods of spatial pattern analysis in the literature use Monte Carlo simulation or randomization tests (Kulldorff, 2006), but asymptotic distributions are available for the NNCT tests. Among the clustering tests above, univariate tests are not comparable with NNCT tests; Moran's I and Whittemore's tests are shown to perform poorly in detecting some kind of clustering (Song & Kulldorff, 2003) and most of the tests require Monte Carlo simulation or randomization to attach significance to their results. Hence in this article NNCT tests are compared only with Cuzick–Edward's k-NN tests and their combined versions in an extensive simulation study. NNCT tests are also compared with Ripley's L-function and Diggle's D-function for the appropriate null hypotheses in the examples. These second-order methods are also based on Monte Carlo simulation, but they are perhaps the most widely used tests for spatial interaction at various scales.

For simplicity, the spatial point patterns for two classes are described; the extension to the multi-class case is straightforward. Furthermore, only the *completely mapped data* type

© 2009 Board of the Foundation of the Scandinavian Journal of Statistics.

is considered; i.e. the locations of all events in a defined area are observed. Throughout the article random quantities are denoted by capital letters, whereas fixed quantities are denoted by lower-case letters. The NNCTs and related tests in the literature are described in section 2, and the new versions of segregation tests in section 3. Empirical significance levels and an empirical power analysis of the tests are given respectively, in, sections 4 and 5. Two illustrative examples are given in section 6. Finally, discussion and some conclusions with guidelines in using the tests, are given in section 7.

2. NNCTs and related tests in the literature

In the case of two classes of points, NNCT is a 2×2 table of frequencies $N_{ij}, i, j \in \{1, 2\}$, of cases, where a point of type j is the NN of a point of type i (Table 1). Under segregation, the diagonal entries, N_{ii} for $i = 1, 2$, tend to be larger than expected; under association, the off-diagonals tend to be larger than expected under CSR independence or RL.

Pielou (1961) used the usual Pearson's chi-squared test of independence for detecting the segregation of the two classes:

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

When NNCT is based on a random sample of (base, NN) pairs (with fixed row and column sums, n_i and c_j, respectively), $\mathbf{E_P}[N_{ij}] = n_i c_j / n$ where $n = n_1 + n_2$ is the (fixed) total number of observations and $\mathcal{X}_P$ is approximately distributed as χ^2_1 (i.e. χ^2 distribution with 1 degree of freedom) for large n_i (Ceyhan, 2008b). When the NNCT is constructed under CSR independence or RL, Pielou's test is not appropriate (Meagher & Burdick, 1980; Dixon, 1994) and is shown to be liberal (Ceyhan, 2008b).

Dixon (1994) proposed a series of tests for segregation based on NNCTs. For Dixon's tests, the probability of an individual from class j serving as a NN of an individual from class i depends only on the class sizes (i.e. row sums), but not the total number of times class j serves as NNs (i.e. column sums). The level of segregation is estimated by comparing the observed cell counts with the expected cell counts under RL of 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 (frequencies) in a NNCT (Dixon, 1994, 2002a).

Under RL,

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

The cell-specific test statistics suggested by Dixon are given by

$$Z^D_{ij} = \frac{N_{ij} - \mathbf{E}[N_{ij}]}{\sqrt{\mathrm{var}[N_{ij}]}}, \tag{2}$$

Table 1. *Nearest neighbour contingency tables for two classes*

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

© 2009 Board of the Foundation of the Scandinavian Journal of Statistics.

where

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

with p_{xx}, p_{xxx} and p_{xxxx} being the probabilities that a randomly picked pair, triplet, or quartet of points, respectively, are in the indicated classes, and are given by Ceyhan (2008a). Furthermore, Q is the number of points with shared NNs, which occur when two or more points share a NN and R is twice the number of reflexive pairs. Then $Q=2(Q_2+3Q_3+6Q_4+10Q_5+15Q_6)$, where Q_k is the number of points that serve as a NN to other points k times. One- and two-sided tests are possible for each cell (i, j) using the asymptotic normal approximation of Z^D_{ij} given in (2) which is the same as Dixon's Z_{AA} when $i=j=1$ and Z_{BB} when $i=j=2$ (Dixon, 1994). Note also that in (2) there seems to be four different test statistics as there are four cells; but there are essentially two different tests, as $Z^D_{11}=-Z^D_{12}$ and $Z^D_{22}=-Z^D_{21}$.

Dixon's overall test of segregation tests the hypothesis that expected cell counts in the NNCT are as in (1). In the two-class case, he calculates Z^D_{ii} for both $i\in\{1,2\}$ and combines these test statistics into a statistic that is asymptotically distributed as χ^2_2 under RL (Dixon, 1994). Under RL with the two classes, the suggested test statistic is given by

$$\mathcal{X}_D=\mathbf{Y}'\mathbf{\Sigma}^{-1}\mathbf{Y}=\begin{bmatrix}N_{11}-\mathbf{E}[N_{11}]\\ N_{22}-\mathbf{E}[N_{22}]\end{bmatrix}'\begin{bmatrix}\mathrm{var}[N_{11}] & \mathrm{cov}[N_{11},N_{22}]\\ \mathrm{cov}[N_{11},N_{22}] & \mathrm{var}[N_{22}]\end{bmatrix}^{-1}\begin{bmatrix}N_{11}-\mathbf{E}[N_{11}]\\ N_{22}-\mathbf{E}[N_{22}]\end{bmatrix}, \quad (4)$$

where $\mathbf{E}[N_{ii}]$ are as in (1), $\mathrm{var}[N_{ii}]$ are as in (3) and

$$\mathrm{cov}[N_{11},N_{22}]=(n^2-3n-Q+R)p_{1122}-n^2p_{11}p_{22}. \quad (5)$$

Dixon's statistic given in (4) can also be written as

$$\mathcal{X}_D=\frac{(Z^D_{11})^2+(Z^D_{22})^2-2rZ^D_{11}Z^D_{22}}{1-r^2},$$

where $r=\mathrm{cov}[N_{11},N_{22}]/\sqrt{\mathrm{var}[N_{11}]\mathrm{var}[N_{22}]}$ (Dixon, 1994). For more than two classes, $\mathcal{X}_D=\mathbf{N}'\mathbf{\Sigma}^-\mathbf{N}$, where $\mathbf{N}$ is the vector of N_{ij} values concatenated row-wise and $\mathbf{\Sigma}^-$ is the generalized inverse of variance–covariance matrix of $\mathbf{N}$ (Searle, 2006).

3. New overall segregation tests based on NNCTs

In defining the new segregation tests, a track similar to that of Dixon (1994) is followed. For each cell, a new type of cell-specific test statistic is defined, and then these four tests are combined into an overall test.

3.1. Version I of the new segregation tests

Consider the test statistics, $T^I_{ij}=N_{ij}-n_iC_j/n$, for cells in the NNCT. Under RL and conditional on $C_j=c_j$, we define

$$N^I_{ij}:=\frac{T^I_{ij}}{\sqrt{n_ic_j/n}}=\frac{N_{ij}-n_ic_j/n}{\sqrt{n_ic_j/n}},$$

then $\mathcal{X}_P=\sum_{i=1}^2\sum_{j=1}^2\left(N^I_{ij}\right)^2$. Moreover, we have

$$\mathbf{E}[T^I_{ij}]=\begin{cases}\dfrac{n_i(n_i-n)}{n(n-1)} & \text{if } i=j,\\[2ex] \dfrac{n_in_j}{n(n-1)} & \text{if } i\neq j.\end{cases}$$

© 2009 Board of the Foundation of the Scandinavian Journal of Statistics.

Hence under RL, we have $\mathbf{E}[T^I_{ij}]\neq 0$ which implies that $\mathbf{E}[N^I_{ij}]\neq 0$. Furthermore,

$$\lim_{n_i,n_j\to\infty} \mathbf{E}[T^I_{ij}] = \begin{cases} \nu_i(1-\nu_i) & \text{if } i=j, \\ \nu_i\nu_j & \text{if } i\neq j, \end{cases}$$

where ν_i is the probability that an individual being from class i and $n_i, n_j \to \infty$ means that $\min(n_i, n_j)\to\infty$. Although $\mathbf{E}[N^I_{ij}]$ is not analytically tractable, we can find its limit. Because $\sqrt{n_i c_j/n^2} \to \sqrt{\nu_i^2} = \nu_i$ which implies $1/\sqrt{n_i c_j/n} \to 0$ as $n_i, n_j \to \infty$, we have $\lim_{n_i,n_j\to\infty} \mathbf{E}[N^I_{ij}] = 0$.

Let $\mathbf{N_I}$ be the vector of N^I_{ij} values concatenated row-wise and let $\boldsymbol{\Sigma}_I$ be the variance–covariance matrix of $\mathbf{N_I}$ based on the correct sampling distribution of the cell counts. That is, $\boldsymbol{\Sigma}_I = (\text{cov}[N^I_{ij}, N^I_{kl}])$ where, $\text{cov}[N^I_{ij}, N^I_{kl}] = (n/\sqrt{n_i c_j n_k c_l})\,\text{cov}[N_{ij}, N_{kl}]$ with, $\text{cov}[N_{ij}, N_{kl}]$ being as in (3) if $(i,j)=(k,l)$ and as in (5) if $(i,j)=(1,1)$ and $(k,l)=(2,2)$. Since $\boldsymbol{\Sigma}_I$ is not invertible, its generalized inverse, $\boldsymbol{\Sigma}_I^-$ is used. Then the proposed statistic for testing overall segregation is the quadratic form

$$\mathcal{X}_I = \mathbf{N}_\mathbf{I}' \boldsymbol{\Sigma}_I^- \mathbf{N}_\mathbf{I}, \tag{6}$$

which asymptotically has a χ^2_1 distribution. The test statistic $\mathcal{X}_I$ can be obtained by adding a correction term to $\mathcal{X}_P$ as $\mathcal{X}_I = \mathbf{N}_\mathbf{I}'\boldsymbol{\Sigma}_I^-\mathbf{N}_\mathbf{I} = \mathbf{N}_\mathbf{I}'(\boldsymbol{\Sigma}_I^- - I_2 + I_2)\mathbf{N}_\mathbf{I} = \mathbf{N}_\mathbf{I}'\mathbf{N}_\mathbf{I} + \mathbf{N}_\mathbf{I}'(\boldsymbol{\Sigma}_I^- - I_2)\mathbf{N}_\mathbf{I} = \mathcal{X}_P + \Delta_c$, where $\Delta_c = \mathbf{N}_\mathbf{I}'(\boldsymbol{\Sigma}_I^- - I_2)\mathbf{N}_\mathbf{I}$ and I_2 is the 2×2 identity matrix. Furthermore, $\boldsymbol{\Sigma}_I$ can be obtained from $\boldsymbol{\Sigma}$ by multiplying it entry-wise with the matrix $\mathbf{C}^I_M = (n/\sqrt{n_i c_j n_k c_l})$.

3.2. Version II of the new segregation tests

For large n, we have $n_i c_j/n \approx n\nu_i\kappa_j$, where κ_j is the probability that a NN is of class j. Under CSR independence or RL, $\nu_i = \kappa_i$ for $i=1,2$, and then $n_i c_j/n \approx n\nu_i\nu_j$ for large n. This suggests the test statistics, $T^{II}_{ij} = N_{ij} - n_i n_j/n$, for the four cells. Let

$$N^{II}_{ij} = \frac{T^{II}_{ij}}{\sqrt{n_i n_j/n}} = \frac{N_{ij} - n_i n_j/n}{\sqrt{n_i n_j/n}}.$$

Under RL,

$$\mathbf{E}[T^{II}_{ij}] = \begin{cases} \dfrac{n_i(n_i-n)}{n(n-1)} & \text{if } i=j, \\ \dfrac{n_i n_j}{n(n-1)} & \text{if } i\neq j, \end{cases} \longrightarrow \begin{cases} \nu_i(1-\nu_i) & \text{if } i=j, \\ \nu_i\nu_j & \text{if } i\neq j, \end{cases} \quad \text{as } n_i, n_j \to \infty.$$

Notice that under RL, $\mathbf{E}[T^{II}_{ij}] = \mathbf{E}[T^I_{ij}] \neq 0$ which implies that $\mathbf{E}[N^{II}_{ij}]\neq 0$. Furthermore,

$$\mathbf{E}[N^{II}_{ij}] = \begin{cases} \dfrac{(n_i-n)}{\sqrt{n}(n-1)} & \text{if } i=j, \\ \dfrac{\sqrt{n_i n_j}}{\sqrt{n}(n-1)} & \text{if } i\neq j. \end{cases}$$

Thus, $\lim_{n_i,n_j\to\infty} \mathbf{E}[N^{II}_{ij}] = 0$. Note that the sum of the squares of N^{II}_{ij} does not equal $\mathcal{X}_P$. Let $\mathbf{N}_{\text{II}}$ be the vector of N^{II}_{ij} values and

$$\boldsymbol{\Sigma}_{II} = (\text{cov}[N^{II}_{ij}, N^{II}_{kl}]) = \left(\frac{n}{\sqrt{n_i n_j n_k n_l}}\,\text{cov}\left[N_{ij}, N_{kl}\right] \right)$$

be the variance–covariance matrix of $\mathbf{N_{II}}$. Similarly as in section 3.1, the proposed statistic for testing overall segregation is the quadratic form

$$\mathcal{X}_{II} = \mathbf{N}_\mathbf{II}' \boldsymbol{\Sigma}_{II}^- \mathbf{N}_\mathbf{II}, \tag{7}$$

© 2009 Board of the Foundation of the Scandinavian Journal of Statistics.

which asymptotically has a χ^2_2 distribution under RL. This version of the segregation test is asymptotically equivalent to Dixon's test in (4).

3.3. Version III of the new segregation tests

Consider the following test statistics which use both the row and column sums and are not conditional on the column sums:

$$T^{III}_{ij} = \begin{cases} N_{ij} - \dfrac{(n_i-1)}{(n-1)} C_j & \text{if } i=j, \\ N_{ij} - \dfrac{n_i}{(n-1)} C_j & \text{if } i \neq j. \end{cases}$$

Note that $\mathbf{E}[T^{III}_{ij}]=0$ under RL, but the sum of the squares of T^{III}_{ij} does not equal $\mathcal{X}_P$. Let $\mathbf{N_{III}}$ be the vector of T^{III}_{ij} values concatenated row-wise and $\boldsymbol{\Sigma}_{III}$ be the corresponding variance–covariance matrix. Similarly as in section 3.1, the proposed test statistic for overall segregation is the quadratic form

$$\mathcal{X}_{III} = \mathbf{N}'_{\mathbf{III}} \boldsymbol{\Sigma}^{-}_{III} \mathbf{N_{III}}, \tag{8}$$

which asymptotically has a χ^2_1 distribution. $\boldsymbol{\Sigma}_{III} = (\text{cov}[T^{III}_{ij}, T^{III}_{kl}])$ where $\text{cov}[T^{III}_{ij}, T^{III}_{kl}]$ are provided in Ceyhan (2008a).

Remark 3.1. The status of Q and R under CSR independence and RL: Q and R are fixed under RL, but random under CSR independence. Hence under the CSR independence pattern, the asymptotic distributions of $\mathcal{X}_I, \mathcal{X}_{II}$ and $\mathcal{X}_{III}$ are conditional on Q and R. Moreover, $\mathcal{X}_I$ is also conditional on the column sums (i.e. on $C_j = c_j$). Cuzick–Edward's tests are also appropriate for the RL of the points in a study area but are conditional on the locations of the points under CSR independence.

The unconditional variances and covariances can be obtained by replacing Q and R with their expectations (Ceyhan, 2009a). Unfortunately, given the difficulty of calculating the expectation of Q under CSR independence, it is reasonable and convenient to use test statistics employing the conditional variances and covariances even when assessing their behaviour under the CSR independence model. Cox (1981) calculated analytically that $\mathbf{E}[R \mid N=n] = 0.6215n$ for a planar Poisson process where N is the (random) total sample size. Alternatively, one can estimate the expected values of Q and R empirically. For example, for homogeneous planar Poisson pattern, we have $\mathbf{E}[Q \mid N=n] \approx 0.6328n$ and $\mathbf{E}[R \mid N=n] \approx 0.6211n$ (estimated empirically by 1,000,000 Monte Carlo simulations for various values of n on the unit square). Notice that $\mathbf{E}[R \mid N=n]$ agrees with the analytical result of Cox (1981). When Q and R are replaced by $0.63n$ and $0.62n$, respectively, the so-called *QR-adjusted* tests are obtained. However, QR adjustments do not improve on the unadjusted NNCT tests (Ceyhan, 2008c).

Remark 3.2. Extension of NNCT tests to the multi-class case: For q classes with $q>2$, the NNCT will be of dimension $q \times q$. Pielou's test readily extends to $q \times q$ contingency tables, but it will still be inappropriate (i.e. liberal). The cell counts for the diagonal cells have asymptotic normality. As for the off-diagonal cells, although the asymptotic normality is supported by Monte Carlo simulation results (Dixon, 2002a), it is not rigorously proven yet. Nevertheless, if the asymptotic normality held for all q^2 cell counts in the NNCT, under RL, Dixon's test and version II of the new tests would have $\chi^2_{q(q-1)}$ distribution, versions I and III of the new tests would have $\chi^2_{(q-1)^2}$ distribution asymptotically. Under CSR independence, these tests will have the corresponding asymptotic distributions conditional on Q and R.

© 2009 Board of the Foundation of the Scandinavian Journal of Statistics.

3.4. Cuzick–Edward's k-NN tests

Cuzick–Edward's k-NN test is defined as $T_k = \sum_{i=1}^{n} \delta_i d_i^k$, where

$$\delta_i = \begin{cases} 1 & \text{if } z_i \text{ is a case,} \\ 0 & \text{if } z_i \text{ is a control,} \end{cases}$$

with z_i being the ith point and d_i^k is the number of cases among the k NNs of the ith point. As in practice, the correct choice of k is not known in advance, Cuzick & Edwards (1990) also suggest combining information for various T_k values. Assuming multivariate normality of T_k values and T_k being a mixture of shifts all in the same direction under an alternative, the combined test statistic is given by

$$T_S^{\text{comb}} = \mathbf{1}'\boldsymbol{\Sigma}^{-1/2}\mathbf{T},$$

where $S = \{k_1, k_2, \ldots, k_m\}, \mathbf{T} = (T_{k_1}, T_{k_2}, \ldots, T_{k_m})'$ (i.e. T_S^{comb} is the test obtained by combining T_k whose indices are in S), $\mathbf{1}' = (1, 1, \ldots, 1)$ and $\boldsymbol{\Sigma} = \text{cov}[\mathbf{T}]$ is the variance–covariance matrix of $\mathbf{T}$. Under H_o: *RL of cases and controls to the given locations in the study region*, T_k converges in law to $N(\mathbf{E}[T_k], \text{var}[T_k]/n_0)$; similarly, T_S^{comb} converges in law to $N(\mathbf{E}[T_S^{\text{comb}}], \text{var}[T_S^{\text{comb}}])$ when the number of cases n_0 goes to infinity. The expected values $\mathbf{E}[T_k]$ and $\mathbf{E}[T_S^{\text{comb}}]$ and variances $\text{var}[T_k]$ and $\text{var}[T_S^{\text{comb}}]$ are provided in Cuzick & Edwards (1990).

3.5. Second-order analysis by Ripley's L-function and related functions

When the NNCT tests and k-NN tests indicate significant segregation or clustering, one might also be interested in the (possible) causes of the segregation and the type and level of interaction between the classes at different scales (i.e. inter-point distances). To answer such questions, we consider Ripley's (univariate) L-function, denoted by $L_{ii}(t)$, and bivariate L-function, denoted by $L_{ij}(t)$, for classes i and j. Ripley's L is estimated as

$$\hat{L}_{ii}(t) = \sqrt{\left(\hat{K}_{ii}(t)/\pi\right)},$$

where t is the distance from a randomly chosen event (i.e. location of a tree) and $\hat{K}_{ii}(t)$ is an estimator of

$$K_{ii}(t) = \lambda^{-1}\mathbf{E}\ [\#\text{ of extra events within distance } t \text{ of a randomly chosen event}]$$

with λ being the density (number per unit area) of events and is calculated as

$$\hat{K}_{ii}(t) = \hat{\lambda}^{-1} \sum_i \sum_{j \neq i} w(i, d_{ij}) \mathbf{I}(d_{ij} < t)/N, \tag{9}$$

where $\hat{\lambda} = N/A$ is an estimate of density (N is the observed number of points and A is the area of the study region), d_{ij} is the distance between points i and j, $\mathbf{I}(\cdot)$ is the indicator function, and $w(i, d_{ij})$ is the proportion of the circumference of the circle centred at point i with radius d_{ij} that falls in the study area, which corrects for the boundary effects, see Diggle (2003) for more details. Also provided is the Ripley's bivariate L-function, denoted by $L_{ij}(t)$ for classes i and j, and estimated as $\hat{L}_{ij}(t) = \sqrt{(\hat{K}_{ij}(t)/\pi)}$, where $\hat{K}_{ij}(t)$ is an estimator of

$$K_{ij}(t) = \lambda_j^{-1}\mathbf{E}\ [\#\text{ of extra type } j \text{ events within distance } t \text{ of a randomly chosen type } i \text{ event}]$$

with λ_j being the density of type j events. $\hat{K}_{ij}(t)$ is calculated as

$$\hat{K}_{ij}(t) = \left(\hat{\lambda}_i \hat{\lambda}_j A\right)^{-1} \sum_i \sum_j w(i_k, d_{i_k, j_l}) \mathbf{I}(d_{i_k, j_l} < t),$$

© 2009 Board of the Foundation of the Scandinavian Journal of Statistics.

where d_{i_k,j_l} is the distance between the kth type i and the lth type j points, and $w(i_k, d_{i_k,j_l})$ is the proportion of the circumference of the circle centred at kth type i point with radius d_{i_k,j_l} that falls in the study area, which is used for edge correction, see Diggle (2003) and Ceyhan (2008a) for more details.

However, Ripley's K-function is cumulative, so interpreting the spatial interaction at larger distances is problematic (Stoyan & Penttinen, 2000; Wiegand *et al.*, 2007). The pair correlation function $g(t)$ is better for this purpose (Stoyan & Stoyan, 1994). The pair correlation function of a (univariate) stationary point process is defined as

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

where $K'(t)$ is the derivative of $K(t)$. However, the pair correlation function estimates might have critical behaviour for small t if $g(t) > 0$, as the bias is considerably large. This problem gets worse especially in cluster processes (Stoyan & Stoyan, 1996). On the other hand, the pair correlation function is more reliable for larger distances and it might be safer to use $g(t)$ for distances larger than the average NN distance in the data set (Ceyhan, 2009b). Therefore, we only compare Ripley's L-function with NNCT-tests here.

When the null case is the RL of points from an inhomogeneous Poisson process, Ripley's K- or L-function in the general form is not appropriate to test for the spatial clustering of the cases (Kulldorff, 2006). However, Diggle (2003) suggests a version based on Ripley's univariate K-function as $D(t) = K_{11}(t) - K_{22}(t)$. In this set-up, 'no spatial clustering' is equivalent to RL of case and control labels on the locations in the sample, which implies $D(t) = 0$, as $K_{22}(t)$ measures the degree of spatial aggregation of the controls (i.e. the population at risk), while $K_{11}(t)$ measures this same spatial aggregation plus any additional clustering due to the disease. The test statistic $D(t)$ is estimated by $\hat{D}(t) = \hat{K}_{11}(t) - \hat{K}_{22}(t)$, where $\hat{K}_{ii}(t)$ is as in (9).

4. Empirical significance levels of the tests under CSR independence

The points from two classes are generated under H_o: *CSR independence* as follows. Consider the classes 1 and 2 (i.e. X and Y) of sizes n_1 and n_2, respectively. At each of $N_{mc} = 10{,}000$ replicates, we generate data for some pairs of (n_1, n_2) with $n_1, n_2 \in \{10, 30, 50, 100\}$ points independently and identically distributed (i.i.d.) from $\mathcal{U}((0,1)\times(0,1))$, the uniform distribution on the unit square. and compute the NNCT tests and Cuzick–Edward's k-NN and combined tests. The empirical sizes are calculated as the ratio of the number of significant results to the number of Monte Carlo replications.

As Pielou's test is extremely liberal in rejecting H_o, the empirical sizes for other NNCT tests considered here, as well as for the Cuzick–Edward's k-NN tests T_k for $k \leq 5$ and four combined tests T_S^{comb}, are presented in Fig. 1 for $(n_1, n_2) \in \{(30,30), (30,50), (50,30), (50,50)\}$. The asymptotic normal approximation to proportions is used in determining the significance of the deviations of the empirical sizes from 0.05. For these proportion tests, we also use $\alpha = 0.05$ as the significance level. The NNCT tests are about the desired level (or size) when both samples are large, and are mostly conservative otherwise.

In the simulated patterns under CSR independence, class X represents the cases and class Y represents the controls in the context of Cuzick–Edward's tests. However, the simulated patterns are not realistic for the case/control framework, as X and Y points are from a homogeneous Poisson process, whereas in practice case and control locations usually exhibit inhomogeneity. So Cuzick–Edward's tests are not used in the conventional sense here, but instead are used to test deviations of the classes from CSR independence. The actual values of empirical significance levels are provided by Ceyhan (2008a).

© 2009 Board of the Foundation of the Scandinavian Journal of Statistics.

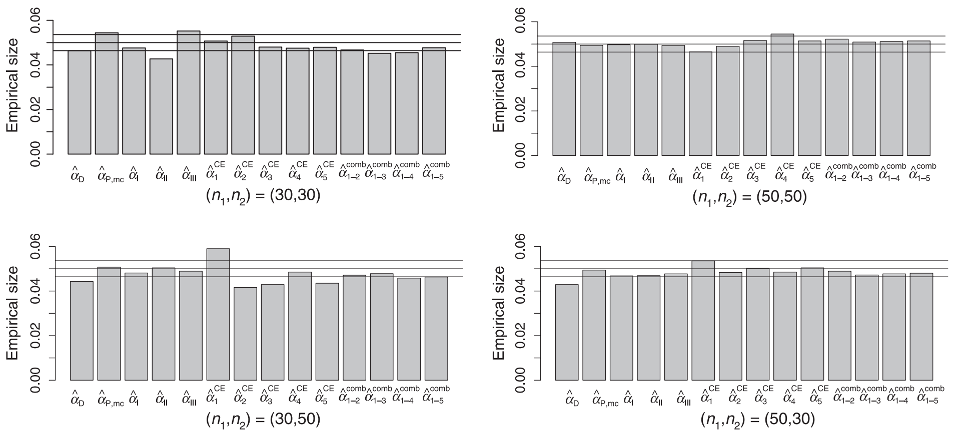

Fig. 1. The empirical size estimates for the nearest neighbour contingency tables (NNCT) tests based on 10,000 Monte Carlo replicates under the complete spatial randomness independence of the two classes, i.e. uniform data from two classes on the unit square for various (n_1, n_2) values. The horizontal lines are located at 0.0464 (upper threshold for conservativeness), 0.0500 (nominal level) and 0.0536 (lower threshold for liberalness). $\hat{\alpha}_D$ stands for the empirical significance level for Dixon's test, $\hat{\alpha}_I, \hat{\alpha}_{II}$ and $\hat{\alpha}_{III}$ for versions I, II and III of the new tests, respectively, and $\hat{\alpha}_{P,mc}$ for the Monte Carlo-corrected version of Pielou's test as $\mathcal{X}_{P,mc} := \mathcal{X}_P + 0.013/1.643$ (Ceyhan, 2008a), $\hat{\alpha}_k^{CE}$ stands for the empirical size for Cuzick–Edward's k-NN test for $k = 1, 2, \ldots, 5$, and $\hat{\alpha}_S^{\text{comb}}$ for the combined test T_S^{comb} for $S \in \{\{1, 2\}, \{1, 2, 3\}, \{1, 2, 3, 4\}, \{1, 2, 3, 4, 5\}\}$. For brevity in notation, the index set S is written as $1 - j$ for $j = 2, 3, 4, 5$.

Observe that as k increases, the size performance of Cuzick–Edward's k-NN test T_k seems to get better, since the size gets closer to the desired nominal level of 0.05. However for most sample sizes T_4 seems to have the best empirical size performance. On the other hand, among the combined tests, T_S^{comb} tests usually have better size performance compared with each T_k for $k \in S$. Furthermore, as the number of combined tests (i.e. the size of the index set S) increases, the empirical size gets closer to the nominal level, hence the T_{1-5}^{comb} exhibits the best size performance.

Remark 4.1. Empirical significance levels under RL: To evaluate empirical size performance of the tests under RL, Monte Carlo simulations under various RL cases are also performed (Ceyhan, 2008a). The RL cases are more appropriate for the case/control framework of Cuzick–Edward's tests compared to the CSR independence cases. In the RL cases, the locations of the points could represent the locations of the $n = (n_1 + n_2)$ subjects so that n_1 of them are patients (i.e. cases) whereas the remaining n_2 are controls. It is shown that the tests are usually about the desired level. However, Dixon's and corrected version of Pielou's test are more severely influenced by the clustering of the points. Cuzick–Edward's 1-NN test seems to be the most sensitive to the differences in the relative abundance of the classes.

5. Empirical power analysis

5.1. Empirical power analysis under segregation alternatives

For the segregation alternatives, we generate $X_i \overset{\text{i.i.d.}}{\sim} \mathcal{U}((0, 1 - s) \times (0, 1 - s))$ for $i = 1, 2, \ldots, n_1$ and $Y_j \overset{\text{i.i.d.}}{\sim} \mathcal{U}((s, 1) \times (s, 1))$ for $j = 1, 2, \ldots, n_2$ with $s \in (0, 1)$. In the pattern generated, appro-

© 2009 Board of the Foundation of the Scandinavian Journal of Statistics.

priate choices of s will imply X_i and Y_j to be segregated. The three values of s constitute the three segregation alternatives:

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

From H_S^I to H_S^{III} (i.e. as s increases), the segregation between X and Y gets stronger. By construction, with respect to the unit square these alternative patterns are examples of departures from first-order homogeneity. The simulated segregation patterns are symmetric in the sense that, X and Y points are equally segregated (or clustered) from each other. Hence, although class X stands for the 'cases' in Cuzick–Edward's tests, the results would be similar if class Y is chosen instead.

The power estimates under the segregation alternatives H_S^I and H_S^{II} for $(n_1, n_2) \in \{(30, 30), (30, 50)\}$ are presented in Fig. 2 where the empirical power estimate for Pielou's test is not presented as it is misleading, see Ceyhan (2008b). Furthermore, the power estimates under H_S^{III} are not presented either, as they are almost 1 for each test. The power estimates for other sample size combinations and under H_S^{III} are presented in Ceyhan (2008a). Among the NNCT tests, version II of the new tests has the highest power estimates. Among Cuzick–Edward's k-NN tests, T_3, T_4, and T_5 have the highest power estimates which are virtually indistinguishable. The power estimates for T_k tests seem to be higher than those of NNCT-tests presented here. As for the combined tests T_S^{comb}, the more successive T_k tests are combined from $1, 2, \ldots, k$, the higher the power estimates for T_S^{comb}.

5.2. *Empirical power analysis under association alternatives*

For the association alternatives, we generate $X_i \overset{\text{i.i.d.}}{\sim} \mathcal{U}((0,1) \times (0,1))$ for $i = 1, 2, \ldots, n_1$. Then Y_j for $j = 1, 2, \ldots, n_2$ are generated as follows. For each j, an i is picked at random, then Y_j are generated as $X_i + R_j(\cos T_j, \sin T_j)'$ where $R_j \overset{\text{i.i.d.}}{\sim} \mathcal{U}(0, r)$ with $r \in (0, 1)$ and $T_j \overset{\text{i.i.d.}}{\sim} \mathcal{U}(0, 2\pi)$. In the pattern generated, appropriate choices of r will imply Y_j and X_i are associated. The

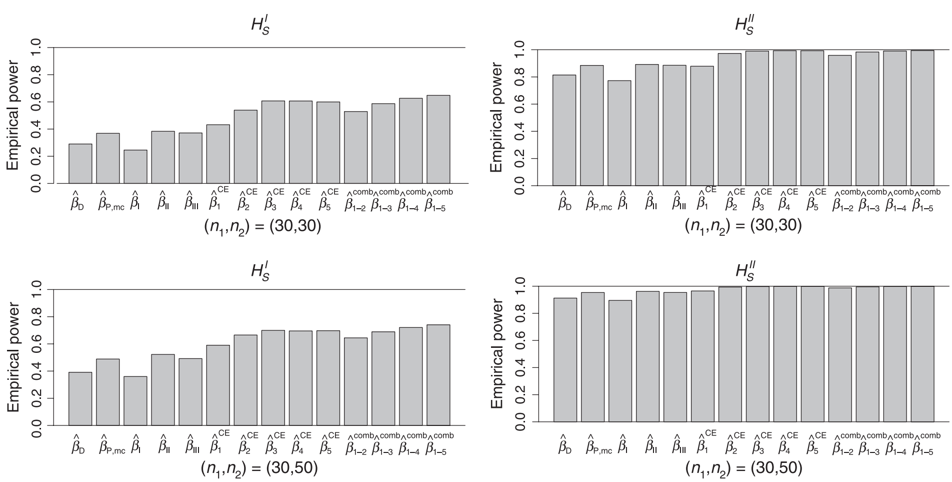

Fig. 2. The empirical power estimates of the tests under the segregation alternatives H_S^I (left) and H_S^{II} (right) based on 10,000 Monte Carlo replicates for $(n_1, n_2) \in \{(30, 30), (30, 50)\}$. $\hat{\beta}_D$ stands for the empirical estimate for Dixon's test, $\hat{\beta}_I$, $\hat{\beta}_{II}$ and $\hat{\beta}_{III}$ for versions I, II and III of the new tests, respectively, and $\hat{\beta}_{P,mc}$ for the Monte Carlo-corrected version of Pielou's test, $\hat{\beta}_k^{CE}$ for Cuzick–Edwards k-NN test for $k = 1, 2, \ldots, 5$ and $\hat{\beta}_{1-j}^{\text{comb}}$ for Cuzick–Edwards T_{1-j}^{comb} for $j = 1, 2, 3, 4$.

© 2009 Board of the Foundation of the Scandinavian Journal of Statistics.

three values of r constitute the three association alternatives:

$$H_A^I : r = 1/4, \quad H_A^{II} : r = 1/7, \quad H_A^{III} : r = 1/10.$$

From H_A^I to H_A^{III} (i.e. as r decreases), the association between X and Y gets stronger. By construction, conditional on sample sizes, X points are from a homogeneous Poisson process with respect to the unit square, while Y points exhibit inhomogeneity in the same region. Furthermore, these alternative patterns are examples of departures from second-order homogeneity which implies association of the class Y with class X. The simulated association patterns are contrary to the case/control framework of Cuzick–Edward's tests, since class X is used for the case class and class Y points are clustered around X points. However, Cuzick–Edward's tests are still included in our analysis to evaluate their performance under this type of deviation from the CSR independence pattern.

The power estimates under the association alternatives H_A^{II} and H_A^{III} for $(n_1, n_2) \in \{(30, 30), (30, 50)\}$ are plotted in Fig. 3. The power estimates under H_A^I and for other sample size combinations are presented in Ceyhan (2008a). Considering the NNCT tests, for most sample size combinations, version III of the new tests has the highest power estimate. Hence version III of the new tests is recommended for large samples, and Monte Carlo randomization is recommended for the NNCT-tests for small samples.

Considering Cuzick–Edwards k-NN tests, the power estimates tend to decrease as k increases. Among Cuzick–Edward's combined tests, usually T_{1-5}^{comb} has the best power performance. Considering all tests together, still T_{1-5}^{comb} has the best power performance under the association alternatives, hence it can be recommended for use against this type of association. However, given the computational cost of combined tests, one might prefer version III of the new tests under the association alternatives for larger samples, as its power is very close to that of T_{1-5}^{comb}. For smaller samples, either Monte Carlo randomization for NNCT tests, or asymptotic approximation for T_{1-4}^{comb} or T_{1-5}^{comb} can be used.

Remark 5.1. Edge correction for NNCT-tests: Edge effects are not a concern for testing against the RL pattern. However, the CSR independence pattern assumes that the study region is unbounded for the analysed pattern, which is not the case in practice. So it might

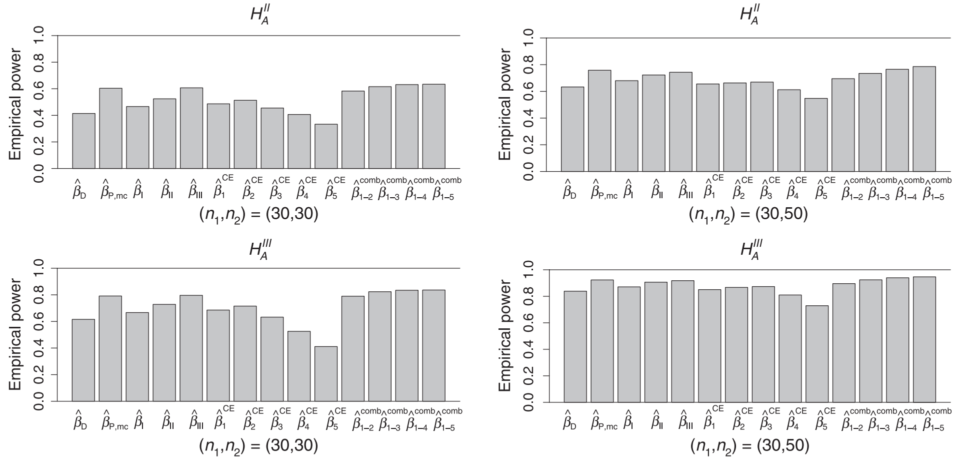

Fig. 3. The empirical power estimates of the tests under the association alternatives H_A^{II} and H_A^{III}, based on 10,000 Monte Carlo replicates for $(n_1, n_2) \in \{(30, 30), (30, 50)\}$. The empirical power notation is as in Fig. 2.

© 2009 Board of the Foundation of the Scandinavian Journal of Statistics.

be necessary to correct for edge effects under CSR independence (Dixon, 2002b; Yamada & Rogersen, 2003). Two correction methods for the edge effects on NNCT-tests, namely, buffer zone correction and toroidal correction, are investigated in Ceyhan (2007, 2008b) where it is shown that the empirical sizes of the NNCT tests are not affected by the toroidal edge correction under CSR independence. However, toroidal correction is biased for non-CSR patterns. In particular if the pattern outside the plot (which is often unknown) is not the same as that inside, it yields questionable results (Haase, 1995; Yamada & Rogersen, 2003). Under CSR independence, the (outer) buffer zone edge correction method does not change the sizes significantly for most sample size combinations. This is in agreement with the findings of Barot *et al.* (1999) who say NN methods only require a small buffer area around the study region. Once the buffer area extends past the likely NN distances (i.e. about the average NN distances), it is not adding much helpful information for NNCTs. Hence we recommend inner or outer buffer zone correction for NNCT tests with the width of the buffer area being about the average NN distance.

6. Examples

6.1. Swamp tree data

Good & Whipple (1982) considered the spatial patterns of tree species along the Savannah River, South Carolina, USA. From this data, Dixon (2002a) used a single $50\,\text{m} \times 20\,\text{m}$ rectangular plot to illustrate his tests. All live or dead trees with 4.5 cm or more dbh (diameter at breast height) were recorded together with their species. The plot contains 13 different tree species, but we conduct a 3×3 NNCT analysis for the three most frequent tree species, namely, water tupelos, black gums and Carolina ashes. The NNCT (Table 2 and Fig. 4) suggests segregation among all tree species. If segregation among the less frequent species were important, a more detailed 12×12 NNCT analysis should be performed.

The locations of the tree species can be viewed *a priori* resulting from different processes so the more appropriate null hypothesis is CSR independence, hence the inference will be a conditional one (see remark 3.1). For this data set, $Q = 472$ and $R = 454$. The test statistics and the associated p-values for the NNCT tests are presented in Table 3, where it is seen that the segregation between the species is significant. As Cuzick–Edward's tests are more sensitive to detect the clustering of the cases (i.e. the first class in the generalized framework), they are applied for each of the six different ordered pairs of tree species and the resulting test statistics and the associated p-values are also presented in Table 3, where it is seen that each species in a pair of species are significantly segregated (i.e. clustered) with respect to the other species. Furthermore, in each pair, the larger sized species seems to be more segregated from

Table 2. *The nearest neighbour contingency tables (NNCT) for swamp tree data and the corresponding percentages (in parentheses), where the cell percentages are with respect to the row sums and marginal percentages are with respect to the grand sum*

		NN			
		WT	BG	CA	Sum
Base	WT	134 (62)	47 (22)	34 (16)	215 (37)
	BG	47 (23)	128 (62)	31 (15)	206 (36)
	CA	34 (22)	27 (17)	96 (61)	157 (27)
	Sum	215 (37)	202 (35)	162 (28)	578 (100)

WT, water tupelos; BG, black gums; CA, Carolina ashes.

© 2009 Board of the Foundation of the Scandinavian Journal of Statistics.

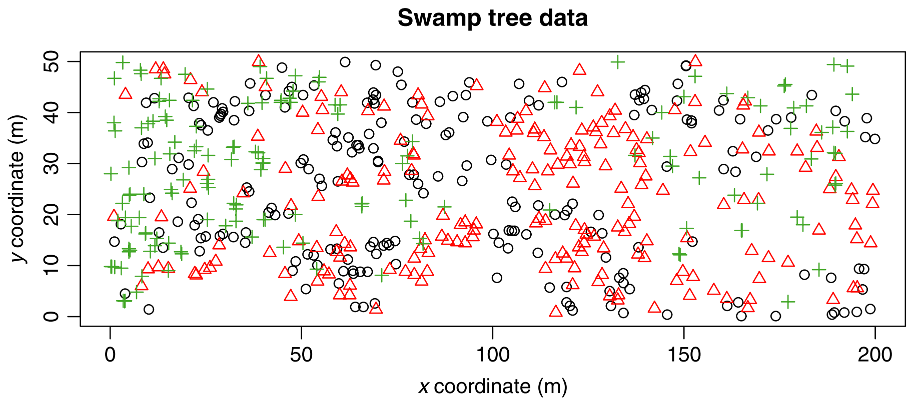

Fig. 4. The scatter plot of the locations of water tupelos (*Nyssa aquatica*, ○), black gum trees (*Nyssa sylvatica*, △), and Carolina ashes (*Fraxinus caroliniana*, +).

Table 3. *Test statistics and the associated p-values* (*in parentheses*) *for nearest neighbour contingency tables* (*NNCT*) *tests and Cuzick–Edward's tests for the swamp tree data*

NNCT tests

$\mathcal{X}_P$	$\mathcal{X}_D$	$\mathcal{X}_I$	$\mathcal{X}_{II}$	$\mathcal{X}_{III}$	$\mathcal{X}_{P,mc}$
212.20 (<0.0001)	133.48 (<0.0001)	132.13 (<0.0001)	132.42 (<0.0001)	133.20 (<0.0001)	129.16 (<0.0001)

Cuzick–Edward's k-NN tests

Species	T_1	T_2	T_3	T_4	T_5
WT versus BG	155 (<0.0001)	309 (<0.0001)	451 (<0.0001)	588 (<0.0001)	703 (<0.0001)
BG versus WT	149 (<0.0001)	279 (<0.0001)	411 (<0.0001)	529 (<0.0001)	650 (<0.0001)
WT versus CA	171 (<0.0001)	337 (<0.0001)	498 (<0.0001)	653 (<0.0001)	812 (<0.0001)
CA versus WT	108 (<0.0001)	213 (<0.0001)	297 (<0.0001)	378 (<0.0001)	455 (<0.0001)
BG versus CA	159 (<0.0001)	303 (<0.0001)	461 (<0.0001)	606 (<0.0001)	755 (<0.0001)
CA versus BG	115 (<0.0001)	216 (<0.0001)	315 (<0.0001)	410 (<0.0001)	511 (<0.0001)

Cuzick–Edward's combined tests

Species	T_{1-2}^{comb}	T_{1-3}^{comb}	T_{1-4}^{comb}	T_{1-5}^{comb}
WT versus BG	7.31 (<0.0001)	8.15 (<0.0001)	8.78 (<0.0001)	9.06 (<0.0001)
BG versus WT	7.27 (<0.0001)	7.95 (<0.0001)	8.36 (<0.0001)	8.70 (<0.0001)
WT versus CA	7.64 (<0.0001)	8.71 (<0.0001)	9.46 (<0.0001)	10.14 (<0.0001)
CA versus WT	7.29 (<0.0001)	7.93 (<0.0001)	8.27 (<0.0001)	8.56 (<0.0001)
BG versus CA	6.80 (<0.0001)	7.83 (<0.0001)	8.55 (<0.0001)	9.23 (<0.0001)
CA versus BG	7.96 (<0.0001)	8.83 (<0.0001)	9.41 (<0.0001)	9.57 (<0.0001)

$\mathcal{X}_D$ stands for Dixon's test of segregation, $\mathcal{X}_P$ for Pielou's test, $\mathcal{X}_{P,mc}$ for Pielou's test by Monte Carlo simulations, $\mathcal{X}_I$ for version I as in (6), $\mathcal{X}_{II}$ for version II as in (7), and $\mathcal{X}_{III}$ for version III as in (8). Furthermore, T_k stands for Cuzick–Edward's k-NN test for $k = 1, 2, 3, 4, 5$ and T_{1-j}^{comb} stands for the combined tests for $j = 2, 3, 4, 5$. WT, water tupelos; BG, black gums; CA, Carolina ashes.

the smaller sized species. However, the results of NNCT tests pertain to small-scale interaction at about the average NN distances; and the results of Cuzick–Edward's k-NN tests pertain to interaction at about the average k-NN distances. For the swamp tree data average NN distance (±SD) is about 1.93(±1.17) m. Further, the average k-NN distances ± SD for $k = 2, 3, 4, 5$ are $2.94 \pm 1.36, 3.81 \pm 1.41, 4.47 \pm 1.45$ and 5.10 ± 1.49, respectively. To find out the possible causes of the segregation and the type and level of interaction between the

© 2009 Board of the Foundation of the Scandinavian Journal of Statistics.

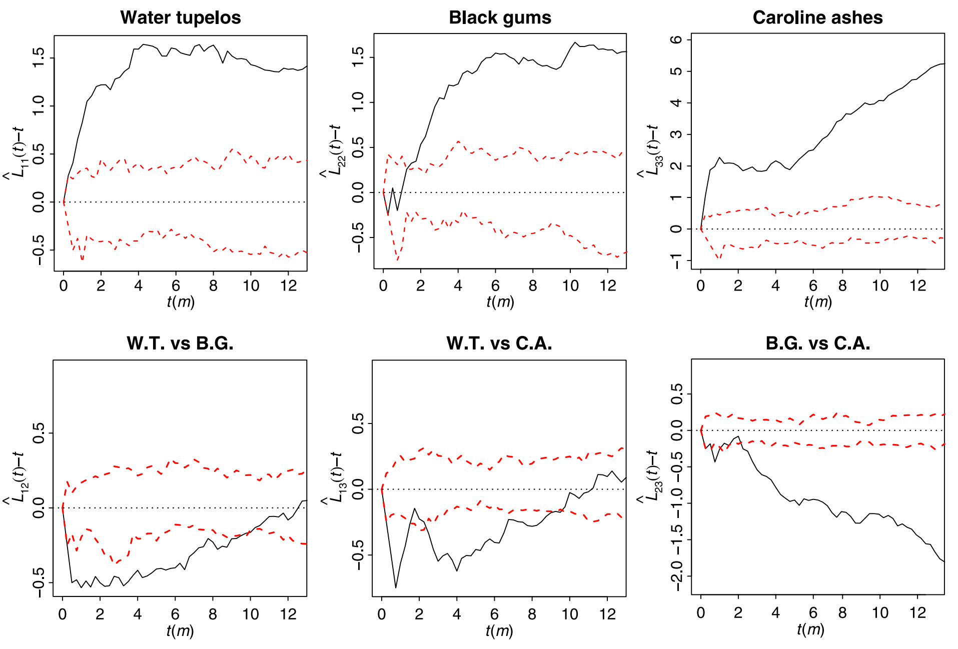

Fig. 5. Ripley's univariate L-functions $\hat{L}_{ii}(t)-t$ for $i=1,2,3$ for each species in the swamp tree data (top row) and Ripley's bivariate L-functions $\hat{L}_{ij}(t)-t$ for $i,j=1,2,3$ for the three pairs of species in the swamp tree data (bottom row). The dashed lines around 0 (which is the theoretical value for Ripley's L) are the upper and lower (pointwise) 95 per cent confidence bounds based on Monte Carlo simulation under the complete spatial randomness independence pattern. $i=1$ stands for water tupelos (WT), $i=2$ for black gums (BG) and $i=3$ for Carolina ashes (CA). Note also that the vertical axes are differently scaled.

tree species at different scales (i.e. distances between the trees), the second-order analysis of the swamp tree data is also performed. We present Ripley's (univariate) L-function (Fig. 5, top). As Ripley's L-function provides reliable results for distances smaller than the average NN distance (Ceyhan, 2009b), we only consider distances up to around 2m for the swamp tree data. Black gums seem less aggregated at small distances than other species. Hence segregation of the species might be because of different levels and types of aggregation of the species in the study region.

Ripley's bivariate L-function, $\hat{L}_{ij}(t)$, for each pair of tree species are also calculated and presented in Fig. 5 (bottom). Owing to the theoretical symmetry of $L_{ij}(t)$, only plots for three different pairs are presented. The other three plots are omitted since they are very similar in practice also. Again we only consider distances up to around 2 m. Observe that water tupelos and black gums exhibit significant segregation for $t \lesssim 2$ m; water tupelos and Carolina ashes are significantly segregated up to about $t \approx 1.8$ m; black gums and Carolina ashes are significantly segregated for $t<1$ m. The bivariate pair correlation functions are presented in Ceyhan (2008a).

6.2. Leukaemia data

Cuzick & Edwards (1990) considered the spatial locations of 62 cases of childhood leukaemia in the North Humberside region in United Kingdom, between the years 1974 and 1982 (inclusive) (Fig. 6, left). A sample of 143 controls are selected using the completely randomized design from the same region. It is reasonable to assume that some process affects

© 2009 Board of the Foundation of the Scandinavian Journal of Statistics.

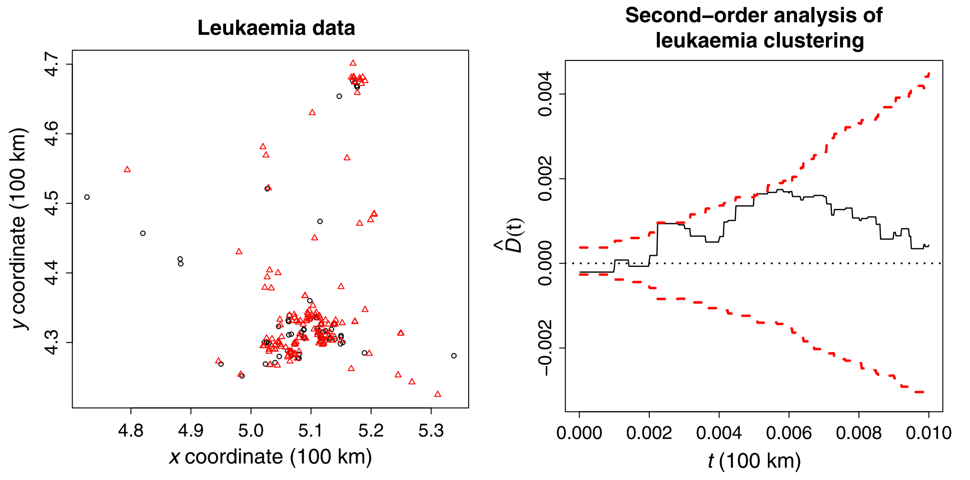

Fig. 6. To the left are the scatter plots of the locations of cases (circles ∘) and controls (triangles △) in the North Humberside leukaemia data set. To the right is Diggle's modified bivariate K-function $\hat{D}(t)=\hat{K}_{11}(t)-\hat{K}_{22}(t)$ with $i=1$ for controls and $i=2$ for leukaemia cases. The dashed lines around 0 are the upper and lower (pointwise) 95 per cent confidence bounds for the $\hat{D}(t)$ under random labelling of cases and controls.

Table 4. *The nearest neighbour contingency table (NNCT) for the North Humberside leukaemia data and the corresponding percentages (in parentheses)*

		NN		
		Case	Control	Sum
Base	Case	25 (38)	41 (62)	66 (30)
	Control	39 (26)	113 (74)	152 (70)
	Sum	64 (29)	154 (71)	218 (100)

a posteriori the population of North Humberside region so that some of the individuals get to be cases, while others continue to be healthy (i.e. they are controls). So the more appropriate null hypothesis is the RL pattern. The percentages in the diagonal cells of the NNCT (Table 4) are about the same as the marginal (row or column) percentages of the subjects in the study, which might be interpreted as the lack of any deviation from RL for both classes. Figure 6 (right) is also supportive of this observation. For this data set, we calculate $Q=152$ and $R=142$. None of the NNCT tests yields a significant result (Table 5). On the other hand, Cuzick–Edwards T_k are all significant for $k>1$, and so are all combined tests T_S^{comb}.

Based on these tests, it is seen that the cases and controls do not exhibit significant clustering (i.e. segregation). However, there seems to be a large scale segregation of the cases at about k-NN distances with $2\leq k\leq 5$. The locations of the subjects in this population (cases and controls together) seem to be from an inhomogeneous Poisson process (see also Fig. 6). Hence we also provide Diggle's D-function (Fig. 6, right) which indicates mild clustering of diseases (i.e. segregation of cases from controls) at distances about 200 and 600 m. At smaller scales, the plot in Fig. 6 is consistent with the results of the NNCT analysis. The average k-NN distances ± SDs for $k=1,2,\ldots,5$ are $700\pm1400, 1342\pm2051, 1688\pm2594, 2152\pm3188$ and 2495 ± 3810 m, respectively. Notice that the distribution of NN distances is very skew in this case, as their SDs are greater than their means. As NN- and distance-based methods

© 2009 Board of the Foundation of the Scandinavian Journal of Statistics.

Table 5. *Test statistics and the associated p-values for nearest neighbour contingency table (NNCT) tests and Cuzick–Edward's k-NN and combined tests for North Humberside leukaemia data. The notation for the tests is as in Table* 3

NNCT tests

$\mathcal{X}_P$	$\mathcal{X}_D$	$\mathcal{X}_I$	$\mathcal{X}_{II}$	$\mathcal{X}_{III}$	$\mathcal{X}_{P,mc}$
3.31	2.25	1.98	2.10	2.13	2.02
(0.0687)	(0.3249)	(0.1599)	(0.3505)	(0.1449)	(0.1547)

Cuzick–Edward's tests

T_1	T_2	T_3	T_4	T_5	T^{comb}_{1-2}	T^{comb}_{1-3}	T^{comb}_{1-4}	T^{comb}_{1-5}
25	53	78	95	116	2.12	2.46	2.53	2.58
(0.0647)	(0.0043)	(0.0014)	(0.0093)	(0.0099)	(0.0170)	(0.0068)	(0.0057)	(0.0048)

simply address different features, it is likely for them to yield different results. In this particular case, Cuzick–Edward's tests appear to be more powerful (than either NNCT tests or the D-function).

7. Discussion and conclusions

In this article, multivariate clustering (or segregation) tests based on NNCTs are discussed. Three new overall segregation tests are proposed using the correct (asymptotic) distribution of cell counts in NNCTs. It is shown that NNCT tests are conditional on Q and R under CSR independence, but unconditional under RL. When sample sizes are small (i.e. the corresponding cell counts are ≤ 5), the asymptotic approximation of the NNCT tests is not appropriate. So Dixon(1994) recommends Monte Carlo randomization for his test when some cell count(s) are ≤ 5 in a NNCT. This recommendation extends for all the NNCT tests introduced in this article. For Cuzick–Edward's tests, Monte Carlo randomization is recommended when $n_1 < 10$; otherwise asymptotic approximations can also be employed. Furthermore, k-NN tests for $k > 1$ and combined tests attain the normal approximation at smaller sample size combinations (i.e. their distribution approaches to normality faster) compared to that of the NNCT tests. The newly proposed NNCT tests and Dixon's test address the same question about the spatial interaction between the classes. On the other hand, Cuzick–Edward's test is designed for the clustering of the cases (or the first class in the general framework), so NNCT tests and Cuzick–Edward's tests answer somewhat different questions.

Considering the empirical significance levels and power estimates, among the NNCT tests, version III of the new NNCT tests is recommended when testing for segregation or association. When all tests are considered, under segregation, T^{comb}_{1-4} or T^{comb}_{1-5} have the best performance, hence either can be recommended against the segregation alternatives. However given the computational cost of combined tests for larger k values, T_k with $k=4$ or 5 is recommended for the segregation alternatives instead of T^{comb}_{1-4} or T^{comb}_{1-5}. Under association, the power estimates of Cuzick–Edwards k-NN tests tend to decrease as k increases. Among Cuzick–Edwards combined tests, usually T^{comb}_{1-5} has the best power performance. Considering all tests together, still T^{comb}_{1-5} has the best power performance. However, given the computational cost of combined tests, one might prefer version III of the NNCT tests under the association alternatives for larger samples, as its power is very close to that of T^{comb}_{1-5}.

When testing against CSR independence or RL, NNCT tests provide information about the spatial interaction for small scales up to about the average NN distance. Cuzick–Edward's k-NN tests provide information about the k-NN distance. On the other hand, the pair correlation function $g(t)$ and Ripley's classical K- or L-function and other variants provide information

© 2009 Board of the Foundation of the Scandinavian Journal of Statistics.

on the pattern at various scales. Furthermore, Diggle's D-function provides the level of spatial interaction (or clustering of a class when compared with another) at all scales when used against the RL pattern. Ripley's classical K- or L-function can be used when the null pattern assumes first-order homogeneity for each class. When the null pattern is the RL of points from an inhomogeneous Poisson process they are not appropriate (Kulldorff, 2006); Cuzick–Edward's k-NN tests are designed for testing bivariate spatial interaction and are mostly used for spatial clustering of cases in epidemiology; Diggle's (2003) D-function adjusts for any inhomogeneity in the locations of, for example, cases and controls. Furthermore, there are variants of $K(t)$ that explicitly correct for inhomogeneity (Baddeley *et al.*, 2000). The NNCT tests can be used for the multivariate spatial interaction between two or more classes. Ripley's L-function has univariate and bivariate versions while Diggle's D-function is designed for bivariate pattern analysis. A practical concern about these tests is the lack of code in some statistical software in a way that others could use. The methods outlined here have been implemented in *R* version 2.6.2, and the relevant code is available from the author upon request.

The NNCT tests and Ripley's L-function provide similar information in the two-class case at small scales. For $q>2$ classes, overall tests provide information on the (small-scale) multivariate spatial interaction in one compound summary measure; while the Ripley's L-function requires performing all bivariate spatial interaction analysis. There are class-specific NNCT tests and cell-specific NNCT tests that can serve as *post hoc* analysis only when the overall test is significant (Ceyhan, 2009a,b). To the author's knowledge, all other multivariate spatial pattern tests are designed to test the spatial interaction between two classes, hence can serve as pairwise interaction tests for $q>2$ classes. Usually, the tests involve a parameter, which forces the user to adjust for multiple testing, but the NNCT tests by construction avoid the problem of multiple testing. As pairwise analysis of q classes with 2×2 NNCTs might yield conflicting results compared to $q\times q$ NNCT analysis (Dixon, 2002a), Ripley's L-function and NNCT tests might also yield conflicting results. Hence Ripley's L-function and NNCT tests may provide similar but not identical information about the spatial pattern and the latter might detect small-scale interaction that is missed by the former. As the pair correlation functions are derivatives of Ripley's K-function, most of the above discussion holds for them also, except that $g(t)$ is reliable only for large scale interaction analysis (Stoyan & Stoyan, 1996). Hence, NNCT tests and the pair correlation function are not comparable but provide complementary information about the pattern in question.

The order of classes in the construction of the NNCTs is irrelevant for the values of the test statistics hence for the results of the NNCT tests. However, by construction, Cuzick–Edward's tests are more sensitive for clustering of the cases in a case/control framework (or the first class that is treated as cases in the generalized two-class framework). That is, if one reverses the role of cases and controls in a data set, the test statistics might give different results. In theory, Ripley's bivariate K- or L-function is symmetric in the classes it pertains to; but due to edge corrections, it might be slightly asymmetric in practice.

For a data set for which CSR independence is the reasonable null pattern, we recommend the overall segregation test if the question of interest is the spatial interaction at small scales (i.e. about the mean NN distance). If it yields a significant result, then to determine which pairs of classes have significant spatial interaction, the class- and cell-specific tests can be performed. One can also perform Ripley's K- or L-function and only consider distances up to around the average NN distance and compare the results with those of the NNCT analysis. If the spatial interaction at higher scales is of interest, the pair correlation function is recommended (Stoyan & Stoyan, 1996; Ceyhan, 2009b). On the other hand, if the RL pattern is the reasonable null pattern for the data, we recommend the NNCT tests if the small-scale interaction is of interest and Diggle's D-function if the spatial interaction at higher scales is also of interest.

© 2009 Board of the Foundation of the Scandinavian Journal of Statistics.

Acknowledgements

The author thanks the editors and anonymous referees whose constructive remarks and suggestions greatly improved the presentation and flow of this article. Most of the Monte Carlo simulations presented in this article were executed on the Hattusas cluster of Koç University High Performance Computing Laboratory.

References

Baddeley, A., Møller, J. & Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statist. Neerlandica* **54**, 329–350.

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

Ceyhan, E. (2007). Edge correction for segregation tests based on nearest neighbor contingency tables. In *Proceedings of the Applied Statistics 2007 International Conference*. Ribno (Bled), Slovenia.

Ceyhan, E. (2008a). New tests for spatial segregation based on nearest neighbor contingency tables; arXiv:0808.1409v1 [stat.ME]. Technical Report # KU-EC-08-6. Koç University, Istanbul, Turkey.

Ceyhan, E. (2008b). On the use of nearest neighbor contingency tables for testing spatial segregation. *Environ. Ecol. Stat.* doi:10.1007/s10651-008-0104-x.

Ceyhan, E. (2008c). *QR*-adjustment for clustering tests based on nearest neighbor contingency tables; arXiv:0807.4231v1 [stat.ME]. Technical Report # KU-EC-08-5. Koç University, Istanbul, Turkey.

Ceyhan, E. (2009a). Class-specific tests of segregation based on nearest neighbor contingency tables. *Statist. Neerlandica* **63**, 149–182.

Ceyhan, E. (2009b). Overall and pairwise segregation tests based on nearest neighbor contingency tables. *Comput. Statist. Data Anal.* **53**, 2786–2808.

Coomes, D. A., Rees, M. & Turnbull, L. (1999). Identifying aggregation and association in fully mapped spatial data. *Ecology* **80**, 554–565.

Cox, T. F. (1981). Reflexive nearest neighbors. *Biometrics* **37**, 367–369.

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

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

Dixon, P. M. (1994). Testing spatial segregation using a nearest-neighbor contingency table. *Ecology* **75**, 1940–1948.

Dixon, P. M. (2002a). Nearest-neighbor contingency table analysis of spatial segregation for several species. *Ecoscience* **9**, 142–151.

Dixon, P. M. (2002b). Nearest neighbor methods. In *Encyclopedia of Environmetrics*, vol. **3** (eds A. H. El-Shaarawi & W. W. Piegorsch), 1370–1383. John Wiley & Sons Ltd., New York.

Good, B. J. & Whipple, S. A. (1982). Tree spatial patterns: South Carolina bottomland and swamp forests. *Bull. Torrey Botanical Club* **109**, 529–536.

Goreaud, F. & Pélissier, R. (2003). Avoiding misinterpretation of biotic interactions with the intertype K_{12}-function: population independence vs. random labelling hypotheses. *J. Veg. Sci.* **14**, 681–692.

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

Kulldorff, M. (1997). A spatial scan statistic. *Comm. Statist. Theory Methods* **26**, 1481–1496.

Kulldorff, M. (2006). Tests for spatial randomness adjusted for an inhomogeneity: a general framework. *J. Amer. Statist. Assoc.* **101**, 1289–1305.

van Lieshout, M. N. M. & Baddeley, A. J. (1996). A nonparametric measure of spatial interaction in point patterns. *Statist. Neerlandica* **50**, 344–361.

van Lieshout, M. N. M. & Baddeley, A. J. (1999). Indices of dependence between types in multivariate point patterns. *Scand. J. Statist.* **26**, 511–532.

Meagher, T. R. & Burdick, D. S. (1980). The use of nearest neighbor frequency analysis in studies of association. *Ecology* **61**, 1253–1255.

Moran, P. A. P. (1948). The interpretation of statistical maps. *J. Roy. Statist. Soc. Ser. B* **10**, 243–251.

Perry, G., Miller, B. & Enright, N. (2006). A comparison of methods for the statistical analysis of spatial point patterns in plant ecology. *Plant Ecol.* **187**, 59–82.

Pielou, E. C. (1961). Segregation and symmetry in two-species populations as studied by nearest-neighbor relationships. *J. Ecol.* **49**, 255–269.

Ripley, B. D. (2004). *Spatial statistics*. Wiley-Interscience, New York.

© 2009 Board of the Foundation of the Scandinavian Journal of Statistics.

Searle, S. R. (2006). *Matrix algebra useful for statistics*. Wiley-Intersciences, New York.

Song, C. & Kulldorff, M. (2003). Power evaluation of disease clustering tests. *International J. Health Geogr.*, **2**, 9–16.

Stoyan, D. & Penttinen, A. (2000). Recent applications of point process methods in forestry statistics. *Statist. Sci.* **15**, 61–78.

Stoyan, D. & Stoyan, H. (1994). *Fractals, random shapes and point fields: methods of geometrical statistics*. John Wiley and Sons, New York.

Stoyan, D. & Stoyan, H. (1996). Estimating pair correlation functions of planar cluster processes. *Biom. J.* **38**, 259–271.

Wiegand, T., Gunatilleke, S. & Gunatilleke, N. (2007). Species associations in a heterogeneous Sri Lankan dipterocarp forest. *Am. Nat.* **170**, 77–95.

Yamada, I. & Rogersen, P. A. (2003). An empirical comparison of edge effect correction methods applied to K-function analysis. *Geogr. Anal.* **35**, 97–109.

Received December 2006, in final form May 2009

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

© 2009 Board of the Foundation of the Scandinavian Journal of Statistics.