Ripley's K function

Philip M. Dixon

Department of Statistics

Iowa State University

20 December 2001

Ripley's K(t) function is a tool to analyze completely mapped spatial point process data. Spatial point process data consists of the locations of events. These are usually recorded in 2 dimensions, but they may be locations along a line or in space. Here I will only describe K(t) for 2 dimensional spatial data. Completely mapped data includes the locations of all events in a predefined study area. Ripley's K(t) function can be used to summarize a point pattern, test hypotheses about the pattern, estimate parameters and fit models. Bivariate or multivariate generalizations can be used to describe relationships between two or more point patterns. Applications include spatial patterns of trees [29, 20, 10], herbaceous plants [28], bird nests [11], and disease cases [7]. Details of various theoretical aspects of K(t) are in books by Ripley [26], Diggle [6], Cressie [4], Stoyan and Stoyan [31]. Examples of computation and interpretation can be found in those books and also Upton and Fingleton [32].

Theoretical K(t) function

The K function is

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

[23, 24], where λ is the density (number per unit area) of events.

K(t) describes characteristics of the point processes at many distance scales. Alternative summaries (e.g. mean nearest-neighbor distance or the c.d.f. of distance from random points to their nearest neighbors, see [nearest neighbor methods]), do not have this property. Many ecological point patterns show a combination

of effects, e.g. clustering at large scales and regularity at small scales. The combination can be seen as a characteristic pattern in a plot of the K(t) function.

K(t) does not uniquely define the point processes in the sense that two different processes can have the same K(t) function [1, 15]. Also, while K(t) is related to the nearest-neighbor distribution function [26, p. 158], the two functions describe different aspects of a point process. In particular, processes with the same K(t) function may have different nearest-neighbor distribution functions, G(t), and vice versa. K(t) is also closely related to the pair correlation function, $g(t)$ [31, p. 249]. Stoyan and Penttinen [29] summarize the relationships between K(t) and other statistics for spatial point processes.

Although it is usual to assume stationarity, K(t) is interpretable for non-stationary processes because K(t) is defined in terms of a randomly chosen event. It is also customary to assume isotropy, i.e. that 1 unit of distance in the Y direction has the same effect as 1 unit of distance in the X direction. If the degree of anisotropy is known, the definition of the distance t can be adjusted.

Models for K(t)

For many point processes, the expectation in the numerator of the K(t) function (equation 1) can be analytically evaluated, so the K(t) function can be written down in closed form. The simplest, and most commonly used, is K(t) for a homogeneous Poisson process, also known as complete spatial randomness:

$$K(t) = \pi t^2 \tag{2}$$

A variety of processes can be used to model small scale regularity. Hard core processes are those in which events can not occur within some minimum distance of each other. In a Matern hard core process [17, pp. 47-48], locations are non-random thinning of a homogeneous Poisson process with intensity ρ. Any pair of events separated by less than a critical distance, δ, is deleted. The remaining events are a realization of a hard core

process. The K(t) function for this process is [4]

$$K(t) = \frac{2\rho\pi}{\exp(-\rho\pi\delta^2)} \int_0^t u\, k(u)\, \mathrm{d}u \tag{3}$$

where $k(u)$ describes the probability of retaining pair of events separated by a distance u:

$$k(u) = \begin{cases} 0 & h < \delta \\ \exp(-\rho V(h,\delta)) & h \geq \delta \end{cases} \tag{4}$$

$V(h,\delta)$ is the area of intersection of two circles, each of radius δ, with centers separated by a distance h. A sequential variant of this hard core process has a different K(t) function [4, p. 670].

Soft-core processes are those where the number of neighbors within some critical distance, δ, is smaller than expected under CSR, but the number is not zero. One example is a Strauss process [30], in which a fraction, $1-\gamma$, of the events within the critical distance, δ, are deleted. An approximation to the K(t) function for this process is [13]:

$$K(t) = \begin{cases} \gamma\pi t^2 & 0 < t \leq \delta \\ \pi t^2 - (1-\gamma)\pi\delta^2 & t \geq \delta \end{cases} . \tag{5}$$

Events may also be spatially clustered. One process that generates clustered locations is a Neyman-Scott process in 2 dimensions. 'Parent' events are a realization of a homogeneous Poisson process with intensity ρ. Each parent event, i, generates a random number of 'offspring' events, N_i, where N_i has a Poisson distribution with mean m. The locations of the offspring, relative to the parent individual, have a bivariate Gaussian distribution with mean $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and variance $\begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$. When locations of the parent events are ignored, locations of the clustered offspring events are a realization of a Neyman-Scott process [19]. The K(t) function for this process is [6]:

$$K(t) = \pi t^2 + (1 - e^{-t^2/4\sigma^2})/\rho \tag{6}$$

For a general Poisson cluster process with an arbitrary distributions for the number of offspring per parent, N, and their distance from the parent,

$$K(t) = \pi t^2 + \mathrm{E}[N(N-1)]\, F(t)/\rho\mu, \tag{7}$$

where $F(t)$ is the c.d.f. of the distance between two offspring from the same parent and μ is the mean number

of offspring per parent. K(t) functions can also be written down for other clustered and regular processes, see [6, pp 46-69], [4, pp 650-695], or [31, pp. 307-334] for details.

Estimating K(t)

Given the locations of all events within a defined study area, how can $\hat{K}(t)$ be estimated? K(t) is a ratio of a numerator and the density of events, λ. The density, $\hat{\lambda}$ can be estimated by N/A, where N is the observed number of points and A is the area of the study region. It is customary to condition on N, so the uncertainty in $\hat{\lambda}$ can be ignored, although unconditionally unbiased estimators have been suggested [9]. If edge effects are ignored, the numerator can be easily estimated by $N^{-1} \sum_i \sum_{j \neq i} I(d_{ij} < t)$, where d_{ij} is the distance between the i'th and j'th points, and $I(x)$ is the indicator function, with the value 1 if x is true and 0 otherwise. However, the boundaries of the study area are usually arbitrary. Edge effects arise because points outside the boundary are not counted in the numerator, even if they are within distance t of a point in the study area. Ignoring edge effects biases $\hat{K}(t)$, especially at large values of t.

A variety of edge-corrected estimators have been proposed. The most commonly used one is due to Ripley [23]:

$$\hat{K}(t) = \hat{\lambda}^{-1} \sum_i \sum_{j \neq i} w(l_i, l_j)\, I(d_{ij} < t)/N. \tag{8}$$

As before, d_{ij} is the distance between the i'th and j'th points, and $I(x)$ is the indicator function. The weight function, $w(l_i, l_j)$, provides the edge correction. It has the value of 1 when the circle centered at l_i and passing through the point l_j (i.e. with a radius of d_{ij}) is completely inside the study area. If part of the circle falls outside the study area (i.e. d_{ij} is larger than the distance from l_i to at least one boundary), $w(l_i, l_j)$ is the proportion of the circumference of that circle that falls in the study area. The effects of edge corrections are more important for large t because large circles are more likely to be outside the study area. Other edge corrected estimators have been proposed. They and their properties are summarized by [4, pp 616-618] and [31, pp 279-284]. Although $\hat{K}(t)$ can be estimated for any t, it is common practice to consider only $t <$ one-half the shortest dimension of the study area, if the study area is approximately rectangular, or $t < \sqrt{A/2}$, where A is the area of the study region.

$\hat{K}(t)$ is easy to compute, except perhaps for the geometric aspects of the edge corrections. Edge-corrected estimators are available in at least three Splus librarys (Splancs [27], Spatial [33], and S+SpatialStats [14]) and at least one SAS macro is available [18].

Evaluating spatial models

The simplest use of Ripley's K(t) function is to test complete spatial randomness, i.e. test whether the observed events are consistent with a homogeneous Poisson process. If so, $K(t) = \pi t^2$ for all t. In practice, it is easier to use $L(t) = (K(t)/\pi)^{1/2}$ and its estimator:

$$\hat{L}(t) = (\hat{K}(t)/\pi)^{1/2} \tag{9}$$

because Var $\hat{L}(t)$ is approximately constant under CSR [25]. Under CSR, $L(t) = t$. Deviations from the expected value at each distance, t, are used to construct tests of CSR. One approach is to test $L(t) - t = 0$ at each distance, t. Another is to combine information from a set of distances into a single test statistic, such as $\hat{L}_m = \sup_t \mid \hat{L}(t) - t \mid$ or $\hat{L}_s = \sum_t \mid \hat{L}(t) - t \mid$. Critical values can be computed by Monte-Carlo simulation [3] or approximated. For $\hat{L}_m$, approximate 5% and 1% critical values are $1.42\sqrt{A}/N$ and $1.68\sqrt{A}/N$ [25].

More complicated spatial processes, e.g. a Neyman-Scott process (6) or Strauss process (5), can be tested by similar comparisons, if all the parameters of that process were known. Usually, parameter values for more complicated spatial processes are not known *a-priori* and must be estimated. One reasonable estimator is to find the values, θ, that minimize a discrepancy measure between the observed $\hat{K}(t)$ and the theoretical $K(t, \theta)$. Diggle [6] suggests $\int_0^{t_0} [\hat{K}(t)^{0.25} - K(t)^{0.25}]^2 \, \mathrm{d}t$, which can be approximated by

$$D(\theta) = \sum_t [\hat{K}(t)^{0.25} - K(t, \theta)^{0.25}]^2, \tag{10}$$

where the sum is over a subset of values of t between 0 and t_0. The exponent of 0.25 is empirically chosen to give reasonable results for a variety of random and aggregated patterns [6, p 74]. The upper limit, t_0, is chosen to span the biologically important spatial scales. Large values of t_0 relative to the size of the study area should be avoided because of the large uncertainty in $\hat{K}(t)$ for large t. This model fitting approach can be extended to fit processes for which K(t) can not be written in closed form, so long as the process can be simulated [8].

Diagnostics for fitted models include estimation of residuals [4, p 656-657, for details] or comparison with simulated data. Given estimates of the parameters and an algorithm to simulate data from a particular spatial process, $\hat{K}(t)$ can be estimated for a set of simulated realizations. If the fitted model is reasonable, the observed $\hat{K}(t)$ function should be similar to the $\hat{K}(t)$ from the simulated data. A two-sided 95% pointwise confidence band for the fitted model can be estimated by simulating 39 realizations of the spatial pattern, then computing the minimum and maximum values of $\hat{K}(t)$ at a set of t values. If a one-sided bound is desired, it can obtained as the maximum (or minimum) from 19 simulations. The sampling uncertainty in these bounds can be reduced by increasing the number of simulations and calculating the appropriate quantiles of $\hat{K}(t)$ for each of value of t. This approach tends to overstate the confidence in the fit because the fit is evaluated using the the same data and loss function that were used to estimate the parameters [31, p 305].

K(t) functions for multivariate spatial patterns

The previous analyses considered only the location of an event; they ignored any other information about that event. Many point patterns include biologically interesting information about each point, e.g. species identifiers (if the points include more than one type of species), whether the individual survived or died (for spatial patterns of trees or other plants), and whether a location is a disease case or a randomly selected control. Such data are examples of multivariate spatial point patterns, which are examples of marked point patterns that have a small number of discrete marks. In the previous examples, the marks are the species identifier, the fate (live or dead) or the disease status (case or control). The univariate methods in the previous section can be used to analyze or model the spatial pattern of all individuals (ignoring the marks) or the separate patterns in each type of mark. However, many biological questions concern the relationships between marks, for which the multivariate methods described in this section are needed.

The generalization of K(t) to more than one type of point (a multivariate spatial point process) is

$$K_{ij}(t) = \lambda_j^{-1}\mathrm{E}\ \#\ \text{type } j \text{ events within distance } t \text{ of a randomly chosen type } i \text{ event} \tag{11}$$

When there are g types of events, there are g^2 K functions, $K_{11}(t), K_{12}(t), \cdots K_{1g}(t), K_{21}(t), \cdots K_{2g}(t), \cdots Kgg(t)$.

It is helpful to distinguish the cross-K functions, $K_{ij}(t)$ where $i \neq j$, from the self-K functions, $K_{ii}(t)$. Analytical expressions for $K_{ij}(t)$ are known for various multivariate point processes, see [6, pp. 90-103] or [4, pp. 699-707].

Estimators of each bivariate $K_{ij}(t)$ function are similar to estimators of univariate $K(t)$ functions. If edge corrections are not needed, $\hat{K}_{ij}(t) = (\hat{\lambda}_i \hat{\lambda}_j A)^{-1} \sum_k \sum_l I(d_{i_k,j_l} < t)$, where d_{i_k,j_l} is the distance between the k'th location of type i and the l'th location of type j and A is the area of the study region. Various edge corrections have been suggested; one common one is the extension of Ripley's estimator [12]:

$$\hat{K}_{ij}(t) = (\hat{\lambda}_i \hat{\lambda}_j A)^{-1} \sum_i \sum_j w(i_k, j_l) I(d_{i_k,j_l} < t), \tag{12}$$

where $w(i_k, j_l)$ is the fraction of the circumference of a circle centered at the k'th location of process i with radius d_{i_k,j_l} that lies inside the study area.

If the spatial process is stationary, corresponding pairs of cross K functions are equal, i.e. $K_{12}(t) = K_{21}(t), K_{ij}(t) = K_{ji}(t)$. When edge corrections are used, $\hat{K}_{ij}(t)$ and $\hat{K}_{ji}(t)$ are positively correlated but not equal, which suggests the use of a more efficient estimator, $K^*_{ij}(t) = (\hat{\lambda}_j \hat{K}_{ij}(t) + \hat{\lambda}_i \hat{K}_{ji}(t))/(\hat{\lambda}_i + \hat{\lambda}_j)$ [16], although other linear combinations of $\hat{K}_{ij}(t)$ and $\hat{K}_{ji}(t)$ may have even smaller variance.

Questions about the relationship between two spatial processes can be asked in two different ways. The independence approach [16] conditions on the marginal structure of each process and asks questions about the interaction between the two processes. The random labelling approach [5] conditions on the observed locations and ask questions about the process that assigns labels to points. The distinction between independence and random labelling of two spatial processes requires some care and consideration. When there is no relationship between two processes, the two approaches lead to different expected values of the cross K function, $K_{12}(t)$ and different nonparametric test procedures.

Under independence, the cross-type K function, $K_{12}(t) = \pi t^2$, without regard to the individual univariate spatial patterns of the two types of events. It is easier to work with the corresponding $L^*_{ij}(t) = (K^*_{ij}(t)/\pi)^{1/2}$ function, because the variance of $\hat{K}^*_{12}(t)$ is approximately constant. Under independence, $L^*_{12}(t) = t$. Values of $\hat{L}_{12}(t) - t > 0$ indicate attraction between the two processes at distance t; values < 0 indicate repulsion. As with the univariate functions, tests can be based on the distribution of $\hat{L}_{12}(t)$ (or $\hat{K}^*_{12}(t)$) at each distance t, or

on summary statistics such as $\max_{0<t\leq t_0} \mid \hat{L}_{12}(t) - t \mid$. Determining critical values for a test of independence is more difficult than in the univariate setting because inferences are conditional on the marginal structure of each type of event [16]. This requires maintaining the univariate spatial pattern of each process, but breaking any dependence between them. If both univariate spatial patterns can be described by parametric models, it is easy to estimate the critical values by simulating independent realizations of each parametric spatial process.

The method of toroidal shifts provides a nonparametric way to test independence when the study area is rectangular. All the locations for one type of event are displaced by a randomly chosen $(\Delta X, \Delta Y)$. The study area is treated as a torus, so the upper and lower edges are connected and the right and left edges are connected. $\hat{K}^*_{12}(t)$ and the desired test statistics are computed from the randomly shifted data. Random displacement and estimation of the test statistic(s) are repeated a large number of times to estimate critical values for the test statistic(s). In practice, the toroidal shift method appears to be sensitive the assumption that the multivariate spatial process are stationary.

Under random labelling, of $K_{12}(t) = K_{21}(t) = K_{11}(t) = K_{22}(t) = K(t)$, i.e. all four bivariate K(t) functions equal the K function for all events, ignoring their labels, because each type of event is a random thinning of all events. Departure from random labelling can be examined using pairwise differences between K functions. Each pairwise difference evaluates different biological effects. $\hat{K}_{11}(t) - \hat{K}_{22}(t)$ evaluates whether one type of event is more (or less) clustered than the other. Diggle and Chetwynd [7] use this to examine disease clustering. $K_{11}(t) - K^*_{12}(t)$ and $K_{22}(t) - K^*_{12}(t)$ evaluate whether one type of event tends to be surrounded by other events of the same type. Gaines et al. [11] use this to examine spatial segregration of waterbird foraging sites.

Inference is based either on Monte-Carlo simulation or a normal approximation. The appropriate simulation is to fix the combined set of locations and the number of each type of event, then randomly assign labels to locations. In general, the variance of any of the three differences increases with t, so summary statistics should be based on the studentized difference, e.g. $\max_t \left[\mid \hat{K}_{11}(t) - \hat{K}_{22}(t) \mid / sd(\hat{K}_{11}(t) - \hat{K}_{22}(t)) \right]$. The variance can be calculated given t, the number of each type of point, and the spatial pattern of the combined set of locations.

Example: trees in a swamp hardwood forest

I will illustrate the use of Ripley's K(t) functions to examine spatial patterns using the data set described in nearest neighbor methods. These data are the locations of all 630 trees (stems > 11.5 cm dbh) in a 1 ha plot of swamp hardwood forest in South Carolina, USA. These trees represent 13 species, but most (over 75%) are black gum, *Nyssa sylvatica*, water tupelo, *Nyssa aquatica*, or bald cypress, *Taxodium distichum*. Visually (Figure 1 of nearest neighbor methods), trees seem to be scattered randomly throughout the plot, but cypress trees appear to be clustered. Ripley's K(t) functions provide a way to summarize those spatial patterns, fit models to describe the patterns, and compare the patterns of different species.

The spatial pattern of all 630 trees and the spatial pattern of the 91 cypress trees can be described using univariate K(t) statistics. Because the plot is 50m x 200m, K(t) is estimated for distances up to 35m in 1m increments. $\hat{K}(t)$ for all trees lies above the expected value of πt^2 for all distances between 1 and 10 m, but it is hard to see the effects because of the large range of the Y axis (Figure 1a). The patterns are much clearer in the plot of L(t)-t vs. distance (Figure 1b). There is evidence of weak, but statistically significant, clustering of trees at distances up to 17m. $\hat{L}(t) - t$ lies above the upper 97.5% quantile for all distances up to 17m and above the expected value of 0 for all distances up to 35 m.

Although the deviation from complete spatial randomness is statistically signficant, its magnitude is small. A biologically relevant summary of the clustering is to compute the proportion of excess trees in a specified circle around a randomly chosen tree. This is estimated by $\hat{K}(t)/\mathrm{E}\ \hat{K}(t) - 1$ at a specific distance t. For all trees, this proportion is small (5.6%) for 6m radius circles.

For cypress trees, the plot of their $\hat{L}(t) - t$ (Figure 1c) indicates two different departures from randomness. At very short distances ($\leq$ 2m), $\hat{L}(t) - t$ is less than 0, indicating spatial regularity. At longer distances ($\geq$ 3m), $\hat{L}(t) - t$ is larger than 0, indicating spatial clustering. The observed $\hat{L}(t) - t$ curve is much larger than the pointwise 0.975 quantiles for distances from 4m to 27m and both the max and mean summary statistics are highly significant (p = 0.001). This clustering represents a biologically large effect. In a 6m radius circle, each cypress tree is surrounded by an estimated 88% more cypress trees around it than expected if cypress were

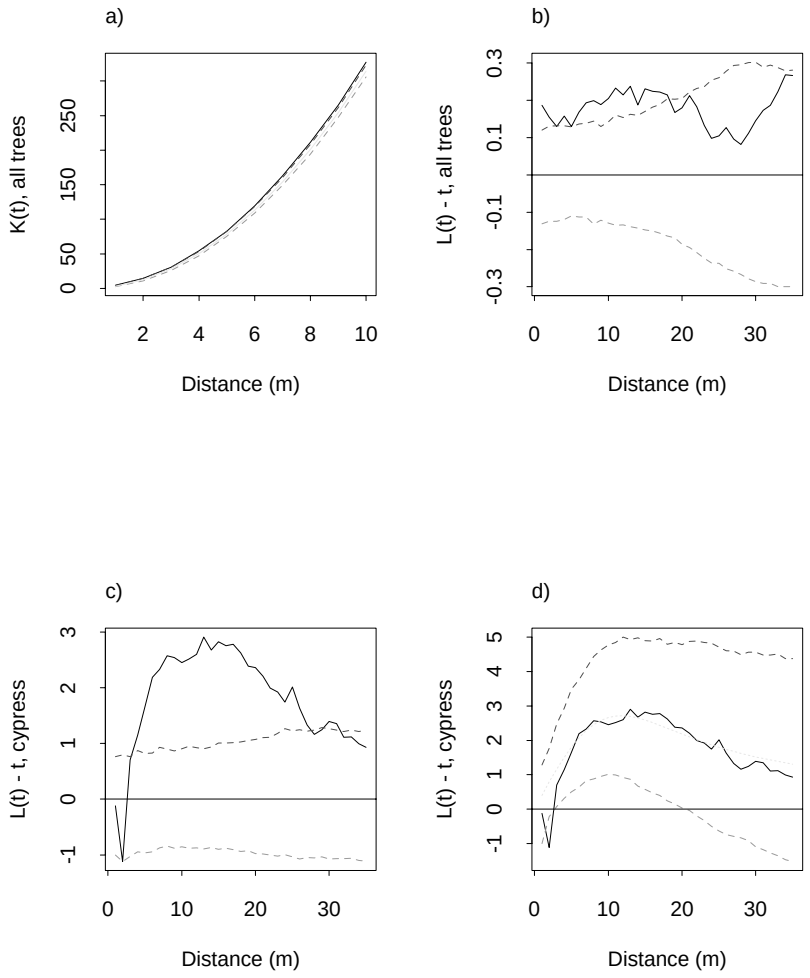

Figure 1: K(t) and L(t) plots for swamp trees.

a) Plot of $\hat{K}(t)$ against distance up to 10m for all 630 trees.

b) Plot of $\hat{L}(t) - t$ for all 630 trees. Solid horizontal line provides a reference for L(t) under complete spatial randomness. Dashed lines are 0.025 and 0.975 quantiles of $L(t) - t$ estimated from 999 simulations.

c) Plot of $\hat{L}(t) - t$ for 91 cypress trees. Line markings are the same as in b).

d) Plot of $\hat{L}(t) - t$ for 91 cypress trees fit to a Neyman-Scott process (equation 6). Solid horizontal line provides a reference under complete spatial randomness. Dotted line is $L(t) - t$ using the estimated parameters. Dashed lines are the 0.025 and 0.975 quantiles of $L(t) - t$ estimated from 999 simulations of a Neyman-Scott process.

randomly distributed.

The larger scale clustering pattern can be described by fitting a Neyman-Scott process (equation 6) to the locations. The Neyman-Scott process describes a clustered set of locations in terms of an unknown number of randomly distributed 'mothers', each with a random number of 'daughters' distributed around the 'mother'. Parameters are estimated by minimizing the loss function given by equation (10) for distances from 5m to 35m. Shorter distances were excluded because I was uninterested in the small scale spatial regularity. The choices of 5m and 35m are arbitrary, but other reasonable values gave similar results. The estimated parameters are the variance of daughter locations, $\hat{\sigma}^2$, = 24.1 m^2 and the density of 'mothers', $\hat{\rho} = 0.0034$. The fitted K(t) function is very close to $\hat{K}(t)$ for distances larger than 5m (Figure 1d). Point-wise 95% confidence bounds for the fitted K(t) function are computed by repeatedly simulating the Neyman-Scott process using the estimated $\hat{\sigma}^2$ and $\hat{\rho}$, then estimating the 0.025 and 0.975 quantiles of $\hat{K}(t)$ at each distance. The observed $\hat{K}(t)$ curve falls well inside the bounds except at 2m. This deviation is due to the small scale regularity. It is possible to fit a more complicated process that combines small scale regularity and larger scale clustering, similar to the more biologically detailed processes fit by Rathbun and Cressie [22], but the theoretical K(t) function would have to be estimated by simulation [8].

Even though a Neyman-Scott process describes the spatial pattern quite well, it is not appropriate to conclude that it is the mechanism responsible for the clustering. Other mechanisms can lead to exactly the same pattern [2]. The plot, like most of the swamp, is not a homogeneous environment. In particular, some areas are above the mean water level, others are in shallow water, and still others are deep channels. Cypress are known to be most successful in parts of the swamp with shallow to moderately deep water. Other trees (e.g. black gum) prefer drier areas. The clustering of cypress could simply be a response to a heterogeneous environment; this hypothesis could be tested if environmental data such as water depth or elevation were available [21].

Patterns with small scale regularity and large scale clustering are quite common for ecological data, especially when individuals are large, as cypress trees can be. Diameter at breast height for the 91 cypress trees in the plot ranges from 15 cm to 180 cm, with a median of 105 cm. It is physically impossible for two median-sized cypress to be closer than 1 m. However, this small scale separation of stems occurs in conjunction with a larger scale

clustering of individuals into patches. The K(t) and L(t) statistics provide evidence of both ecological processes.

Visually, cypress and black gum trees appear to be spatially segregated, i.e. cypress tend to be found in patches of mostly cypress and black gum tend to be found in patches of mostly black gum. This pattern can be described and evaluated using the bivariate K statistics. I will use the subscripts C to represent cypress patterns and G to represent black gum patterns. As described above, two different hypotheses (random labelling and independent processes) could be used to describe the absence of dependence between cypress and black gum.

Under random labelling, $K_{CC}(t) = K_{CG}(t) = K_{GG}(t)$. If cypress trees tend to occur in patches of other cypress trees, then $K_{CC} > K_{CG}$, while if black gums tend to occur in patches of other black gums, $K_{GG} > K_{GC}$. Each species can be evaluated by estimating differences of K functions and their uncertainty under randomly labelling. The plot of $\hat{K}_{CC}(t) - \hat{K}^*_{CG}(t)$ is above zero and well outside the 95% quantiles for all distances larger than 3 m (Figure 2a). The plot of $\hat{K}_{GG}(t) - \hat{K}^*_{CG}(t)$ is above zero for all distances larger than 2 m and well outside the 95% quantiles for all distances larger than 3 m (Figure 2b). Summary statistics combining tests at all distances are highly significant (P value < 0.001). These two species are not randomly labelled; instead, both are spatially segregated.

The two sets of locations are also not spatially independent. If they were, $K_{CG}(t) = \pi t^2$ at all distances, t. As with univariate tests of randomness, it is easier to visualize patterns in the equivalent $\hat{L}_{CG}(t) - t$ plot (Figure 2c). For cypress and black gum, $\hat{L}_{CG}(t) - t$ is less than 0 for all distances and below the lower 0.025 quantile for most distances larger than 3 m (Figure 2c). The number of black gum trees in the neighborhood of cypress (or equivalently the number of cypress trees in the neighborhood of black gums) is less than expected. The observed value of $\hat{L}_{CG}(t) - t$ under toroidal rotation is not as extreme a value as those seen under random labelling. The point-wise two-sided p-values for the test of independence range from 0.002 to 0.082 for distances from 3m to 35m. The conclusion is the spatial pattern of cypress trees is not independent of the black gum spatial pattern.

The hypotheses of independent processes and random labelling are not equivalent. However, when both hypotheses are appropriate, which test is the more powerful? A detailed comparison has not been made, but it is possible to compare distributions of $\hat{K}^*_{CG}(t)$ using specific data sets. $\hat{K}^*_{CG}(t)$ for random labelling is less variable

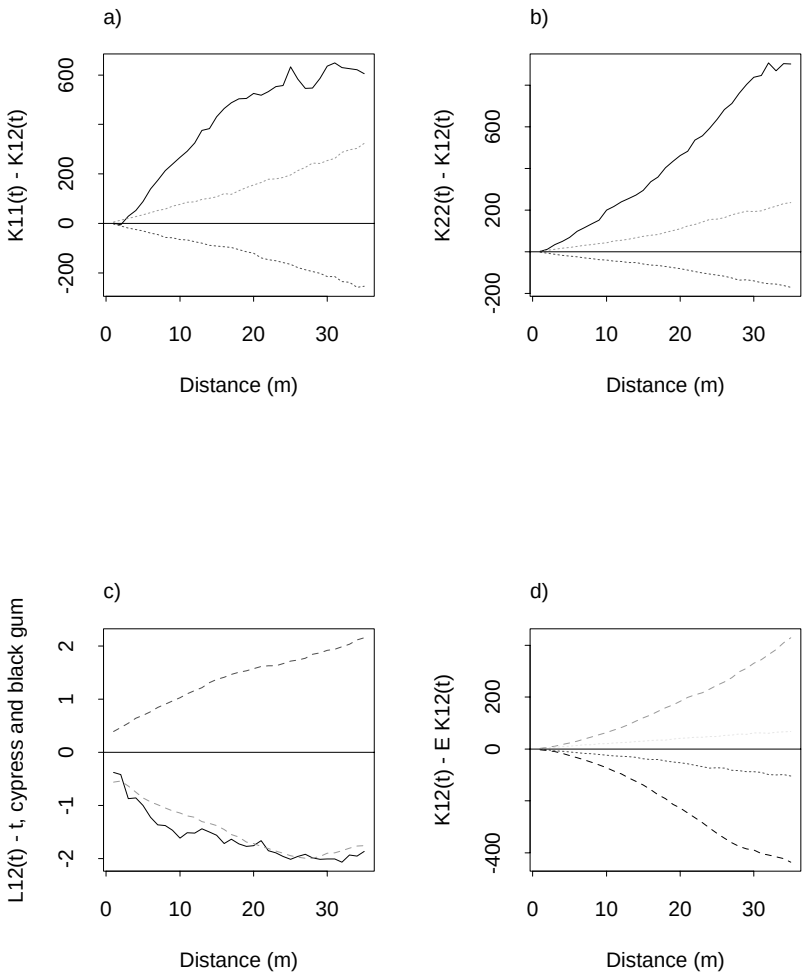

Figure 2: Bivariate K(t) plots to evaluate the spatial relationship between cypress and black gum trees.

a) Plot of $\hat{K}_{11}(t) - \hat{K}_{12}(t)$ for cypress trees. Solid horizontal line at 0 provides a reference for random labelling. Dotted lines are 0.025 and 0.975 quantiles of $L(t) - t$ estimated from 999 random relabellings .

b) Plot of $\hat{K}_{22}(t) - \hat{K}_{12}(t)$ for black gum trees. Solid horizontal line at 0 provides a reference for random labelling. Dotted lines are 0.025 and 0.975 quantiles of $L(t) - t$ estimated from 999 random relabellings .

c) Plot of $\hat{L}_{12}(t) - t$ for cypress and black gum trees. Solid horizontal line at 0 provides a reference for independence of the two spatial processes. Dotted lines are 0.025 and 0.975 quantiles of $L(t) - t$ estimated from 999 random toroidal shifts.

d) Comparison of 0.025 and 0.975 quantiles for $\hat{K}_{12}(t)$ computed by random labelling (dotted lines) and random toroidal shifts (dashed lines).

than $\hat{K}^*_{CG}(t)$ for toroidal rotation. This is illustrated using the 0.025 and 0.975 quantiles of $\hat{K}^*_{CG}(t)$ (Figure 2d). The random labelling quantiles are considerably less extreme than the toroidal rotation quantiles.

References

[1] Baddeley, A. J. and Silverman, B. W. 1984, A cautionary example on the use of second-order methods for analyzing point patterns. *Biometrics* 40, 1089-1093.

[2] Bartlett, M. S. 1964, The spectral analysis of two-dimensional point processes. Biometrika 51, 299-311.

[3] Besag, J. and Diggle, P. J. 1977, Simple Monte Carlo tests for spatial pattern. *Applied Statistics* 26, 327-333.

[4] Cressie, N. A. C. 1991, *Statistics for Spatial Data*, Wiley, New York.

[5] Cuzick, J. and Edwards, R. 1990, Spatial clustering for inhomogeneous populations (with discussion), *Journal of the Royal Statistical Society, Series B* 52, 73-104.

[6] Diggle, P. J. 1983, *Statistical Analysis of Spatial Point Patterns*, Academic Press, New York.

[7] Diggle, P. J. and Chetwynd, A. G. 1991, Second-order analysis of spatial clustering for inhomogeneous populations, *Biometrics* 47, 1155-1163.

[8] Diggle, P. J. and Gratton, R. J. 1984, Monte Carlo methods of inference for implicit statistical models (with discussion). *Journal of the Royal Statistical Society, Series B* 46, 193-227.

[9] Doguwa, S. I. and Upton, G. J. G. 1989, Edge-corrected estimators for the reduced second moment measure of point processes. *Biometrical Journal* 31, 563-576.

[10] Duncan, R. P. 1993. Testing for life historical changes in spatial patterns of four tropical tree species in Westland, New Zealand, *Journal of Ecology* 81, 403-416.

[11] Gaines, K.F., Bryan, A.L. Jr, and Dixon, P.M. 2000. The effects of drought on foraging habitat selection in breeding wood storks in coastal Georgia. *Waterbirds* 23, 64-73.

[12] Hanisch, K. H. and Stoyan, D. 1979, Formulas for second-order analysis of marked point processes, *Mathematische Operationsforschung und Statistik, Series Statistics* 14, 559-567.

[13] Isham, V., 1984, Multitype Markov point processes: some applications*, Proceedings of the Royal Society of London, Series A*, 391, 39-53.

[14] Kaluzny, S. P., Vega, S. C., Cardoso, R. P. and Shelly, A. A. 1996. *S+SPATIALSTATS User's Manual, Version 1.0*. MathSoft Inc, Seattle, WA.

[15] Lotwick, H. W. 1984, Some models for multitype spatial point processes, with remarks on analysing multitype patterns, *Journal of Applied Probability*, 21, 575-582.

[16] Lotwick, H. W. and Silverman, B. W. 1982, Methods for analysing spatial processes of several types of points, *Journal of the Royal Statistical Society, Series B*, 44, 406-413.

[17] Matern, B. 1986, *Spatial Variation, 2nd edition*. Lecture Notes in Statistics, number 36, Springer-Verlag, Berlin.

[18] Moser, E. B. 1987, The analysis of mapped spatial point patterns. *Proceedings of the 12th Annual SAS Users Group International Conference* 12, 1141-1145.

[19] Neyman, J. and Scott, E. L. 1952, A theory of the spatial distribution of galaxies, *Astrophysical Journal* 116, 144-163.

[20] Peterson, C. J. and Squiers, E. R. 1995 An unexpected change in spatial pattern across 10 years in an Aspen-White Pine forest, *Journal of Ecology* 83, 847-855.

[21] Rathbun, S. L. 1996. Estimation of Poisson intensity using partially observed concomitant variables. *Biometrics* 52, 226-242.

[22] Rathbun, S. L. and Cressie, N. 1994, A space-time survival point process for a longleaf pine forest in southern Georgia, *Journal of the American Statistical Association* 89, 1164-1174.

[23] Ripley, B. D. 1976, The second-order analysis of stationary point processes. *Journal of Applied Probability* 13, 255-266.

[24] Ripley, B. D. 1977, Modelling spatial patterns, *Journal of the Royal Statistical Society, Series B* 39, 172-192.

[25] Ripley, B. D. 1979, Tests of 'randomness' for spatial point patterns. *Journal of the Royal Statistical Society, Series B* 41, 368-374.

[26] Ripley, B. D. 1981, Spatial Statistics. Wiley, New York.

[27] Rowlinson, B.S. and Diggle, P. J. 1992, *Splancs: Spatial Point Pattern Analysis Code in S-Plus*. Technical Report 92/63, Lancaster University, U.K.

[28] Stamp, N. E. and Lucas, J. R. 1990, Spatial patterns and dispersal distance of explosively dispersing plants in Florida sandhill vegetation, *Journal of Ecology* 78, 589-600.

[29] Stoyan, D. and Penttinen, A. 2000, Recent applications of point process methods in forestry statistics. *Statistical Science* 15, 61-78.

[30] Strauss, D. J. 1975, A model for clustering, *Biometrika* 62, 467-475.

[31] Stoyan, D. and Stoyan, H. 1994. *Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics*. Wiley, Chichester.

[32] Upton, G. J. G. and Fingleton, B. 1985. *Spatial Data Analysis by Example, Volume 1. Point pattern and quantitative data* Wiley, Chichester.

[33] Venables, W. N. and Ripley, B. D. 1994, *Modern Applied Statistics with S-Plus*, Springer-Verlag, New York.