1-2015

Restricted random labeling: testing for between-group interaction after controlling for joint population and within-group spatial structure

Barry J. Kronenfeld
Eastern Illinois University

Timothy F. Leslie
George Mason University

Restricted Random Labeling: Testing for between-group interaction after controlling for joint population and within-group spatial structure

Barry J. Kronenfeld and Timothy F. Leslie

Publication preprint. The published version is available at Springer via
http://dx.doi.org/10.1007/s10109-014-0206-y

Abstract

Statistical measures of spatial interaction between multiple types of entities are commonly assessed against a null model of either toroidal shift (TS), which controls for spatial structure of individual subpopulations, or random labeling (RL) which controls for spatial structure of the joint population. Neither null model controls for both types of spatial structure simultaneously, although this may sometimes be desirable when more than two subpopulations are present. To address this, we propose a flexible framework for specifying null models that we refer to as restricted random labeling (rRL). Under rRL, a specified subset of individuals is restricted, and other individuals are randomly relabeled. Within this framework two specific null models are proposed for pairwise analysis within populations consisting of three or more subpopulations, to simultaneously control for spatial structure in the joint population and one or the other of the two subpopulations being analyzed. Formulas are presented for calculating expected nearest neighbor counts and co-location quotients within the proposed framework. Differences between TS, RL and rRL are illustrated by application to six types of generating processes in a simulation study, and to empirical datasets of tree species in a forest and crime locations in an urban setting. These examples show that rRL null models are typically stricter than either TS or RL, which often detect "interactions" that are an expected consequence either of the joint population pattern or of individual subpopulation patterns.

Keywords:

spatial interaction; random labeling; toroidal shift; null models; cross *K*-function

1 Introduction

The spatial interaction between populations of individuals belonging to different categories is a fundamental, but often overlooked aspect of geographical analysis (Cromley et al. 2014). It has been used to understand patterns of a wide variety of spatial entities, including biological species (Wiegand et al. 2007), disease occurrences (Souris and Bichaud 2011), businesses (Leslie and Ó hUallacháin 2006, Ó hUallacháin and Leslie 2013), and even astronomical objects (Stoica 2010). The concept goes by several different names in the literature, including spatial dependence (Liu et al. 2007), spatial interaction (Goreaud and Pélissier 2003, Fuller and Enquist 2012), bivariate spatial association (Souris and Bichaud 2011) and co-location (Huang et al. 2006, Leslie and Kronenfeld 2011). The term spatial segregation (Ceyhan 2009) is also

used to refer to negative interaction. In what follows, the term spatial interaction is used generally due to its popularity in the literature, but we also refer to a specific metric of co-location as defined in Leslie and Kronenfeld (2011). The concept of spatial interaction as used here is distinct from cross-correlation in that it relates to qualitative marks (i.e. categories, species or subpopulations) rather than numerical values.

An important challenge for researchers investigating patterns of spatial interaction is determining an appropriate null model. When three or more categorical populations are present, null model selection is complicated by the existence of multiple, confounding processes, including processes that affect the entire joint population (Ceyhan 2008, Getzin et al. 2008), subpopulations individually (Wiegand et al. 2007) and pairs of subpopulations (e.g. spatial interaction). An appropriate null model should randomize those elements of the problem that pertain to the research question, while holding unrelated elements constant. Failure to isolate the specific effect of interest to the researcher will result in a hypothesis test that is likely too liberal, leading to Type I statistical errors (Fuller and Enquist 2012).

Two commonly used null models are toroidal shift (*TS*) and random labeling (*RL*), which constrain the characteristics of individual subpopulations and the joint population, respectively. These null models are straightforward in their definition and easy to implement. When more than two subpopulations are present, however, it may be necessary to constrain characteristics of both types of patterns simultaneously. For example, a researcher investigating the spatial interaction between two tree species in a heterogeneous forest might wish to exclude the effects of the overall forest pattern, as well as the tendency for one species to grow in clusters. In such situations, an approach that combines the characteristics of *TS* and *RL* would be useful. At present, this can only be accomplished with more sophisticated methods such as pattern reconstruction (Wiegand and Moloney 2014), which requires detailed modeling of the spatial characteristics of the joint population and each subpopulation. We seek a simpler generic approach that is broadly applicable to exploratory and confirmatory analysis.

Randomization restrictions are a powerful tool for developing alternative null models (Manly 1991). In a Monte Carlo simulation environment, randomization defines the process by which the null sampling distribution is determined empirically. Thus, changes to the randomization process have the effect of changing the null model. *RL* restricts the randomization process to retain the locations in the joint pattern. Fuller and Enquist (2012) further restrict label randomization to predefined cells to identify departures from randomness at a local rather than global scale. In a study of plant-plant interactions, Wiegand et al. (2007) restricted one species while randomly relocating plants of other species to distinguish between first- and second-order effects. These randomization restrictions have enabled a more nuanced interpretation of spatial pattern than *CSR* or *RL*. Nevertheless, a general theory of randomization restrictions and their use in null model specification is lacking.

In this study, we introduce a generalized method of *restricted random labeling (rRL)* that allows specification of a range of null models for co-location analysis. *rRL* allows for any subset of the population to be restricted, and a general formula is presented to compute the expected

nearest neighbor count for any restriction set. We suggest two useful specifications of *rRL* appropriate for populations consisting of more than 2 subpopulations, which entail restricting the first and second subpopulation, respectively, in each pairwise analysis. Envisioning the set of pairwise analyses as a matrix with subpopulations assigned to rows and columns, we refer to these two specifications as $rRL^{\boxed{row}}$ and $rRL^{\boxed{col}}$. These variants provide a useful combination of the properties of *TS* and *RL*, exposing only those interaction patterns that are not logically implied by either the joint population pattern or the patterns of individual subpopulations.

To demonstrate the proposed null models, we first review methods of co-location analysis for marked point patterns in Section 2. Next we define *rRL* and derive a general formula to compute expected nearest neighbor counts under each model in Section 3. Specific formulas and significance testing procedures for $rRL^{\boxed{row}}$ and $rRL^{\boxed{col}}$ are also introduced. To illustrate the type of knowledge that can be gained from *rRL* analysis, Sections 4 and 5 describe a case study comparing *rRL* with both *TS* and *RL* for three simulated pattern types and two empirical datasets. Spatial interaction is reported as a restricted co-location quotient comparing expected nearest neighbor counts under each null model with observed counts, along the lines of Leslie and Kronenfeld (2011), and significance testing is implemented through Monte Carlo simulation.

We conclude with a discussion of potential uses and limitations of *rRL* and suggest opportunities for future research. The methods presented in this paper are implemented in a software package that is freely available from the authors upon request.

2 Spatial Interaction Statistics

Different approaches have been developed to measure co-location for data with different geometries. For example, the join count statistic is commonly used to measure co-location in polygon data (Cliff and Ord 1981). We are interested in qualitatively marked point patterns, where individual points are classified using a finite set of labels. Examples of marked point patterns used in co-location analysis include locations of different types of businesses (Leslie and Ó hUallacháin 2006), biological species (Law et al. 2009) and disease occurrences (Leibovici et al. 2011, Souris and Bichaud 2011). For our purposes, we refer to a marked point pattern as a *(joint) population* consisting of several *groups* or *subpopulations* of identically marked individuals.

2.1 Pairwise Metrics

Two common and intuitive measures of co-location can be derived from alternate definitions of proximity. If proximity is defined in terms of raw distance, co-location measurement entails counting the number of individuals from one group that are found within a specified distance *h* of another group. This is the basis for the cross *K-function*, a variation of Ripley's *K-function* (Lotwick and Silverman 1982). More will be said about the cross *K-function* and its variants below.

An alternative to distance-based neighborhood definition is to define proximity in terms of nearest neighbor relations (Pielou 1961, Cuzick and Edwards 1990). The observed frequency of nearest-neighbor instances among all pairs of groups can be captured in a nearest neighbor contingency table (*NNCT*) (Dixon 1994). A single cell in the *NNCT* denotes the nearest neighbor count between the row and column subpopulations, or more specifically, the number of individuals in the row subpopulation that have an individual in the column subpopulation as their nearest neighbor. Leslie and Kronenfeld (2011) extend this definition to allow for multiple equidistant nearest neighbors: if an individual has n equidistant nearest neighbors, then each contributes $1/n$ to the corresponding cell in the *NNCT*. In this manner, all row individuals effectively contribute one neighbor to the total nearest neighbor count. As noted by Pielou (1961), for a given individual *I*, there is a maximum of five other individuals for whom *I* will be the (exclusive) nearest neighbor due to geometric constraints. This affects the maximum degree of spatial interaction that can be observed using *NNCT* analysis (Leslie and Kronenfeld 2011).

Conventional reporting methods for distance- and neighbor-based metrics of co-location differ. Results of cross *K-function* analysis are typically presented for one pair of classes at a time as a graph over the range of distance values, overlaid onto upper and lower bounds of the simulation. The result is a set of three lines, one representing the observed level of co-location at different scales, and the others showing the simulation envelope under the null model. Observed values outside the envelope provide evidence to reject the null hypothesis, though proper inferential analysis must take into account the problem of multiple hypothesis testing (Loosmore and Ford, 2006). In contrast, *NNCT* counts are usually presented in tabular form to show all pairwise relations simultaneously. To aid in interpretation, a co-location quotient (*CLQ*) for each cell can be computed as the ratio of observed to expected counts (Leslie and Kronenfeld 2011). The *CLQ* always has the same expected value (=1) allowing for consistent interpretation across class pairs, and is analogous to the *location quotient* from which its name is derived (Florence 1944).

Despite differences in reporting methods, Dixon (1994) noted that distance- and neighbor-based definitions of proximity are related and usually yield similar results, especially when the overall population is evenly distributed over space. However, some important differences exist. When density is inhomogeneous, isolated individuals will have relatively less influence on the cross *K-function* than the *CLQ*, while individuals in clusters will have relatively more influence. Also, unlike the cross *K-function*, the *CLQ* is theoretically asymmetrical: the observed number of *A* individuals with *B* as their nearest neighbor can be different from the number of *B* individuals with *A* as their nearest neighbor. This asymmetry may seem counterintuitive at first, but effectively captures asymmetrical relations between groups that differ substantially in size or spatial isolation (Leslie and Kronenfeld 2011).

2.2 Null Models

To determine if a bivariate pattern of spatial association is significantly different from chance, it is necessary to compare an observed metric value with the corresponding expectation under an appropriate null model. Although the expectation can sometimes be calculated analytically, this is not always the case and in practice Monte Carlo simulation is typically used to determine the

sampling distribution (e.g. Liu et al. 2007). Three common null models for determining expected association patterns are complete spatial randomness (*CSR*), population independence (*PI*) and random labeling (*RL*). Of these, *CSR* is often considered unrealistic, and *PI* and *RL* are generally favored.

In *PI*, the null hypothesis is that the processes generating each population are independent (Lotwick and Silverman 1982). One common method for implementing *PI* is toroidal shift (*TS*), in which each subpopulation is shifted by a randomly distance and direction, treating the study area as a torus (Lotwick and Silverman 1982). *TS* is limited in that the study area must be rectangular, and even then can produce artifacts at study area edges (Wiegand and Moloney 2014).

A more flexible approach to implementing *PI* is to use a specific point generating process fitted to reproduce the structure of each subpopulation independently. Parametric point processes can be used (e.g. Wiegand and Moloney 2004, John et al. 2007), but it can be quite challenging to identify a process that provides a good fit to the observed pattern. An alternative approach is to attempt to reconstruct the pattern by identifying important summary statistics and then using a non-parametric algorithm to generate point patterns that match these statistics within some defined degree of allowable deviation. Wiegand and Moloney (2014) show that such an approach can successfully match a wide array of empirical point patterns when conditioned on an intensity function to capture broad-scale spatial structure. Further, they found that the required summary statistics were largely consistent across patterns, meaning that the approach can be generally applied.

While the reconstruction approach is more generally applicable than parametric process fitting, it does has some drawbacks. Considerable expertise is required to select summary statistics, designate an acceptable deviation limit and implement an annealing algorithm. Conditioning on the intensity function is necessary to replicate inhomogeneous patterns, but the results are affected by the scale at which the intensity function is defined. Also, the algorithm requires iterative replacement of individual points and thus is computationally demanding, which may limit application to large datasets.

In *RL*, the null hypothesis is that locations are generated by a single common process but individuals are randomly assigned to subpopulations (Goreaud and Pélissier 2003). Unlike the null hypothesis of *PI*, which can be implemented in many ways including but not limited to *TS*, the concept and implementation of *RL* are rarely separated. The name indicates the simulation procedure in which observed locations are held constant, but class marks are randomly reassigned to each location.

Some variations on *RL* have been developed that have similarities to the general method we propose below. Trivariate random labeling (de la Cruz et al. 2008, Wiegand and Moloney 2014) involves random labeling of two groups while holding a third group constant. Proposed as a means to investigate whether a pattern of event labels (e.g. mortality) in one group (e.g. seedlings) is affected by the locations of another group (e.g. mature trees), the method has not been generalized or extended to scenarios involving more than three groups. Researchers have

also combined elements of *RL* and *PI*. For example, Wiegand et al. (2007) developed a procedure to assess 2nd order spatial interaction by restricting the locations of individuals of one species and randomizing the locations of the other using a heterogeneous Poisson process. As with the reconstruction approach described above, this procedure requires expertise in determining an appropriate intensity function for the joint population.

2.3 Choosing between PI and RL

Demonstrating substantial differences between *TS* and *RL* for several simulated and empirical datasets, Goreaud and Pélissier (2003) proposed a framework for choosing between *PI* and *RL* that forms the basis for null model selection in many research studies (de la Cruz et al. 2008), based on the temporal sequence in which locations and marks are established in the hypothesized generating process. In *a priori* marked point processes, individual marks are determined prior to location establishment, whereas in *a posteriori* marked processes categories are assigned (or re-assigned) after locations have already been established. Under this framework, *PI* and *RL* are considered appropriate for *a priori* and *a posteriori* marked processes, respectively.

Although the distinction between *a priori* and *a posteriori* marking processes is useful for selecting a null model, limitations exist. Not all patterns can be clearly identified as resulting from *a priori* or *a posteriori* processes. Ambiguities can arise due to lack of knowledge as well as the presence of multiple processes at work simultaneously (Goreaud and Pélissier 2003). These issues can be subtle and may be perspective dependent. For example, if commercial real estate transactions are analyzed, commercial industry analysts might consider the business activity (i.e. marks) as being determined *a priori*, and wish to evaluate the locations in which businesses choose to site these activities. From the perspective of a city planner, however, it may be the locations that are seen as being determined *a priori*. In any case, the distinction is largely made on theoretical grounds. As far as we are aware, no empirical test exists to discriminate between *a priori* and *a posteriori* marked patterns.

Even if a pattern can be clearly identified as resulting from an *a priori* process, it can be difficult to implement the *TS* model due to boundary effects. *TS* effectively splits each subpopulation along the original boundaries, which cross through the interior of the subpopulation after it has been shifted. When the study area is not rectangular, the method becomes ill-defined. Goreaud and Pélissier (2003) suggest shifting only one subpopulation in a pairwise analysis, but it is not clear how this addresses the problem of defining a toroid from a non-rectangular region.

The above limitations suggest the need for additional null models to handle real-world data. To motivate our own proposal, we offer an alternative interpretation of the empirical results of Goreaud and Pélissier (2003), and in particular the large differences they found between *TS* and *RL* for some datasets.

2.4 Joint Population vs. Subpopulation Patterns

Our interpretation is based on the well-established concept of restricted randomization (Manly 1991), in which a null model is generated by randomizing some aspects of the pattern of interest while holding other aspects constant. In the context of multitype marked point patterns, three aspects should be considered: (a) the pattern of the joint population, (b) the patterns of each individual subpopulation, and (c) patterns of interaction between subpopulations. Patterns of type (c) are of primary interest in spatial interaction analysis and so must be randomized, but can be strongly affected by patterns of types (a) and (b). The null models of *TS* and *RL* differ as to whether (a) or (b) are restricted during the randomization process. Under *RL* the pattern of the joint population is held constant, since point locations are not altered. However, the patterns of individual subpopulations are randomized in the relabeling process. In contrast, except for the effect of the study area boundary mentioned above, patterns of individual subpopulations are held constant under *TS* while the pattern of the joint population is allowed to vary.

To support the utility of this interpretation, we use it to interpret the results of Goreaud and Pélissier (2003). First, we consider the expected value of the interaction metric L_{12} they used, estimated empirically as

$$\hat{L}_{12} = \sqrt{\frac{C_{12}(R)}{\pi N_1 \lambda_2}} - r\,. \qquad (1)$$

where N_1 is the number of Type 1 individuals, $\lambda_2 = N_2/S$ is density of Type 2 individuals within the study area (of size S), and $C_{12}(R)$ is the number of Type 1 – Type 2 pairs of individuals within distance r of each other; we omit edge correction factors for simplicity. Under *TS* the expectation of $\hat{L}_{12}$ is zero because following randomization of pattern 1, each Type 1 individual will have on average $\pi r^2 \lambda_2$ Type 2 individuals within radius r. Under *RL*, on the other hand, for a binary marked population the expectation of $\hat{L}_{12}$ is entirely determined by the pattern of the joint population. Thus, one would expect the envelopes of $\hat{L}_{12}$ under *TS* vs. *RL* to be similar when the joint population is random, but to differ substantially when the joint pattern departs strongly from *CSR*. Indeed, we note that in all examples in which Goreaud and Pélissier (2003) found substantial differences between *TS* and *RL*, the pattern of the joint population appears to be significantly different from random: in their simulated stands *B* and *E*, the joint pattern is highly clustered and the simulated values of L_{12} is much higher for *RL* than *TS*; conversely, their simulated stands *C* and *F* appear more uniform than random, and the simulated value of L_{12} is lower for *RL* than *TS* for these stands. This interpretation is also consistent with Ceyhan's (2008) observation that differences between *CSR* and *RL* are small when the joint population closely resembles a random pattern.

The above interpretation is not entirely dissimilar to the distinction between *a priori* and *a posteriori* marking made by Goreaud and Pélissier (2003). However, it provides more flexibility because the effects of the joint population and individual subpopulations are considered

independently. In particular, if *TS* and *RL* respectively hold constant the patterns of individual subpopulations and of the joint population, the obvious question arises as to whether it is possible to hold both types of patterns constant simultaneously. Such a null model would answer the question, *is the observed pattern of spatial association between A and B different from what would be expected given both the pattern of the joint population and the patterns of each subpopulation individually?*

There are many situations in which such a null model might be useful. For example, consider the relationships between different housing types, businesses or crimes. All of these are strongly related to the artificial landscape, creating *a priori* constraints on the joint population. At the same time, individual types of housing, businesses and crimes often tend to cluster in space. A researcher seeking to understand the spatial interaction between two different types of housing, businesses or crimes would likely want to account for this tendency for individual types to cluster. Similarly, in biological settings, individual subpopulations will obviously cluster in space due to shared biological traits, but there may also be environmental characteristics that affect all subpopulations simultaneously, such as water bodies, variations in soil fertility, etc. Indeed, the need to restrict randomization of the joint population has been acknowledged in contexts that are not strictly *a posteriori* marked. For example, in case-control designs commonly used in epidemiological studies, controls are taken from the population at large rather than the complete spatial domain to account for environmental factors common to all subpopulations (Wiegand and Moloney 2014).

3 Restricted Random Labeling

We seek to understand the effects that spatial structure of the joint population and individual subpopulations has on our understanding of spatial interaction between subpopulations. Our approach builds on the fact that the joint population pattern is not perturbed in random labeling (*RL*). The objective is to further restrict *RL* so as to hold constant the subpopulation pattern(s) of interest.

To begin, let $rRL^{\boxed{S}}$ denote a null model in which individuals in an arbitrarily defined population subset *S* are restricted during the random labeling process. The restriction subset is considered to be a component of the null model. The term *restricted random labeling* (*rRL*) is proposed to refer generally to random labeling with any such restriction set.

As an example of *rRL*, consider the marked point pattern shown in Fig. 1a, in which subpopulation *C* (black circles) is dispersed throughout the lower right half of the study area. When *RL* is used as a null model, the observed pattern is compared to patterns created by randomly rearranging all point labels (see Fig. 1b). In contrast, random labeling with subpopulation *C* restricted (denoted $rRL^{\boxed{C}}$; see Fig. 1c) compares the observed pattern to simulations in which all point labels except for subpopulation *C* (the restriction set) are rearranged. This partially mimics the effect of *TS*, in which each subpopulation's internal structure is held constant, but unlike *TS* the fixed subpopulation is not allowed to shift

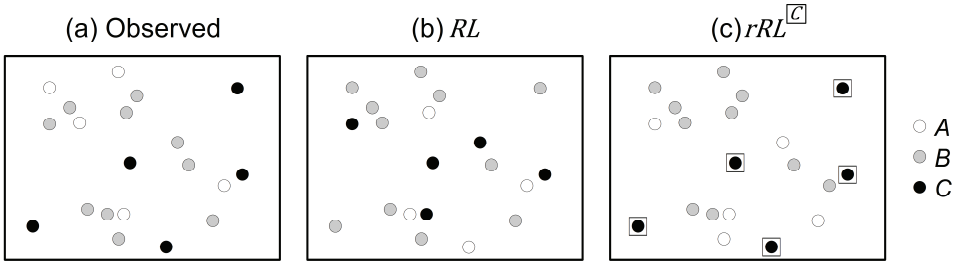

Fig. 1: Illustration of random labeling and restricted random labeling. Given an observed marked point pattern (a), random labeling (b) implies that all point labels are randomly assigned. In contrast, restricted random labeling (c) implies that labels are randomly assigned to all individuals except for those in the restriction set, enclosed in square boxes.

randomly. In effect, rRL answers the question, are the observed co-location patterns unusual *given the spatial pattern of the joint population and the restriction set?* Note that while this example uses a restriction set consisting of a single, entire subpopulation, it is possible to define the restriction set to include any subset of individuals.

3.1 Expected Neighbor Counts

To determine if the observed pattern of spatial association between two subpopulations A and B is higher or lower than expected, a common method is to compare the observed count of B within a defined neighborhood surrounding each A individual with the expected count under the appropriate null model. The following variables are defined similar to Leslie and Kronenfeld (2011):

N	the total number of individuals in all subpopulations,
n_A, n_B	the number of individuals in populations A and B, respectively,
n'_B	$n_B - 1$ if $A = B$; n_B otherwise,
$C_{A \to B}$	the total weighted count of B individuals within the defined neighborhood of all A individuals.

If the neighborhood is defined as the distance to the nearest neighbor and there are no equidistant nearest neighbors, then $C_{A \to B}$ is simply the number of A individuals whose nearest neighbor is B. Equidistant nearest neighbors can be handled by assigning fractional weights to each equidistant nearest neighbor (Leslie and Kronenfeld 2011). Cromley et al (2014) describe the use of weights to define other types of neighborhoods, including distance neighborhoods, topological neighborhoods derived from Thiessen polygons, and k-nearest neighbor neighborhoods. For simplification purposes, we assume a weighting scheme in which the weights of all neighbors of a given point sum to one. The term n'_B indicates the number of B individuals that can potentially be in the neighborhood of each A individual, and is one less than n_B when $A = B$ because an individual cannot be it's own neighbor.

	B restricted	B unrestricted
A restricted	$C_{\boxed{A}\to\boxed{B}}$	$\left(C_{\boxed{A}\to\breve{P}}\right)\frac{n_{\breve{B}}}{N_{\breve{P}}}$
A unrestricted	$\left(C_{\breve{P}\to\boxed{B}}\right)\frac{n_{\breve{A}}}{N_{\breve{P}}}$	$\left(n_{\breve{A}}\times\frac{C_{\breve{P}\to\breve{P}}}{N_{\breve{P}}}\right)\frac{n'_{\breve{B}}}{N_{\breve{P}}-1}$

Table 1: Expected counts of $A\to B$ nearest neighbor pairs for combinations of restricted and unrestricted portions of each subpopulation.

To distinguish between restricted and unrestricted subsets of the population and of individual subpopulations, additional notation is necessary. We use a box (e.g. $\boxed{A}$) to denote sets of restricted individuals and a breve (e.g. $\breve{A}$) to denote sets of unrestricted individuals in subpopulations (e.g. A and B) as well as the joint population (P), so that:

$\boxed{A}, \boxed{B}, \boxed{P}$ the sets of restricted individuals in *A, B* and *P*, respectively,

$\breve{A}, \breve{B}, \breve{P}$ the sets of unrestricted individuals in *A, B* and *P*, respectively.

This notation is used in subscripts as above. For example, $n_{\boxed{A}}$ denotes the number of restricted individuals in subpopulation A, and $C_{\breve{P}\to\boxed{B}}$ denotes the number of unrestricted individuals in the entire population P whose nearest neighbor is a restricted individual from subpopulation B. Note that any (sub)population can be partitioned into restricted and unrestricted subsets, so that $n_A = n_{\boxed{A}} + n_{\breve{A}}$ and $N_P = N_{\boxed{P}} + N_{\breve{P}}$.

Using this notation, the expected neighbor count from A to B can be derived from permutation theory as follows:

$$E(C_{A\to B}) = C_{\boxed{A}\to\boxed{B}} + \left(C_{\boxed{A}\to\breve{P}}\right)\frac{n_{\breve{B}}}{N_{\breve{P}}} + \left(C_{\breve{P}\to\boxed{B}}\right)\frac{n_{\breve{A}}}{N_{\breve{P}}} + \left(n_{\breve{A}}\times\frac{C_{\breve{P}\to\breve{P}}}{N_{\breve{P}}}\right)\frac{n'_{\breve{B}}}{N_{\breve{P}}-1}. \quad (2)$$

Each of the four terms in Eq. (2) represents the expected weighted count of $A\to B$ neighbor pairs in one of four non-overlapping subsets: the sets of restricted and unrestricted individuals of subpopulation A whose neighbors are restricted or unrestricted individuals of subpopulation B. Within each such subset, the randomization process is equivalent to a random sample, without replacement, of a set of labeled elements (Dixon 1994). The portions of Eq. (2) pertaining to each subset are shown in Table 1. A more detailed derivation and sample calculation are provided in the appendices.

3.2 Restriction Options

Equation (2) is generally applicable to any restriction set, and can be used to test null models tailored to specific research questions. To address the goal of ascertaining (sub)population independence in situations with three or more subpopulations, a natural approach is to restrict

one of the two subpopulations for each pairwise interaction. The first option, designated $rRL^{\boxed{row}}$ (equivalently $rRL^{\boxed{A}}_{A\to B}$), is to restrict the row subpopulation in each pairwise test. This is similar to the restricted randomization described in Wiegand et al. (2007), but without requiring a spatial distribution model to randomize locations. Specifically, $rRL^{\boxed{row}}$ tests the hypothesis that row-category individuals tend to locate near column-category individuals, given the spatial distribution of row-category individuals. The converse, $rRL^{\boxed{col}}$ (equivalently $rRL^{\boxed{B}}_{A\to B}$), restricts the column subpopulation and thus tests the same hypothesis given the spatial distribution of column-category individuals. These null models are only applicable when analyzing spatial interaction between non-identical subpopulations (i.e $row \neq column$).

The expected neighbor counts for $rRL^{\boxed{row}}$ and $rRL^{\boxed{col}}$ can be derived from Eq. (2). Under $rRL^{\boxed{row}}$, only the second term in Eq. (2) is applicable. Given that the restriction set is the first subpopulation, the expected value of the neighbor count $C_{A\to B}$ reduces to

$$E(C_{A\to B}) = (n_A - C_{A\to A})\frac{n_B}{N - n_A}. \tag{3}$$

Similarly, the expected value under $rRL^{\boxed{col}}$ is

$$E(C_{A\to B}) = (n_B - C_{B\to B})\frac{n_A}{N - n_B}. \tag{4}$$

When weights are equal to one (i.e. exactly one neighbor for each point), the distribution of neighbor counts is equivalent to a random sample without replacement of $n_A - C_{A\to A}$ elements from a set with $N - n_A$ elements, of which n_B are from subpopulation *B*. The distribution is hypergeometric with variance under $rRL^{\boxed{row}}$ equal to

$$Var(C_{A\to B}) = (n_A - C_{A\to A})\frac{n_B}{(N - n_A)}\frac{(N - n_A - n_B)}{(N - n_A)}\frac{(N - 2n_A + C_{A\to A})}{(N - n_A - 1)} \tag{5}$$

and variance under $rRL^{\boxed{col}}$ equal to

$$Var(C_{A\to B}) = (n_B - C_{B\to B})\frac{n_A}{(N - n_B)}\frac{(N - n_B - n_A)}{(N - n_B)}\frac{(N - 2n_B + C_{B\to B})}{(N - n_B - 1)}. \tag{6}$$

3.3 Co-Location Quotient

Leslie and Kronenfeld (2011) defined the asymmetrical co-location quotient (*CLQ*) between two subpopulations *A* and *B* under random labeling (*RL*) as

$$CLQ_{A\to B} = \frac{C_{A\to B}}{n_A\, n'_B/(N-1)}. \tag{7}$$

The denominator of Eq. (7) is equal to the expected neighbor count from *A* to *B* under random labeling (*RL*):

$$E(C_{A\to B}) = n_A n'_B/(N-1)\,. \tag{8}$$

Therefore, under *RL* the *CLQ* can be re-written as

$$CLQ_{A\to B} = \frac{C_{A\to B}}{E(C_{A\to B})}. \tag{9}$$

Equation (9) shows that the *CLQ* is naturally interpreted as the ratio between observed and expected neighbor counts, where the expected count is determined under a null model of *RL*.

If a null model other than *RL* is used, the expected neighbor count will differ from Eq. (8), and therefore the *CLQ* as defined by Leslie and Kronenfeld (2011) will not be appropriate. We define the generalized co-location quotient $CLQ^{\boxed{S}}_{A\to B}$ as the ratio between the observed and expected neighbor counts under the null model of $rRL^{\boxed{S}}$. Generalized equations for computing *CLQ*s under $rRL^{\boxed{row}}$ and $rRL^{\boxed{col}}$ are given by:

$$CLQ^{\boxed{A}}_{A\to B} = \frac{C_{A\to B}}{(n_A - C_{A\to A})\frac{n_B}{N-n_A}}, \tag{10}$$

$$CLQ^{\boxed{B}}_{A\to B} = \frac{C_{A\to B}}{(n_B - C_{B\to B})\frac{n_A}{N-n_B}}. \tag{11}$$

The generalized *CLQ* under any null model has an expected value of one. Values greater than one indicate a higher degree of spatial interaction than expected by chance, while values less than one indicate the opposite. In particular, $CLQ^{\boxed{S}}_{A\to B} = \beta$ indicates that β times as many individuals of subpopulation *B* are found within the defined neighborhood of subpopulation *A* as would be expected given the constraint set *S*.

3.4 Significance Testing

Two approaches to significance testing are possible. An analytic approach uses parametric formulas that represent the sampling distribution either approximately or exactly to determine the likelihood of the observed metric under the given null model. Alternatively, a simulation

approach entails generating a large number of potential realizations of the null model to estimate the distribution of the test statistic.

Analytic formulas for sampling distributions of pairwise nearest-neighbor counts under *RL* were given by Dixon (2002), who showed that these are approximately normal when the expected nearest neighbor count is larger than ten. Ceyhan (2008) suggested that Monte Carlo simulation is only necessary when the expected nearest neighbor count is four or less. We are not aware of research indicating how large the expected neighbor count needs to be for the normal approximation to hold in the case of multiple neighbors with varying weights. In any case, Monte Carlo simulation under *rRL* is computationally efficient and can be used in any scenario to estimate statistical significance. To create realizations of the *rRL* null model, marks of unrestricted individuals are randomly shuffled in the same manner as *RL*, and individuals in the restriction set are kept in their original location.

4 Case Studies

To illustrate restricted random labeling and to ascertain its ability to disentangle joint-population, within-group and between-group effects, three types of simulated point patterns and two empirical dataset were analyzed. For comparison purposes, co-location quotients and significance values were calculated under null models of *TS*, *RL*, $rRL^{\boxed{row}}$ and $rRL^{\boxed{col}}$ for all examples. Statistical significance was assessed by Monte Carlo simulation of the corresponding null model, with two-tailed *p*-values calculated as two times the proportion of simulations for which the test statistic was more extreme than the observed statistic. This method has been found to be more conservative than inverting a single two-sided test, especially in situations with small sample sizes (Agresti and Min 2001). For *TS*, we follow Goreaud and Pélissier (2003), and shift all subpopulations except for the first in each pairwise analysis. Since the expected neighbor count cannot be derived analytically for *TS*, we used the simulated mean value.

Although Manly (1991) suggests at least 1,000 simulations for significance testing at α=0.05, we found this impracticable due to the computation time of *TS*. Because *TS* results in a new set of point locations, neighbor relations must be recalculated for each MC simulation, which is computationally expensive especially for large datasets. An advantage of *RL* and *rRL* is that point locations do not change from the original pattern, eliminating the need to recalculate neighbor relations. Therefore, we performed only 100 Monte Carlo simulations for all analyses. This may have increased uncertainty regarding statistical significance in some cases, but our purpose was not to assess statistical significance but rather to compare null models. Variation in results was assessed for a representative sample of simulated patterns and was found to be at least an order of magnitude less than the observed differences between null models.

All analyses were performed using custom software developed in Visual Studio.Net (Microsoft Corp.), available upon request from the authors.

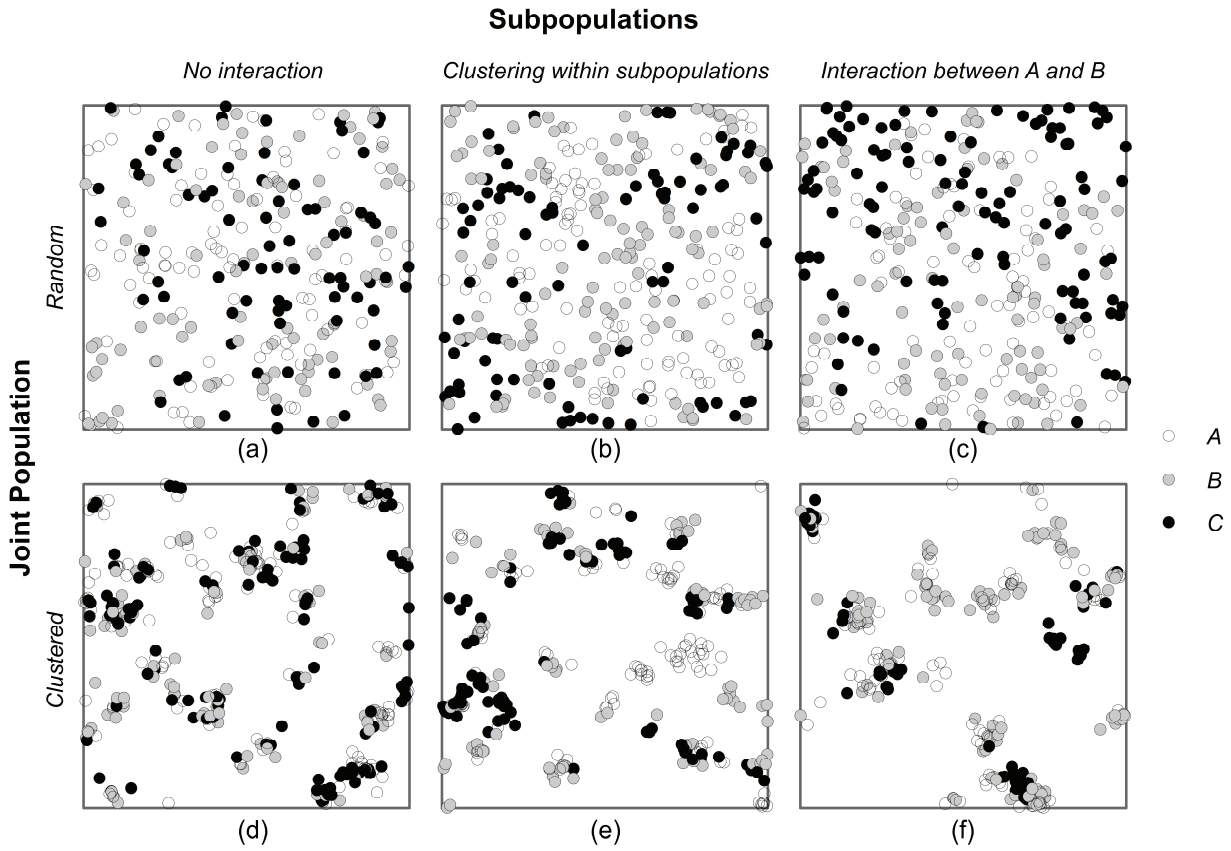

Fig. 2: Examples of simulated patterns used to compare null models of spatial interaction.

4.1 Simulated Patterns

To illustrate different patterns of aggregation in the joint population, aggregation of individual subpopulations, and spatial interaction between subpopulations, six generating processes were defined in a 3-by-2 matrix. An example pattern from each process is shown in Fig. 2. Each process consisted of three subpopulations, and was initialized by construction of one of two types of joint population patterns (random or clustered), followed by a marking process defined to create a pattern with no spatial interaction, clustering of individual subpopulations, or interaction between subpopulations. For each process, 100 pattern instances were generated on a one-unit-square region.

For the joint population, random patterns were generated using a Poisson process with density 300. Clustered patterns were generated with a Thomas process (Illian et al. 2008) with 30 parent clusters per square unit and ten daughters per parent distributed in a Gaussian kernel with σ=0.0325.

To simulate absence of interaction, each point was randomly assigned one of three subpopulations {*A*, *B*, or *C*}. To simulate clustering of individual subpopulations, each point was then selected in sequence and replaced with an individual recruited from the subpopulation of

the nearest neighbor. This was repeated 1,000 times. The effect of this procedure was to create new members of each subpopulation near to the locations of existing members. The nearest neighbor in this process can be thought of as a reproductive agent that reseeds nearby locations with self-similar individuals when they become vacant. Note that, beyond favoring neighbors from the same subpopulation, the process is random with respect to interaction between non-identical subpopulations. Finally, to simulate interaction between subpopulations, the process was defined in a similar manner as above, except that 50% of the time individuals with nearest neighbors from subpopulations *A* were replaced with an individual from subpopulation *B* and vice versa. This introduced an explicit interaction component, simulating positive interaction between *A* and *B* as well as within-group aggregation of all subpopulations.

For analysis purposes, nearest-neighbor analysis was used (i.e. the neighborhood around each point was defined as the distance to the nearest neighbor). It should be noted that the above generating processes conform to an *a posteriori* model according to the classification of Goreaud and Pélissier (2003). It would have been preferable to also include an *a priori* marked generating process designed to create both the desired spatial interactions and a consistent degree of clustering in the joint population. Although in theory this could be achieved using pattern reconstruction methods, it would require a metric of spatial interaction that could be interpreted independently of any null model (since we apply different null models to each pattern and examine the results). Our simulation procedures are designed such that the presence or absence of spatial interaction can be interpreted in context from the procedure itself.

4.2 Savannah River Trees

Next, we evaluate an empirical dataset originally compiled by Good and Whipple (1982), consisting of the locations and species of 734 trees in a one-hectare (50×200m) forest plot along the Savannah River in South Carolina, USA (see Fig. 3). The joint population of this dataset is random at all scales. Tree species are usually considered an *a posteriori* marking and thus *PI* is recommended as a null model (Goreaud and Pélissier 2003), although *RL* can be justified by the recognition of a shared environmental factor (Wiegand and Moloney 2014). Dixon (2002) previously used *RL* to examine spatial interaction of species in this dataset, and found three significant types of patterns: (a) positive within-species interaction for all species except

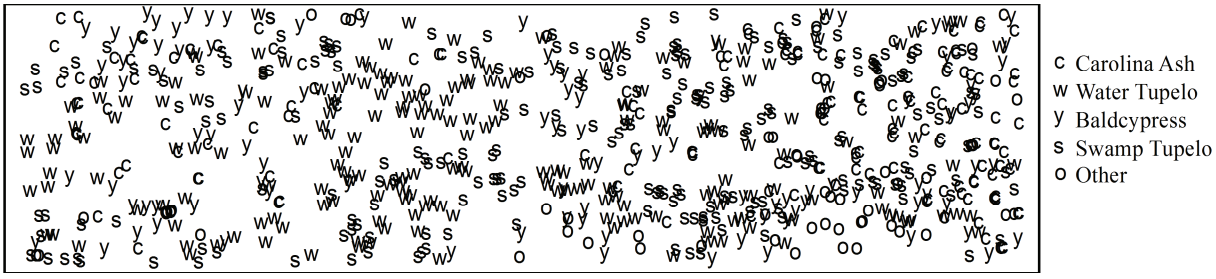

Fig. 3: Locations of tree species in the Savannah River forest plot. Data from Dixon (2002).

Fig. 4: Locations of murders (large red dots) and other documented crimes (gray dots) in Spokane, WA, USA from 2008 to 2012.

baldcypress, (b) negative interaction of most non-identical species pairs, and (c) positive interaction between baldcypress and Carolina ash. We followed Dixon in using nearest neighbor analysis, and compared our *NNCT* counts and *RL* results with his to ensure computational agreement.

4.3 Crimes in Spokane

The last dataset consists of crimes committed between 2008 and 2012 in Spokane, WA, USA (see Fig. 4). Of interest is the possible spatial interaction between murders (n=73) and nine other crime categories. To make the data more compact, 78 crimes in outlying areas were removed, and the study area was defined as the minimum bounding rectangle around the remaining 124,951 crimes. The joint population is highly clustered at local scales and contains overlap, with as many as 973 crimes recorded for the same location. 23.2% of points were coincident with 25 or more other points. Crime density also varies from neighborhood to neighborhood, but generally increases with population density toward the center of the city. In addition, crimes exhibit a pattern clearly related to, but not constrained by the road network. It is not obvious whether crimes should be considered *a priori* or *a posteriori* marked data, since some crimes are premeditated while others are opportunistic in nature.

		No Interaction			Clustering			Interaction		
		A	***B***	***C***	***A***	***B***	***C***	***A***	***B***	***C***
TS	***A***	1.01	1.01	0.98	**1.31**	**0.62**	**0.63**	0.98	**1.44**	**0.50**
	B	0.99	1.02	1.00	**0.63**	**1.29**	**0.65**	**1.44**	0.98	**0.52**
	C	0.99	1.00	1.01	**0.61**	**0.65**	**1.30**	**0.54**	**0.57**	**1.41**
RL	***A***	1.02	1.00	0.98	**2.22**	**0.42**	**0.43**	**1.30**	**1.34**	**0.38**
	B	0.98	1.01	1.01	**0.42**	**2.19**	**0.44**	**1.34**	**1.30**	**0.39**
	C	0.99	1.00	1.01	**0.41**	**0.44**	**2.23**	**0.38**	**0.40**	**2.46**
$rRL^{\boxed{row}}$	***A***	X	1.01	0.99	X	1.00	1.01	X	**1.58**	**0.44**
	B	0.99	X	1.01	1.00	X	1.01	**1.58**	X	**0.45**
	C	1.00	1.00	X	0.97	1.04	X	0.97	1.03	X
$rRL^{\boxed{col}}$	***A***	X	1.00	0.99	X	1.00	1.01	X	**1.56**	0.98
	B	0.99	X	1.01	1.03	X	1.01	**1.57**	X	1.02
	C	1.01	1.00	X	0.99	1.01	X	**0.44**	**0.46**	X

Notes: X indicates $rRL^{\boxed{row}}$ and $rRL^{\boxed{col}}$ not applicable when row = column

Table 2: Average pairwise co-location quotients for three simulated types of interaction within a random joint population pattern under each of four null models. Greyscale indicates number of trials out of 100 in which observed CLQ value was significant: 0 to 5, 6 to 10, 11 to 20, **21 to 50, 51 to 100.**

We assessed spatial interaction between murders (*M*) and each other type of crime (*X*) by calculating pairwise co-location quotients $CLQ_{M \to X}$ under *TS*, *RL* and $rRL^{\boxed{X}}$. Due to the low number and lack of clustering of murders, we did not calculate $rRL^{\boxed{M}}$. Given the large number of coincident points, simple nearest-neighbor analysis would have included anywhere between one and 973 nearest neighbors; to reduce this variability, analysis was performed on a 25-nearest-neighbor neighborhood.

5 Results

5.1 Simulated Patterns

Average pairwise *CLQ* values for the three simulated pattern types derived from a random joint population are shown in Table 2 for each null model. Trials in which significant interaction was found more frequently are indicated by darker font. When subpopulations were distributed randomly (first column), all null models behaved as expected: *CLQ* values were close to unity, and these values were found to be statistically significant (p<0.05) in approx. 5% of trials. When clustering within subpopulations was simulated (second column), both *TS* and *RL* detected positive aggregation within subpopulations but also negative interaction between subpopulations, with more extreme values under *RL*. In contrast, $rRL^{\boxed{row}}$ and $rRL^{\boxed{col}}$ both resulted in *CLQ* values near unity, suggesting that interaction patterns were no different than would be expected given the observed aggregation of individual subpopulations.

When positive interaction between *A* and *B* were specifically simulated (third column), this

		No Interaction			Clustering			Interaction		
		A	***B***	***C***	***A***	***B***	***C***	***A***	***B***	***C***
TS	***A***	**0.42**	**2.99**	**2.97**	**0.82**	**1.97**	**2.02**	**0.54**	**5.08**	1.34
	B	**2.94**	**0.44**	**2.91**	**2.04**	**0.82**	**1.98**	**4.90**	**0.54**	1.32
	C	**2.95**	**2.96**	**0.43**	**1.97**	**2.07**	**0.83**	1.40	1.50	**0.92**
RL	***A***	0.98	1.00	1.03	**2.12**	**0.44**	**0.45**	**1.34**	**1.34**	**0.31**
	B	1.00	1.02	0.99	**0.46**	**2.17**	**0.44**	**1.34**	**1.34**	**0.30**
	C	1.02	0.99	0.99	**0.44**	**0.45**	**2.19**	**0.29**	**0.30**	**2.69**
$rRL^{\boxed{row}}$	***A***	X	0.99	1.02	X	1.00	1.01	X	**1.62**	**0.36**
	B	1.01	X	0.99	1.03	X	0.99	**1.63**	X	**0.35**
	C	1.01	0.99	X	0.98	1.02	X	0.97	1.02	X
$rRL^{\boxed{col}}$	***A***	X	1.00	1.02	X	1.00	1.01	X	**1.63**	1.01
	B	0.99	X	0.98	1.04	X	1.00	**1.63**	X	0.99
	C	1.01	1.00	X	0.96	1.02	X	**0.35**	**0.35**	X

Notes: X indicates $rRL^{\boxed{row}}$ and $rRL^{\boxed{col}}$ not applicable when row = column

Table 3: Average pairwise co-location quotients for three simulated types of interaction within a clustered joint population pattern under each of four null models. Greyscale indicates number of trials out of 100 in which observed *CLQ* value was significant: 0 to 5, 6 to 10, 11 to 20, **21 to 50, 51 to 100.**

interaction was detected by all null models in more than 50% of trials (*TS*=72.5%, *RL*=57%, $rRL^{\boxed{row}}$=95.5%, $rRL^{\boxed{col}}$=96%). The strength of interaction was slightly higher under both forms of *rRL* ($CLQ = 1.56$ to 1.58) than either *TS* ($CLQ = 1.44$) or *RL* ($CLQ = 1.34$). Negative spatial interaction was detected in all pairs involving subpopulation *C* under both *TS* and *RL*. Under *rRL*, negative interaction with subpopulation *C* was found when subpopulation *C* was not restricted, but was not found when subpopulation *C* was restricted.

Results for clustered joint population patterns are shown in Table 3. The results are nearly identical to Table 2 for *RL*, $rRL^{\boxed{row}}$ and $rRL^{\boxed{col}}$. However, under *TS* every pairwise analysis results in substantial negative interaction within subpopulations and positive interaction between subpopulations, with many interactions significant in more than 50% of trials (darkest font). This is due to the fact that Monte Carlo simulations under the *TS* null model effectively dispersed the subpopulations, removing the original aggregation pattern of the joint population (see Fig. 5). Thus, distances between individuals from different subpopulations are increased in the Monte Carlo simulation in comparison to the original pattern, while distances between members of the same subpopulation remain constant by definition. Positive interaction between *A* and *B* was again detected by all null models in more than 50% of trials (*TS*=100%, *RL*=58%, $rRL^{\boxed{row}}$=93.5%, $rRL^{\boxed{col}}$=95.5%).

5.2 Savannah River Trees

The *NNCT* for the Savannah River plot matched Dixon (2002) exactly except for one swamp tupelo tree for which we found two equidistant nearest neighbors. Our computed statistical significance also matched Dixon closely, with slight differences attributable to random variability in the Monte Carlo simulation process.

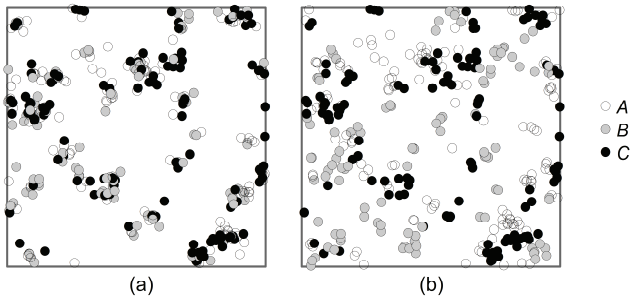

Fig. 5: Effects of toroidal shift on joint population pattern. Original pattern (a) is the same as lower left pattern (d) in Fig. 2. Toroidal shift of subpopulations *A* and *B* (b) results in a joint population that is less aggregated than the original.

CLQ values under each model are shown in Table 4, with darker fonts indicating greater statistical significance. Within-group aggregation was statistically significant for two of five species under *TS* and four of five species under *RL*, with *CLQ* values ranging from 1.3 to 7.0. Under *TS*, positive interaction was found in both directions between *Taxodium distichum* (*TD*) and *Fraxinus caroliniana* (*FX*). Under both $rRL^{\boxed{row}}$ and $rRL^{\boxed{col}}$, this relationship was significant when *FX* was restricted ($p<0.05$) but only marginally significant when *TD* was restricted ($p<0.1$). Since *TD* was less clustered than *FX*, the difference between *RL* and *rRL* was less pronounced when *TD* was restricted than when *FX* was restricted.

In *RL*, the relationship was only marginally significant in one direction ($p<0.1$), and insignificant in the other. In contrast with other null models, negative spatial interaction was found for most other subpopulation pairs, and was statistically significant in half of all pairwise tests ($p<0.05$). This apparent negative interaction can largely be attributed to within-group aggregation, which is accounted for in the other null models but not *RL*.

	TS					*RL*					$rRL^{\boxed{row}}$					$rRL^{\boxed{col}}$				
	FC	NS	NA	TD	OT	FC	NS	NA	TD	OT	FC	NS	NA	TD	OT	FC	NS	NA	TD	OT
FC	1.0	0.9	0.9	1.6	0.9	**2.5**	**0.5**	**0.5**	1.0	0.5	X	0.9	0.8	**1.8**	0.8	X	0.8	0.9	1.4	0.7
NS	0.8	**1.3**	0.8	0.6	0.6	**0.6**	**2.0**	**0.6**	**0.6**	**0.5**	1.0	X	1.1	1.0	0.8	0.8	X	1.1	0.8	0.8
NA	1.0	0.9	1.1	0.9	1.2	**0.6**	**0.7**	**1.8**	0.7	0.8	0.9	1.0	X	1.0	1.2	0.9	1.0	X	0.9	1.3
TD	**1.6**	1.1	0.7	0.7	1.0	1.4	1.1	**0.7**	1.1	0.9	1.4	1.1	**0.7**	X	0.9	**1.9**	**1.6**	1.1	X	1.3
OT	0.7	0.7	0.7	1.3	**1.3**	**0.4**	**0.5**	**0.4**	0.9	**7.0**	0.8	1.0	0.8	1.8	X	0.5	0.7	0.7	1.1	X

Notes: X indicates $rRL^{\boxed{row}}$ and $rRL^{\boxed{col}}$ not applicable when row = column

Table 4: Pairwise co-location quotients for Savannah River dataset under four different null models. FC: *Fraxinus caroliniana* (Carolina ash); NS: *Nyssa sylvatica* (black gum); NA: *Nyssa aquatica* (swamp tupelo); TD: *Taxodium disticum* (baldcypress); OT: all other species. Greyscale indicates *p*-value: $p>0.1$, $0.05<p\leq0.1$, **$0.01<p\leq0.05$**, **$p\leq0.01$**.

For this dataset, *TS* and *rRL* yielded similar results. This can be attributed to the fact that, although *rRL* (but not *TS*) holds the pattern of the joint population constant, the joint population was random and thus generally modeled well by *TS*.

5.3 Crimes in Spokane

Pairwise *CLQ* values between murder and other types of crime are shown in Table 5. Under *TS*, no relationships are significant or marginally significant. This is due to high variability in the nearest neighbor counts under *TS* simulations, which are not constrained to conform to the joint population pattern of city streets and higher crime density downtown. Thus, for example, although the number of murder-assault neighbor pairs in the observed data is 1.78 times higher than the average *TS* simulation, 18% of simulations result in an even higher number of assaults among the 25 nearest crimes around each murder.

Second Subpopulation	*TS*	*RL*	$rRL^{\boxed{col}}$
Arson	0.43	1.26	1.24
Assault	1.78	**1.66**	**1.77**
Burglary	0.74	**0.89**	0.96
Drugs	0.81	**1.31**	**1.39**
Malicious Mischief	0.98	0.96	0.99
Robbery	0.94	**1.44**	**1.46**
Theft	1.23	**0.86**	1.04
Vehicle Prowling	0.69	**0.76**	**0.84**
Vehicle Theft	0.89	**0.81**	0.85

Table 5: Pairwise co-location quotients for murder in the Spokane crime dataset under three different null models. Greyscale indicates p-value: $p>0.1$, $0.05<p\leq0.1$, $0.01<p\leq0.05$, $p\leq0.01$.

Results of *RL* and *rRL* were similar but with some notable differences. Positive associations with assault, drugs and robbery were stronger but less significant under *rRL* in comparison to *RL*, suggesting higher variability of *rRL* simulations. Negative associations significant under *RL* became only marginally significant (vehicle prowling) or disappeared entirely (theft, vehicle theft) under *rRL*. This again suggests that much of the negative spatial interaction detected by *RL* can be explained by aggregation of individual crime types.

In general across all pairs of subpopulatons, *TS* resulted in strongly negative aggregation within subpopulations, with *CLQ* values averaging 0.20 times that of *RL* (data not shown). In contrast, *CLQ* values between non-identical pairs of categories averaged 1.53 times that of *RL* (data not shown). In other words, *TS* suggested that all individual types of crimes were dispersed but most non-identical pairs of crime types had positive spatial association. This runs counter to intuition, but can be explained by the fact that *TS* simulations disperse points across the study region, removing the clustering of the original joint population. This leads to fewer non-identical neighbors in simulations, making the observed non-identical neighbor counts appear large in

comparison. In contrast, *RL* signifies that each crime type is clustered within the already clustered pattern of the joint population.

5.4 Discussion

Null models of population independence and random labeling have provided a foundation for spatial interaction analysis for several decades, but choosing the appropriate null model is not always straightforward. In recent years, the consensus has been to use population independence for *a priori* marking and random labeling for *a posteriori* marking (Goreaud and Pélissier 2003). However, several limitations to this framework exist. Although the distinction is useful, it is not always unambiguous and involves some degree of subjectivity. Implementing population independence using toroidal shift (*TS*) is not straightforward for non-rectangular study regions, and *TS* is computationally intensive for large datasets due to the need to recompute neighborhood relations in each simulation. We also note that toroidal shift produces joint population patterns that differ from the observed pattern in a consistent way, especially when the joint population is clustered. This can lead to unintuitive results.

An alternative conceptual approach to choosing a null model is to randomize those elements of the problem that pertain to the research question while holding unrelated elements constant (Manly 1991, Fuller and Enquist 2012). The functional procedures of *RL* and *TS* provide an obvious foundation for this approach, as they hold constant the patterns of the joint population and of individual subpopulations, respectively. This mirrors the distinction made by some authors in the economic literature between joint-localization and co-localization (see e.g., Duranton and Overman 2005).

Restricted random labeling is a natural extension of *TS* and *RL*, as it seeks to hold constant both the patterns of the joint population and of individual subpopulations simultaneously. It is appropriate in situations where clustering of both the joint population and individual subpopulations is expected *a priori*, and the researcher seeks to understand spatial interactions that are not logically implied by either pattern. Such situations are not uncommon. For example, housing types, businesses and crime are strongly related to the artificial landscape, while different types of housing, businesses and crime all tend to cluster in space. In biological settings, individual subpopulations will obviously cluster in space due to shared biological traits, but there may also be environmental characteristics that affect all subpopulations simultaneously, such as water bodies, variations in soil fertility, etc.

Restricted random labeling is only applicable to point patterns with three or more types of marks (subpopulations), as restriction of one of two subpopulations eliminates all possibility of randomization. In addition, the *rRL* statistic is designed to reveal between-group association or disassociation, but cannot be used to assess within-group aggregation.

In our analyses, *rRL* generally provided a more stringent test of negative pairwise interaction than *RL*, as the latter "discovered" negative interactions that could be interpreted as the natural consequence of aggregation within subpopulations. As a corollary, positive pairwise interaction was identified more often by *rRL* than by *RL*. These differences had a pronounced impact on

interpretation of the Savannah River dataset: whereas *RL* indicates strong aversion of most species pairs, the lack of any such indication under *rRL* suggests that this may simply be the result of aggregation tendencies within species.

rRL was more similar to *TS* than to *RL* when the joint population was randomly distributed (see Table 2 and Table 4). However, when the joint population was clustered, *TS* results were highly variable. These results may be influenced somewhat by the choice of metric (i.e. fixed distance neighborhood vs. number of nearest neighbors), but are ultimately due to the fact that the joint population of simulations under *TS* can have a very different structure than the observed joint population. For example, if the joint population is strongly aggregated, simulations under *TS* will tend to be less aggregated, causing a systematic reduction in simulated neighbor counts and thus resulting in inference of positive spatial interaction. Since such results primarily reflect the clustering of the joint population, employing *rRL* to constrain the joint population pattern will provide a more logical interpretation in many cases.

The present study did not assess other, more sophisticated methods for implementing *PI* such as the use of parametric point processes and non-parametric pattern reconstruction. In theory, pattern reconstruction methods could be designed to simultaneously account for the spatial structure of the individual subpopulations and joint population, thus eliminating the problem of heterogeneity in the joint population pattern observed with *TS*. Pattern reconstruction methods allow locations to vary while preserving the general characteristics of the point pattern, offering analytical flexibility (Wiegand and Moloney 2014). This advantage comes at the expense of greater required expertise and computation time in comparison to *rRL*.

One weakness of *rRL* is that only one subpopulation can be restricted at a time. In simulations, the influence of within-group aggregation was effectively filtered in either $rRL^{\boxed{row}}$ or $rRL^{\boxed{col}}$, but not in both. Greater differences from *RL* are obtained if the subpopulation with the strongest pattern of clustering is restricted. A conservative approach is to consider a relationship statistically significant only if it is significant under both $rRL^{\boxed{row}}$ and $rRL^{\boxed{col}}$.

The examples in this paper employ topological metrics of spatial interaction. Topological metrics are less popular than distance-based metrics, due to the fact that the latter uses a constant neighborhood size. However, nearest neighbor metrics give equal weight to each individual in the base (row) population. In any case, the choice of null model is separate from that of neighborhood definition, and *rRL* could be used to evaluate to distance-based metrics of spatial interaction.

Our presentation of spatial interaction tables follows previous work using nearest neighbor metrics (e.g. Dixon 2002), but differs from the standard presentation of distance-based metrics of spatial interaction over a range of scales in a single graph (e.g. Liu et al. 2007). Such graphs implicitly place the focus on a single pairwise relation, while tabular results emphasize the complete set of pairwise relations but restrict focus to a single scale of analysis.

While *rRL* was able to detect simulated spatial interaction in more than 90% of trials, it may not be the most powerful method of analysis if *a priori* knowledge of certain aspects of the population and subpopulations under investigation, such as first-order spatial trends, allow for randomization of locations as well as labels. Designation of a restriction set effectively reduces sample size and thus decreases statistical power in comparison to *RL*, although not in proportion to the percentage of individuals restricted. Large sample sizes may be crucial to effective analysis and null model comparison, since small sample sizes may have large standard deviations that prevent meaningful comparisons with null models during spatial analysis (Perry et al. 2006), and previous literature has shown that the significance of small departures from the null model should be interpreted with caution (Blanco et al. 2008).

6 Conclusions and suggestions for future research

Restricted random labeling (*rRL*) of each population in pairwise analysis provides a useful mechanism for discerning patterns of spatial interaction when multiple (n>2) populations are constrained by a set of common environmental factors. Our simulation experiment found *rRL* to be unbiased in the presence of non-random structure in either the joint population or individual subpopulations, or both. In addition, *rRL* showed greater statistical power than *TS* and *RL* when these were also unbiased, and is computationally faster than *TS,* which is advantageous for large datasets. A drawback of *rRL* is that only one population can be constrained at a time. Therefore, we recommend that both $rRL^{\boxed{row}}$ and $rRL^{\boxed{col}}$ be assessed unless a reasonable justification exists for disregarding the effect of one population's structure (as in our analysis of Spokane crime data).

One suggestion for future research is to characterize the degree of correspondence between *rRL* and pattern reconstruction methods (Wiegand and Moloney 2014). Although more complex, these can also be used to constrain patterns of both the joint population and individual subpopulations. In particular, the sensitivity of spatial interaction analysis to small perturbations in event locations should be explored.

7 References:

Agresti A, Min Y (2001) On small-sample confidence intervals for parameters in discrete distributions. Biometrics 57(3):963-971

Blanco PD, Rostagno CM, del Valle HF, Beeskow AM, Wiegand T (2008) Grazing impacts in vegetated dune fields: predictions from spatial pattern analysis. Rangeland Ecology and Management 61(2):194-203

Ceyhan E (2009) Class-specific tests of spatial segregation based on nearest neighbor contingency tables. Statistica Neerlandica 63(2):149-182

Ceyhan E (2008) On the use of nearest neighbor contingency tables for testing spatial segregation. Environmental Ecological Statistics 17(3):247-282

Cliff AD, Ord JK (1981) Spatial processes: models and applications. Taylor and Francis, Abingdon
Cromley RG, Hanink DM, Bentley GC (2014) Geographically weighted colocation quotients: specification and application. The Professsional Geographer 66(1):138-148
Cuzick J, Edwards R (1990) Spatial clustering for inhomogeneous populations. Journal of the Royal Statistical Society, Series B 52(1):73-104
De la Cruz M, Romao RL, Escudero A, Maestre FT (2008) Where do seedlings go? A spatio-temporal analysis of seedling mortality in a semi-arid gypsophyte. Ecography 31(6): 1-11
Dixon PM (1994) Testing spatial segregation using a nearest-neighbor contingency table. Ecology 75(7):1940-1948
Dixon PM (2002) Nearest-neighbor contingency table analysis of spatial segregation for several species. Ecoscience 9(2):142-151
Duranton G, Overman HG (2005) Testing for localization using micro-geographic data. Review of Economic Studies 72(4), 1077–1106.
Florence PS (1944) The Selection of Industries Suitable for Dispersion into Rural Areas. Journal of the Royal Statistical Society 107(2):93-116
Fuller MM, Enquist BJ (2012) Accounting for spatial autocorrelation in null models of tree species association. Ecography 35(6):510-518
Getzin S, Wiegand T, Wiegand K, He F (2008) Heterogeneity influences spatial patterns and demographics in forest stands. Journal of Ecology 96(4):807-820
Good BJ, Whipple SG (1982) Tree spatial patterns: South Carolina bottomland and swamp forests. Bulletin of the Torrey Botanical Club 109(4):529-536
Goreaud F, Pélissier R (2003) Avoiding misinterpretation of biotic interactions with the intertype K12-function: population independence vs. random labelling hypotheses. Journal of Vegetation Science 14(5):681-692
Huang Y, Pei J, Xiong H (2006) Mining co-location patterns with rare events from spatial data sets. Geoinformatica 10(3):239-260
Illian J, Penttinen A, Stoyan H and Stoyan D (2008) Statistical Analysis and Modelling of Spatial Point Patterns. Wiley, Hoboken NJ
John R, Dalling JW, Harms KE, Yavitt JP, Stallard RF, Mirabello M, Hubbel SP, Valencia R, Navarrete H, Vallejo M, Foster RB (2007) Soil nutrients influence spatial distributions of tropical tree species. Proceedings of the National Academy of Sciences 104(3):864-869
Law R, Illian J, Burslem DFRP, Gratzer G, Gunatilleke CVS, Gunatilleke IAUN (2009) Ecological information from spatial patterns of plants: insights from point process theory. Journal of Ecology 97(4):616-628
Leibovici DG, Bastin L, Anand S, Hobona G, Jackson M (2011) Spatially clustered associations in health related geospatial data. Transactions in GIS 15(3):347-364
Leslie TF, Kronenfeld BJ (2011) The colocation quotient: a new measure of spatial association between categorical subsets of points. Geographical Analysis 43(3):306-326
Leslie TF, Ó hUallacháin B (2006) Polycentric Phoenix. Economic Geography 82(2):167-192
Liu D, Kelly M, Gong P, Guo Q (2007) Characterizing spatial-temporal tree mortality patterns associated with a new forest disease. Forest Ecology and Management 253(1-3):220-231
Loosmore NB, Ford ED (2006) Statistical inference using the G or K point pattern spatial statistics. Ecology, 87(8):1925-1931

Lotwick HW, Silverman BW (1982) Methods for analysing spatial processes of several types of points. Journal of the Royal Statistical Society, Series B (Methodological) 44(3):406-413
Manly BFJ (1991) Randomization, bootstrap, and Monte Carlo methods in biology. Chapman and Hall, London
Ó hUallacháin B, Leslie TF (2013) Spatial Pattern and Order in Sunbelt Retailing: Shopping in Phoenix in the Twenty-First Century. The Professional Geographer 65(3): 396.420
Pielou EC (1961) Segregation and symmetry in two-species populations as studied by nearest-neighbor relationships. Journal of Ecology 49(2):225-269
Perry GLW, Miller BP, Enright NJ (2006) A comparison of methods for the statistical analysis of spatial point patterns in plant ecology. Plant Ecology 187(1):59-82
Souris M, Bichaud L (2011) Statistical methods vor bivariate spatial analysis in marked points. Examples in spatial epidemiology. Spatial and Spatio-temporal Epidemiology 2(4):227-234
Stoica RS (2010) Marked point processes for statistical and morphological analysis of astronomical data. European Physical Journal Special Topics 186(1):123-165
Wiegand T, Gunatilleke S, Gunatilleke N (2007) Species associations in a heterogeneous Sri Lankan dipterocarp forest. American Naturalist 170(4):E77-E95
Wiegand T, Moloney KA (2004) Rings, circles, and null-models for point pattern analysis in ecology. Oikos 104(2):209-229
Wiegand T, Moloney KA (2014) A handbook of spatial point pattern analysis in ecology. Chapman and Hall/CRC press, Boca Raton FL

Appendix A: Detailed Derivation of Expected Nearest Neighbor Counts [Eq. (2)]

The first term in Eq. (2), $C_{\boxed{A}\to\boxed{B}}$, is the number of restricted *A* individuals that have restricted *B* individuals as their nearest neighbor. This is defined by the restriction pattern and remains fixed under randomization. The remaining three terms are derived from counting arguments. The second term derives the expected count $E\left(C_{\boxed{A}\to\check{B}}\right)$ of restricted-*A*-unrestricted-*B* nearest neighbor pairs by taking the number of restricted *A* individuals that have an unrestricted individual of any type as their nearest neighbor $\left(C_{\boxed{A}\to\check{P}}\right)$, and multiplying by the proportion of *B* among all unrestricted individuals $\left(\frac{n_{\check{B}}}{n_{\check{P}}}\right)$. Similarly, $\left(C_{\check{P}\to\boxed{B}}\right)\frac{n_{\check{A}}}{N_{\check{P}}}$ computes the expected count $E\left(C_{\check{A}\to\boxed{B}}\right)$ of unrestricted-*A*-restricted-*B* nearest neighbor pairs by taking the number of restricted *B* individuals that are the nearest neighbor of an unrestricted individual from any subpopulation $(C_{\check{P}\to\boxed{B}})$ and multiplying by the proportion of *A* among all unrestricted individuals $\left(\frac{n_{\check{A}}}{N_{\check{P}}}\right)$. Finally, the last component, $C_{\check{P}\to\check{P}}\frac{N_{\check{A}}}{N_{\check{P}}}\frac{N\prime_{\check{B}}}{(N_{\check{P}}-1)}$, is the expected count $E(C_{\check{A}\to\check{B}})$ of unrestricted-*A*-unrestricted-*B* nearest neighbor pairs, and is computed by taking the number of unrestricted *A* individuals $(n_{\check{A}})$, multiplying by the proportion of unrestricted individuals that have another unrestricted individual as their nearest neighbor $\left(\frac{C_{\check{P}\to\check{P}}}{N_{\check{P}}}\right)$ to determine the expected count $E(C_{\check{A}\to\check{P}})$ of unrestricted *A* individuals that will have another unrestricted individual from any subpopulation as a nearest neighbor, and then multiplying again by the proportion $\left(\frac{n\prime_{\check{B}}}{N_{\check{P}}-1}\right)$ of *B* among all unrestricted individuals, excluding the original *A* individual.

Appendix B: Applied Example

In this example, we investigate a situation where we are interested in the co-location of categories within the dataset given a fixed set of points inside the grey region (Fig. B-1). In this situation we have 20 total points, eight of which are in the locked region (see Table B-1).

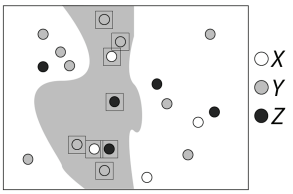

Fig. B-1: Sample spatial population for calculation.

	Restricted	Free
X	2	2
Y	4	7
Z	2	3
P	8	12

Table B-1: Counts of restricted and free individuals in each subpopulation (*X*,*Y*,*Z*) and the joint population (*P*).

	X	Y	Z
X	0	2	2
Y	3	5	3
Z	3	2	0

Table B-2: Actual nearest-neighbor counts.

	X	Y	Z
X	0.17	1.33	1.50
Y	3.17	5.08	2.75
Z	1.67	2.58	0.75

Table B-3: Expected nearest-neighbor counts.

Actual first nearest neighbor counts ($n_{\boxed{X}}$, $n_{\check{X}}$, etc.) are shown in Table B-2. We differentiate which relationships come from restricted points and which from unrestricted using the notation found elsewhere in the paper. Important to this calculation is the distinction of whether the nearest neighbor is also restricted or free.

Expected counts for this data are shown in Table B-3. These are calculated using Eq. (1) from the article. For instance, the calculation of $\mathrm{E}(\mathrm{C}_{X\to Y})$ is derived as follows:

$$\mathrm{E}(\mathrm{C}_{X\to Y}) = \mathrm{C}_{\boxed{X}\to\boxed{Y}} + \left(\mathrm{C}_{\boxed{X}\to\check{P}}\right)\frac{n_{\check{Y}}}{N_{\check{P}}} + \left(\mathrm{C}_{\check{P}\to\boxed{Y}}\right)\frac{n_{\check{X}}}{N_{\check{P}}} + \left(n_{\check{X}} \times \frac{\mathrm{C}_{\check{P}\to\check{P}}}{N_{\check{P}}}\right)\frac{n'_{\check{Y}}}{N_{\check{P}} - 1}$$

$$\mathrm{E}(\mathrm{C}_{X\to Y}) = 0 + (0)\frac{7}{12} + (1)\frac{2}{12} + \left(2 \times \frac{11}{12}\right)\frac{7}{11} = 0 + 0 + \frac{1}{3} + 1 = 1.33$$

The lowest expected count is of *Y* individuals with other *Y* individuals as nearest neighbors, while the weakest is of *X* individuals with other *X* individuals as nearest neighbors. These expectations are cemented partially by the existing restricted spatial structure, and also by the quantity of free *Y* points. Calculating the restricted *CLQ* for this dataset is a matter of dividing the actual counts by the expected counts.