



# Kernel Intensity Estimation of 2-Dimensional Spatial Poisson Point Processes From k-Tree Sampling

## Citation

Ellison, Aaron M., Nicholas J. Gotelli, Natalie Hsiang, Michael Lavine, and Adam B. Maidman. 2014. "Kernel Intensity Estimation of 2-Dimensional Spatial Poisson Point Processes From k-Tree Sampling." JABES (May 1).

## Published Version

doi:10.1007/s13253-014-0175-0

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:12308567>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

# Kernel Intensity Estimation of 2-Dimensional Spatial Poisson Point Processes from $k$ -Tree Sampling

<sup>1</sup>, Aaron M. Ellison<sup>1</sup>, Nicholas J. Gotelli<sup>2</sup>, Natalie Hsiang<sup>3</sup>, Michael Lavine<sup>4</sup>, and Adam B. Maidman<sup>4</sup>

<sup>1</sup>Harvard Forest, Harvard University, 324 North Main Street, Petersham, MA 01366 USA

<sup>2</sup>Department of Biology, University of Vermont, Burlington, VT 05405 USA

<sup>3</sup>Mount Holyoke College, South Hadley, MA, 01075 USA

<sup>4</sup>Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003 USA

April 14, 2014

field work supported by NSF grant DEB 0541680

## Abstract

To estimate the spatial intensity (density) of plants and animals, ecologists often sample populations by pre-specifying a spatial array of points, then measuring the distance from each point to the  $k$  nearest organisms, a so-called  $k$ -tree sampling method. A variety of *ad hoc* methods are available for estimating intensity from  $k$ -tree sampling data, but they assume that two distinct points of the array do not share nearest neighbors. However, nearest neighbors are likely to be shared when the population intensity is low, as it is in our application. The purpose of this paper is twofold: *a*) to derive and use for estimation the likelihood function for a  $k$ -tree sample under an inhomogeneous Poisson point-process model and *b*) to estimate spatial intensity when nearest neighbors are shared. We derive the likelihood function for an inhomogeneous Poisson point-process with intensity  $\lambda(x, y)$  and propose a likelihood-based, kernel-smoothed estimator  $\hat{\lambda}(x, y)$ . Performance of the method for  $k = 1$  is tested on four types of simulated populations: two homogeneous populations with low and high intensity, a population simulated from a bivariate normal distribution of intensity, and a “cliff” population in which the region is divided into high- and low-intensity subregions. The method correctly detected spatial variation in intensity across different subregions of the simulated populations. Application to 1-tree samples of carnivorous pitcher plants populations in four New England peat bogs suggests that the method adequately captures empirical patterns of spatial intensity. However, our method suffers from two evident sources of bias. First, like other kernel smoothers, it underestimates peaks and overestimates valleys. Second, it has positive bias analogous to that of the MLE for the rate parameter of Exponential random variables.

# 1 Introduction

Spatial intensity (commonly, “population density”) is a fundamental property of plant and animal populations, but is challenging to estimate accurately (e.g. Barbour et al., 1999; Byth and Ripley, 1980; Damggard, 2009; Diggle, 1975, 1977; Murdoch, 1994; Patil et al., 1979; Pyle and Ehrlich, 2010). The most straightforward way to estimate spatial intensity is to count all the organisms in a fixed area at a particular time. However, two constraints commonly limit the use of such a “simple” method. First, if the numbers are high, a complete census may not be possible in a limited amount of time even if the sampled area is small. Second, even if time is unlimited, some individuals may not be detected because they are small, hidden, or overlooked (e.g. Mackenzie et al., 2006).

For conspicuous life-history stages of sessile organisms such as most rooted plants, encrusting algae, and many aquatic invertebrates, detection probability may be high but populations are often very dense. Ecologists have developed a number of flexible and cost-effective plotless sampling methods to estimate population density (e.g. Kleinn and Vilčko, 2006) and other parameters (e.g. Augustin et al., 2009). These so-called  $k$ -tree sampling methods are based on a sample of the  $k$  nearest individual organisms to each of a predetermined fixed or randomly located set of  $n$  points (e.g. Diggle, 1975, 1977; Nothdurft et al., 2010; Magnussen et al., 2012). A large number of estimators for population density based on  $k$ -tree samples have been proposed (Magnussen, 2012; Magnussen et al., 2012), but all of them assume that the  $k$  individuals associated with one predetermined point are distinct from the individuals associated with every other fixed point. In practice, this can occur only if sample points are widely spaced relative to the distances between pairs of individuals or, equivalently, if individuals are relatively dense and reliably seen.

The work described here — to estimate spatial intensity of a population of plants from a  $k$ -tree sample with at least one individual plant being nearest-neighbor to at least two sample points — was motivated by an ecological study of carnivorous pitcher plants (*Sarracenia purpurea*) that grow in rain-fed peatlands (bogs) in eastern North America. Bogs are fragile habitats and research within them is carefully regulated by state permitting authorities, so plotless sampling in relatively small areas is preferred because it minimizes damage to the habitat. At the same time, because of common short-range seed dispersal and rare long-distance dispersal (Ellison and Parker, 2002), pitcher plants are often sparsely distributed and spatial intensity can vary dramatically from place to place. The combination of small areas in which sampling is permitted and, on average, low plant population intensity means that even for 1-tree samples, many of the sample points share the nearest neighboring plant: a situation not encountered in the more common applications of  $k$ -tree sampling. Thus, existing methods for estimating intensity from  $k$ -tree samples are inapplicable.

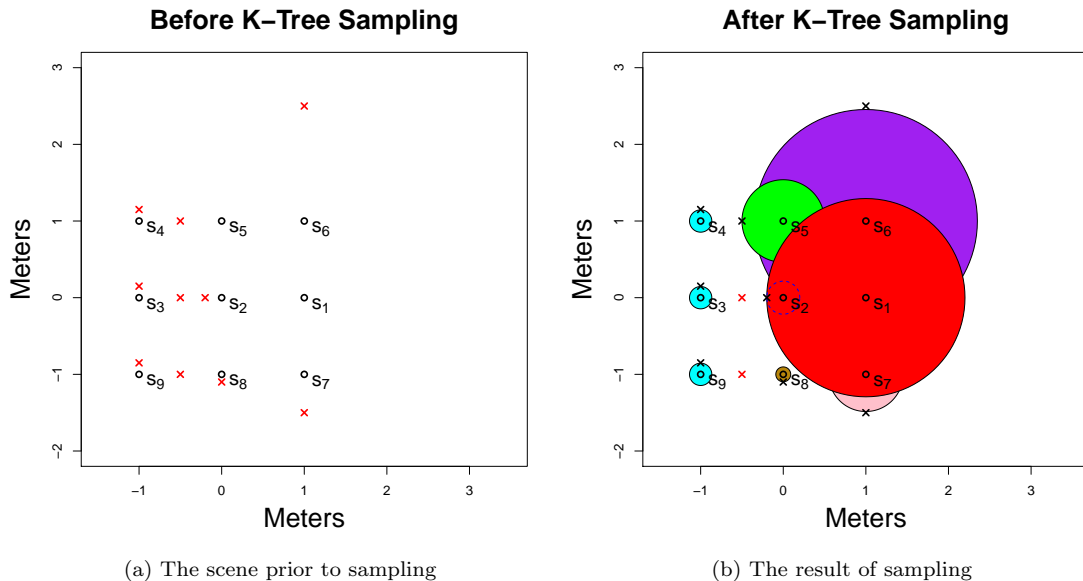


Figure 1: 1-tree sampling. In the left, pre-sampling, panel, objects are denoted by red  $\times$ 's and the pre-determined sampling points are denoted by black circles and subscripts indicating order. In the right, post-sampling, panel, nearest objects are denoted by black  $\times$ 's while non-nearest objects remain red. Colored regions indicate  $A_i$  ( $k = 1$ , so the second subscript is omitted.): red is  $A_1$ , green is  $A_5$ , purple is  $A_6$ , and pink is  $A_7$ .  $A_2$  is empty. The circle around  $s_2$  is denoted by the dotted blue line.

This paper models the locations of the  $k$ -tree sample as arising from an inhomogeneous Poisson process with intensity parameter  $\lambda(x, y)$ . After deriving the likelihood function in Section 3, we show how to compute a kernel estimate  $\hat{\lambda}(x, y)$  in Section 4. The accuracy and bias of our method are examined in a simulation study with four different types of intensity functions in Section 5. Section 6 discusses computational issues and Section 7 applies our method to 1-tree samples of pitcher plant populations in four New England bogs.

## 2 Notation

In  $k$ -tree sampling, we first specify  $n$  sample sites,  $\{(x_1^*, y_1^*), \dots, (x_n^*, y_n^*)\} \equiv \{s_1, s_2, \dots, s_n\} \equiv S$ , then sample the nearest  $k$  objects to each  $s_i$ . The  $s_i$  could be chosen either deterministically, say in a rectangular array, or randomly, say uniformly over a region of interest. Let  $O_{ij} = (x_{ij}, y_{ij})$  be the location of the  $j$ 'th nearest object (nearest neighbor) to  $s_i$  and let  $O = \{O_{ij}\}$ . If  $k = 1$ , then  $j = 1$  by necessity and we omit the second subscript. Some of the  $O_{ij}$ 's may be repeats, as a single object may be a neighbor of both  $s_i$  and  $s_{i'}$ . As both polar and Cartesian coordinates will be needed, let  $(r_{ij}, \theta_{ij})$  denote the polar coordinates of  $O_{ij}$  with respect to an origin at point  $s_i$ . To be clear,  $(r_{ij}, \theta_{ij}) \neq (r_{i'j'}, \theta_{i'j'})$  even if  $O_{ij} = O_{i'j'}$  because the polar

coordinates are with respect to different origins. Referring to Figure 1, the object at  $(-0.2, 0)$  is the nearest object to both  $s_1$  and  $s_2$ ; thus  $O_1 = O_2 = (-0.2, 0)$ , but  $(r_1, \theta_1) = (1.2, \pi)$  and  $(r_2, \theta_2) = (0.2, \pi)$ . (The second subscript is omitted because the figure displays only first-nearest neighbors.)

Here, we consider the 2-dimensional spatial point process describing the distribution of  $O$  in a region encompassing  $S$ . We focus on an inhomogeneous Poisson spatial point process with parameter, or intensity function,  $\lambda(x, y)$ . For any region  $A$ , the number of points  $N_A$  that occur in  $A$  has a Poisson distribution with parameter  $\lambda_A \equiv \int_A \lambda(x, y) dx dy$  and, if  $A$  and  $B$  are disjoint regions,  $N_A$  is independent of  $N_B$ . Our goal is to estimate  $\lambda$ , a function of location  $(x, y)$ .

### 3 The Likelihood Function

The derivation of the likelihood function  $L(\lambda)$ , though not  $L(\lambda)$  itself, depends on the order in which objects are considered. We use lexicographic order and write  $(i, j) < (i', j')$  if either  $i < i'$  or  $i = i'; j < j'$ .

The likelihood function is

$$L(\lambda) = p(O_{11} | \lambda) \times p(O_{12} | O_{11}, \lambda) \times \cdots \times p(O_{nk} | \{O_{ij} : (i, j) < (n, k)\}, \lambda). \quad (1)$$

When we consider  $O_{ij}$ , we learn not only that there is an object at  $O_{ij}$ , but also that there are no other previously undiscovered objects within radius  $r_{ij}$  of sample site  $s_i$ . But, as illustrated in the right-hand side of Figure 1, some of the circle of radius  $r_{ij}$  may already be known to contain no objects. Define  $A_{ij}$  to be that part of the circle of radius  $r_{ij}$  centered at  $s_i$ , not accounted for by  $\{O_{i'j'} : (i'j') < (i, j)\}$ . That is,  $A_{ij}$  is the region we discover to be empty when we consider  $O_{ij}$  and that we did not already discover to be empty by considering objects earlier in the lexicographic order. The colored sub-regions in Figure 1 show  $A_1, \dots, A_9$ . (The figure depicts a case with  $k = 1$ , so the second subscript is omitted.)  $A_{11}$  will always be a circle centered at  $s_1$  with radius  $r_{11}$ . But the other  $A_{ij}$  need not be circles, need not be connected, and could be empty. Let  $A = \bigcup A_{ij}$  be the entire region searched. The individual  $A_{ij}$ 's depend on the order in which sample points are considered, but their union  $A$  does not.

The first term in (1),  $p(O_{11} | \lambda)$ , is the limit, as  $\delta, \epsilon \rightarrow 0$ , of the probability that  $(x_{11}, y_{11})$  is in the small

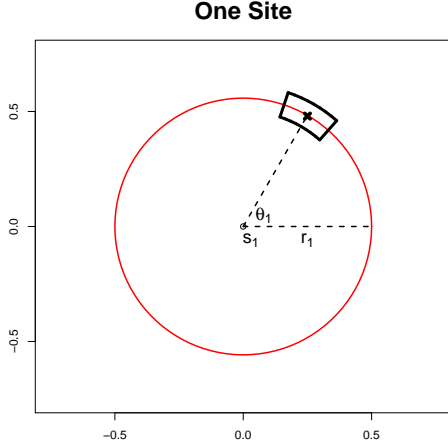


Figure 2: The black  $\times$  is the nearest object to  $s_1$ . The red circle encloses the region containing no objects formed by radius  $r_1$ . The box is defined by  $\{(r_1 - \epsilon, r_1 + \epsilon) \times (\theta_1 - \delta, \theta_1 + \delta)\}$ .

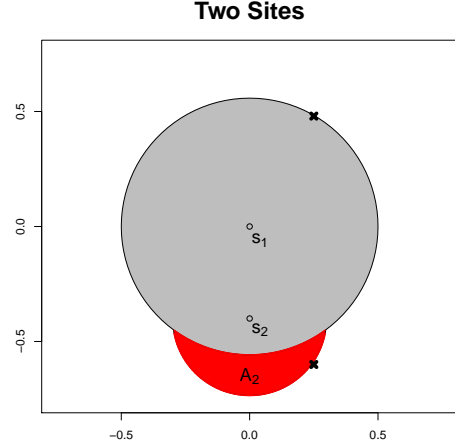


Figure 3: The black  $\times$  at  $(.25, -0.6)$  is the nearest object to  $s_2$ . The red region is  $A_2$ .

black box in Figure 2, divided by the size of that box:

$$\begin{aligned}
p(O_{11} | \lambda) &= p(x_{11}, y_{11} | \lambda) = \lim_{\substack{\epsilon \rightarrow 0 \\ \delta \rightarrow 0}} \frac{\Pr[(x_{11}, y_{11}) \in \text{box} | \lambda]}{\text{size of box}} \\
&= \lim_{\substack{\epsilon \rightarrow 0 \\ \delta \rightarrow 0}} \Pr[\text{no objects with } r < r_{11} - \epsilon | \lambda] \\
&\quad \times \Pr[\text{no objects with } r \in (r_{11} - \epsilon, r_{11} + \epsilon) \text{ and outside of box} | \lambda] \\
&\quad \times \frac{\Pr[\text{one object in box} | \lambda]}{\text{size of box}} \\
&= \lim_{\substack{\epsilon \rightarrow 0 \\ \delta \rightarrow 0}} \exp\left(-\int_{r < r_{11} + \epsilon, \text{ outside of box}} \lambda(x, y) dx dy\right) \\
&\quad \times \frac{\exp\left(-\int_{\text{box}} \lambda(x, y) dx dy\right) \times \int_{\text{box}} \lambda(x, y) dx dy}{\text{size of box}} \\
&= \lim_{\substack{\epsilon \rightarrow 0 \\ \delta \rightarrow 0}} \exp\left(-\int_{r < r_{11} + \epsilon} \lambda(x, y) dx dy\right) \times \frac{\lambda(x_{11}, y_{11}) \times \text{size of box}}{\text{size of box}} \\
&= \lambda(x_{11}, y_{11}) \times \exp\left(-\int_{r < r_{11}} \lambda(x, y) dx dy\right) \\
&= \lambda(O_{11}) \times e^{-\lambda_{A_{11}}}.
\end{aligned} \tag{2}$$

Subsequent terms are of two types. One type occurs when  $O_{ij}$  represents a newly-discovered object, not among the previous  $\{O_{i'j'}\}$ . For such terms, the derivation is similar to (2) and yields  $\lambda(O_{ij})e^{-\lambda_{A_{ij}}}$ . The second type occurs when  $O_{ij}$  represents a previously discovered object. For the second type we learn only



that there are no other objects in  $A_{ij}$  and the corresponding term in the likelihood function is the probability that a particular Poisson random variable equals 0:  $e^{-\lambda A_{ij}}$ . To summarize,

$$p(O_{ij} | \{O_{i'j'} : (i', j') < (i, j)\}, \lambda) = \begin{cases} \lambda(O_{ij})e^{-\lambda A_{ij}} & O_{ij} \text{ is new} \\ e^{-\lambda A_{ij}} & O_{ij} \text{ is old} \end{cases}$$

The likelihood function is the product of all these terms, or

$$L(\lambda) = e^{-\lambda A} \prod_{(i,j) \in U} \lambda(O_{ij}) \quad (3)$$

where  $U$  is the set of  $(i, j)$  for which  $O_{ij}$  is not a duplicate:  $U = \{(i, j) : O_{ij} \neq O_{i'j'} \text{ for any } (i', j') < (i, j)\}$ .

The likelihood function (3) requires only slight modification to handle some common variants.

1. Instead of searching arbitrarily far for a nearest neighbor, we might search only for neighbors that lie within a prespecified distance  $d$  of the original set of points  $S$  or that lie within a prespecified sampling region. In particular, the sampling region might have an boundary and we might be concerned about edge effects. In that case, instead of finding all  $k$  nearest neighbors of  $s_i$ , we may find fewer than  $k$  and some of the ones we find may be more distant than neighbors beyond the search boundary. Then (3) requires no modification — i.e. there are no edge effects — except to note that  $\lambda_A$  refers to the integral of  $\lambda$  over the area searched, which may be less than the area required to find all  $k$  neighbors of every  $s_i$ . Accuracy might be impaired near the boundary because we search less area and find fewer neighbors, but the method requires no adjustment.
2. We might find a chain of nearest neighbors, as in some types of adaptive sampling for rare species (Seber and Salehi, 2012). That is, we begin with either a fixed or random location  $s$ , then find its nearest neighbor  $O_1$ , then find  $O_1$ 's nearest neighbor  $O_2$ , and so on, for  $k$  steps. Because a Poisson process has independent increments, (3) requires no modification.

Two additional points are worth noting.

1. Eq. (3) is the same likelihood that would have been obtained had we decided in advance to sample region  $A$ . It is irrelevant for  $L(\lambda)$ , and hence for inferences in accord with the likelihood principle, whether  $A$  was fixed *a priori* by the experimenter or arose randomly as a result of  $k$ -tree sampling.
2. Though (3) was derived for an inhomogeneous Poisson process, it also applies to the homogeneous

Symbol	Meaning
$S_i$	one of the validation sets, for $i \in 1, \dots, v$
$S_{-i}$	$S \setminus S_i$ : the elements of $S$ not in $S_i$
$O_{S_i}$	the locations of the $k$ -nearest objects to the points in $S_i$
$A^{(S_i)}$	the region searched under $k$ -tree sampling with $S_i$ as the set of predetermined points
$A^{S_i}$	$A^{(S_i)} \setminus A^{(S_{-i})}$
$U_{S_i}$	indices of the unique elements of $O_{S_i} \setminus O_{S_{-i}}$
$\hat{\lambda}_{\sigma^2}^{S_{-i}}$	an estimate of $\lambda$ based on $O_{S_{-i}}$ with tuning parameter $\sigma^2$

Table 1: Notation for  $v$ -fold cross-validation

Poisson process, should we choose to adopt that model. In the homogeneous case two simplifications,

$$\lambda_A = \lambda|A| \quad \text{and} \quad \prod_{(i,j) \in U} \lambda(O_{ij}) = \lambda^t,$$

where  $|A|$  is the area of  $A$ ,  $t$  is the total number of objects found, and  $\lambda$  is the (scalar) rate, lead to the usual formula  $L(\lambda) = \lambda^t e^{-\lambda|A|}$ .

## 4 Estimating the Intensity Function $\lambda$

Eq. (3) is maximized when  $\lambda(O_{ij}) \rightarrow \infty$  at the  $O_{ij}$ ;  $\lambda(x, y) \rightarrow 0$  elsewhere within  $A$ ; and  $\lambda(x, y)$  is arbitrary outside of  $A$ . Such an inference would be implausible in most applications, so regularization is desirable. There are many means to regularization, such as maximizing (3) subject to constraints or imposing a model on  $\lambda$  such as a thin-plate spline or a polynomial of specified degree, but we use kernel estimation guided by cross-validation. Also see Augustin et al. (2009). For an arbitrary location  $(x', y')$  we define

$$\hat{\lambda}(x', y') = \frac{\sum_{(i,j) \in U} K_{(x',y'),\sigma^2}(x_{ij}, y_{ij})}{\int_A K_{(x',y'),\sigma^2}(x, y) dx dy} \quad (4)$$

where  $K$  is a kernel. In what follows,  $K_{(x',y'),\sigma^2}$  is the bivariate Gaussian density with mean  $(x', y')$  and covariance matrix  $\Sigma = \sigma^2 I_2$ .

The estimator (4) depends on the tuning parameter  $\sigma^2$ , which we determine by  $v$ -fold cross validation. The set of predetermined sampling sites  $S$  is partitioned into validation sets  $S_1, S_2, \dots, S_v$ . We use the notation in Table 1, illustrated in Figure 4.

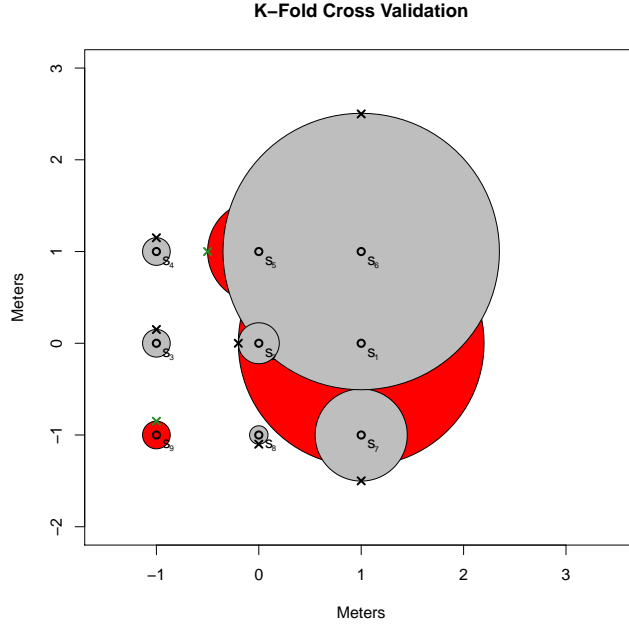


Figure 4:  $S = s_1, \dots, s_9$  is divided into 3 sets with  $S_1 = \{s_1, s_5, s_9\}$ . The red region is  $A^{S_1}$ . The gray region is  $A^{(S_{-1})}$ . The green  $\times$ 's at  $(-0.5, 1)$  and  $(-1, -0.85)$  mark  $U_{S_1}$ . The  $\times$  at  $(-0.2, 0)$  is in  $O_{S_1}$  but not in  $U_{S_1}$ .

In cross-validation, we take each of the  $S_i$  in turn and calculate  $p(O_{S_i} | \hat{\lambda}_{\sigma^2}^{S_{-i}})$ . As in (3),

$$p(O_{S_i} | \hat{\lambda}_{\sigma^2}^{S_{-i}}) = e^{-\hat{\lambda}_{A^{S_i}}} \prod_{(j,k) \in U_{S_i}} \hat{\lambda}(O_{jk}) \quad (5)$$

where we have omitted some super- and subscripts on  $\hat{\lambda}$  for, we hope, clarity. When dealing with validation set  $S_i$  we have already found the nearest objects  $O_{S_{-i}}$  to  $S_{-i}$  and have already searched the region  $A^{(S_{-i})}$  to find them. The region we newly search to find  $O_{S_i}$  is  $A^{S_i} \equiv A^{(S_i)} \setminus A^{(S_{-i})}$ . Therefore, in (5),  $\hat{\lambda}_{A^{S_i}}$  refers to the integral of  $\hat{\lambda}_{\sigma^2}^{S_{-i}}$  over the region  $A^{S_i}$  which, as noted in Table 1, excludes  $A^{(S_{-i})}$ . Similarly,  $U_{S_i}$  is the set of indices  $(i, j)$  for which  $O_{ij}$  is not a duplicate, either of an earlier  $O_{i'j'}$  (as in (3)) or of one of the  $O_{S_{-i}}$ . The tuning parameter  $\sigma^2$  is chosen to maximize  $\prod_{i=1}^v p(O_{S_i} | \hat{\lambda}_{\sigma^2}^{S_{-i}})$ .

As our simulations in Section 5 demonstrate, our estimator is biased. Point-wise bias and MSE can be arbitrarily large. To see why, consider a sequence of intensity functions

$$\lambda_{r,w}(x, y) = \begin{cases} w & x^2 + y^2 < r \\ 0 & x^2 + y^2 \geq r \end{cases}$$

and let  $w \rightarrow \infty$  and  $r \rightarrow 0$  in such a way that  $\pi r^2 w \rightarrow 0$ . Then as  $r$  gets small and  $w$  gets large we expect to see very close to 0 objects in the entire plane and point-wise estimates  $\hat{\lambda}(x, y)$  will converge to 0 everywhere, including at the origin where the true  $\lambda(0, 0)$  is  $w$ , so the bias at  $(0, 0)$  will be close to  $-w$ . The reverse, with  $\lambda(x, y) = 0$  inside a small circle and  $\lambda(x, y) = w$  outside the circle results in a bias of positive  $w$ .

In addition to point-wise estimates, it is often desired to estimate the total number of objects in a region of interest. That estimate, too, can be arbitrarily bad. For example, let

$$\lambda_{r,w}(x, y) = \begin{cases} w & x^2 + y^2 < r \\ c & x^2 + y^2 \geq r \end{cases}$$

for some constant  $c$  and some small radius  $r$ . Let  $|A|$  be the area of the region of interest and let  $w \rightarrow \infty$ . Suppose  $S$  is on the lattice of odd integers, so the  $s_i$ 's closest to the origin are  $(-1, -1)$ ,  $(-1, 1)$ ,  $(1, -1)$ , and  $(1, 1)$ . If  $c$  is reasonably large compared to  $k$ , say  $c \approx 20$  and  $k = 1$ , then the nearest neighbors of  $S$  are very likely to be outside the circle of radius  $r$  in which the intensity is high. Therefore we will sample only that region of the plane where  $\lambda(x, y) = c$  and we will estimate the total number of objects in the area of interest as about  $c|A|$  when the true value is on the order of  $c|A| + w\pi r^2$ . For arbitrarily large  $w$ , our estimate is, with high probability, arbitrarily bad.

Our examples of large bias and poor MSE for point-wise and areal estimates of  $\lambda$  rely on bizarre intensity functions. For estimating bias and MSE in practical settings we recommend simulations with more realistic versions of  $\lambda$ , as illustrated in Section 5.

## 5 Simulation Study

We simulated artificial datasets with four different types of  $\lambda$ 's on the square  $[-1, 12] \times [-1, 12]$ : *a*) homogeneous with  $\lambda = 0.5$ ; *b*) homogeneous with  $\lambda = 4.0$ ; *c*) inhomogeneous with  $\lambda(x, y) = 100 \times \text{the } N_2 \left( \begin{pmatrix} 5 \\ 5 \end{pmatrix}, \begin{pmatrix} 3 & .5\sqrt{6} \\ .5\sqrt{6} & 2 \end{pmatrix} \right)$  density evaluated at  $(x, y)$ ; and *d*) inhomogeneous with  $\lambda(x, y) = 0.2$  for  $x < 6$  and  $\lambda(x, y) = 4$  for  $x > 6$ . This last inhomogeneous case yields a ‘‘cliff’’ population with objects split sharply into two regions—one with high spatial intensity and one with low spatial intensity. Using the `rpoispp` function from the `spatstat` package in R, we simulated 100 populations with each type of  $\lambda$  and analyzed them with 1-tree sampling using prespecified points on the integer lattice  $S = \{1, \dots, 10\} \times \{1, \dots, 10\}$ . We never searched beyond the boundary of a supposed region of interest  $[-1, 12] \times [-1, 12]$ . For each of the four hundred simulated populations we estimated  $\lambda$  at three points near the corner of the region, three points near the midpoint of a

side, and three points near the center. Results are in Tables 2–5 and Figures 5–7. Two points are especially worth noting.

1. Estimates of  $\lambda$  tend to have positive bias. The bias is most easily seen in the homogeneous case where, for both  $\lambda = 0.5$  and  $\lambda = 4.0$ , the mean  $\hat{\lambda}$  was greater than the true  $\lambda$  for all 9 points. The bias can be understood heuristically by considering the one-dimensional case. Let  $X$  be a homogeneous Poisson process on  $\mathbb{R}^+$  with rate  $\lambda$  and  $Y_1$  be the time of the first arrival.  $Y_1$  is also the random variable that would be observed in 1-tree sampling with predetermined point  $S = s_1 = 0$ . It is well-known that  $Y_1 \sim \text{Exp}(\lambda)$ ; the MLE is  $\hat{\lambda} = y_1^{-1}$ ; and that  $E[\hat{\lambda}] = \infty$ . Similarly, if  $Y = Y_1 + \dots + Y_n$  is the time to record  $n$  arrivals, then the MLE is  $\hat{\lambda} = 1/\bar{y}$ , which is positively biased. See also Diggle (1975, 1977).
2. Like all kernel estimators, this one smooths out peaks and troughs. The smoothing is easily seen in the bivariate Normal population. Near the center of the plot, the true intensity is approximately 7 but the estimated intensity is approximately 5. At the corner of the plot, the true intensity is approximately  $10^{-6}$  or  $10^{-7}$  but the estimated intensity is two orders of magnitude larger. Similarly for the cliff population; the cliff at  $x = 6$  is smoothed into a gradual descent from  $x \approx 8$  to  $x \approx 4$ .

$(x, y)$	$\lambda(x, y)$	$\overline{\hat{\lambda}(x, y)}$	RMSE	Bias	Relative RMSE	Relative Bias
(1, 10)	.5	.683	.595	0.183	1.190	.366
(1.5, 10)	.5	.642	.523	0.142	1.046	.284
(1.5, 9.5)	.5	.591	.457	0.091	0.914	.181
(1, 5)	.5	.597	.414	0.097	0.828	.193
(1, 4.5)	.5	.619	.446	0.119	0.892	.237
(1.5, 4.5)	.5	.599	.394	0.099	0.788	.199
(5.5, 5.5)	.5	.540	.344	0.040	0.688	.079
(6, 5.5)	.5	.525	.344	0.025	0.688	.050
(6, 6)	.5	.513	.311	0.013	0.622	.027

Table 2: Simulation results for nine points with a homogeneous  $\lambda = .5$  intensity, where  $\overline{\hat{\lambda}(x, y)}$  is the average value of  $\hat{\lambda}(x, y)$ , RMSE is Root Mean Squared Error, Relative RMSE = RMSE/ $\lambda(x, y)$ , and Relative Bias = Bias/ $\lambda(x, y)$ .

Point $(x, y)$	$\lambda(x, y)$	$\overline{\hat{\lambda}(x, y)}$	RMSE	Bias	Relative RMSE	Relative Bias
(1, 10)	4	5.529	3.964	1.529	0.991	.382
(1.5, 10)	4	5.151	2.995	1.151	0.749	.288
(1.5, 9.5)	4	5.0581	2.452	1.058	0.613	.265
(1, 5)	4	4.550	1.821	.550	0.455	.137
(1, 4.5)	4	4.500	1.878	.500	0.469	.125
(1.5, 4.5)	4	4.354	1.456	.354	0.364	.089
(5.5, 5.5)	4	4.197	1.289	.197	0.322	.049
(6, 5.5)	4	4.127	1.235	.127	0.309	.032
(6, 6)	4	4.177	1.363	.177	0.341	.044

Table 3: Simulation results for nine points with a homogeneous  $\lambda = 4$  intensity, where  $\overline{\hat{\lambda}(x, y)}$  is the average value of  $\hat{\lambda}(x, y)$ , RMSE is Root Mean Squared Error, Relative RMSE = RMSE/ $\lambda(x, y)$ , and Relative Bias = Bias/ $\lambda(x, y)$ .

Point $(x, y)$	$\lambda(x, y)$	$\overline{\hat{\lambda}(x, y)}$	RMSE	Bias	Relative RMSE	Relative Bias
(1, 10)	2.228e-07	2.231e-05	1.086e-04	2.208e-05	487.557	99.112
(1.5, 10)	1.012e-06	6.430e-05	2.660e-04	6.329e-05	262.872	62.511
(1.5, 9.5)	7.940e-06	3.894e-04	1.480e-03	3.814e-04	186.338	48.036
(1, 5)	.214	.427	.371	.212	1.736	.990
(1, 4.5)	.340	.595	.484	.255	1.423	.751
(1.5, 4.5)	.731	.851	.446	.120	0.610	.165
(5.5, 5.5)	6.989	5.092	2.542	-1.897	0.364	-.271
(6, 5.5)	6.333	4.617	2.314	-1.716	0.365	-.271
(6, 6)	5.651	3.939	2.132	-1.713	0.377	-.303

Table 4: Simulation results for nine points with a Normally shaped intensity, where  $\overline{\hat{\lambda}(x, y)}$  is the average value of  $\hat{\lambda}(x, y)$ , RMSE is Root Mean Squared Error, Relative RMSE = RMSE/ $\lambda(x, y)$ , and Relative Bias = Bias/ $\lambda(x, y)$ .



(a) Contour of the mean  $\hat{\lambda}$  when  $\lambda = 0.5$ .

(b) Contour of the mean  $\hat{\lambda}$  when  $\lambda = 4$ .

Figure 5: The estimated contour plots of the homogeneous populations.

Point $(x, y)$	$\lambda(x, y)$	$\overline{\widehat{\lambda}(x, y)}$	RMSE	Bias	Relative RMSE	Relative Bias
(1, 10)	.2	.274	.298	.074	1.492	.371
(1.5, 10)	.2	.242	.219	.042	1.095	.211
(1.5, 9.5)	.2	.240	.202	.040	1.012	.202
(1, 5)	.2	.220	.195	.020	0.975	.099
(1, 4.5)	.2	.224	.200	.024	1.000	.121
(1.5, 4.5)	.2	.210	.176	.010	0.880	.050
(5.5, 5.5)	.2	.713	.561	.513	2.806	2.564
(6, 5.5)	4	1.270	2.749	-2.730	0.687	-.683
(6, 6)	4	1.282	2.737	-2.718	0.684	-.679

Table 5: Simulation results for nine points with a cliff shaped intensity, where  $\overline{\widehat{\lambda}(x, y)}$  is the average value of  $\hat{\lambda}(x, y)$ , RMSE is Root Mean Squared Error, Relative RMSE = RMSE/ $\lambda(x, y)$ , and Relative Bias = Bias/ $\lambda(x, y)$ .

In addition to point-wise estimates of  $\lambda(x, y)$ , we also estimated  $\Lambda \equiv \int_{[-1,12] \times [-1,12]} \lambda(x, y) dx dy$ , or the expected total number of objects within the region of interest. Our estimator is  $\hat{\Lambda} \equiv \int_{[-1,12] \times [-1,12]} \hat{\lambda}(x, y) dx dy$ . That is, we followed the common practice of integrating  $\hat{\lambda}$  over the entire region of interest; we did not integrate only over the unsearched region and add the result to the observed total  $t$  in the searched region. Results are in Table 6. The positive bias shown in Table 6 of  $\hat{\Lambda}$  for the homogeneous populations is consistent with the positive biases in Tables 2 and 3 and Figure 5. Similarly, the negative bias of  $\hat{\Lambda}$  is consistent with the large, negative point-wise biases in Table 4 and Figure 6. The bias for the cliff population comes from

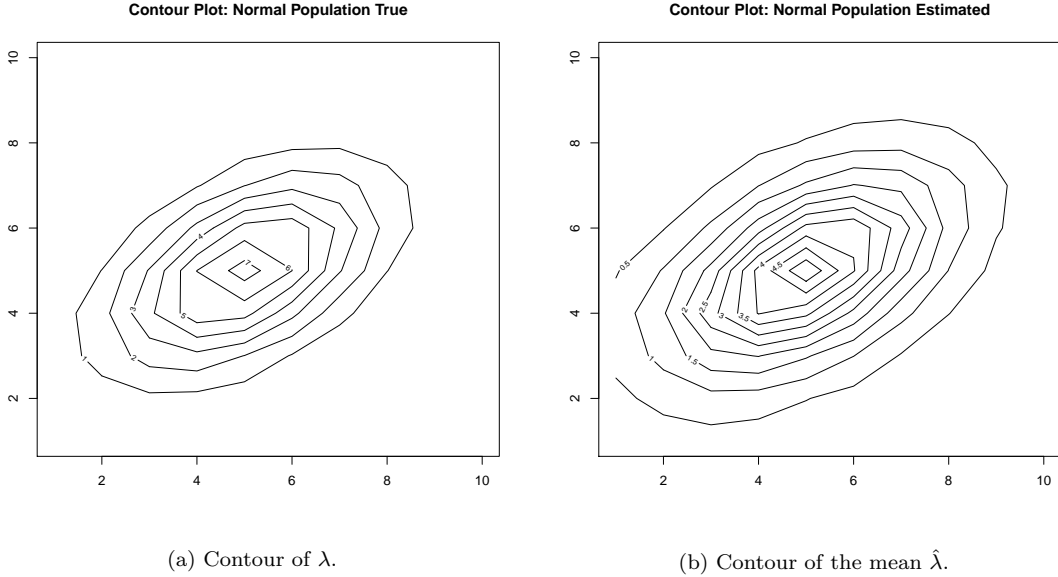


Figure 6: The true and estimated contour plots of the normal population.

positive bias for  $x > 8$  and  $x < 6$  coupled with negative bias for  $x \in (6, 8)$ .

$\lambda$ Type	$\Lambda$	$\bar{\hat{\Lambda}}$	RMSE	Bias	Relative RMSE	Relative Bias
homogeneous $\lambda = 0.5$	84.5	106.748	28.263	22.248	0.334	0.263
homogeneous $\lambda = 4.0$	676	823.504	223.907	147.504	0.331	0.218
cliff	330.2	411.056	171.200	80.856	0.518	0.245
Normal	99.97	82.913	21.715	-17.057	0.217	-0.171

Table 6: Simulation results for expected population totals with 4 types of intensity, where  $\bar{\hat{\Lambda}}$  is the average value of  $\hat{\Lambda}$ , RMSE is Root Mean Squared Error, Relative RMSE = RMSE/ $\Lambda$ , and Relative Bias = Bias/ $\Lambda$ .

Table 7 shows summary statistics of the  $\hat{\sigma}^2$ 's, as determined by cross-validation, that were used to estimate  $\lambda$  and  $\Lambda$  in our simulations.

## 6 Computation

Evaluating expressions like (3) and (5) requires  $\exp(-\lambda_A)$  or  $\exp(-\hat{\lambda}_{A^S_i})$ : the integral of  $\lambda$  or  $\hat{\lambda}$  over a complicated region. Evaluating (4) requires integrating a kernel over the same region. The region is known so in principle the integral can be evaluated analytically. But in practice, the region is sufficiently complicated that we prefer Monte Carlo integration. We distribute  $npts$  points uniformly over the region of interest:  $[-1, 12]^2$  in our simulations. For each  $(x, y)$ , it is easy to evaluate  $\hat{\lambda}(x, y)$  and to determine whether  $(x, y)$



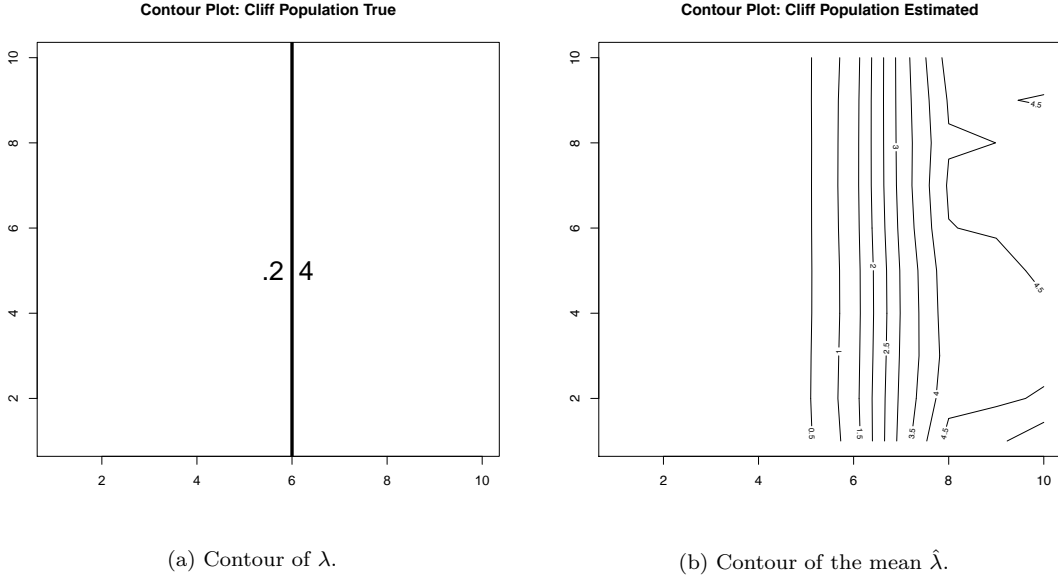


Figure 7: The true and estimated contour plots of the cliff population.

$\lambda$ Type	$\overline{\sigma^2}$	$\text{Var}(\hat{\sigma}^2)$
homogeneous $\lambda = 0.5$	0.536	0.001
homogeneous $\lambda = 4.0$	1.571	1.038
cliff	0.691	0.026
Normal	0.555	0.002

Table 7: Simulation results: mean and variance of the  $\sigma^2$  determined by cross-validation for 4 types of population intensity

is in  $A^{S_i}$ . To choose  $npts$ , we simulated 1-tree sampling on one region and estimated the area searched,  $A$ , many times by Monte Carlo integration, each time using a different set of  $npts$  points. Table 8 shows the SD and the total computing time, on a computer one of us used at the time, for four values of  $npts$ . Results will vary for different problems on different computers. Based on Table 8, we set  $npts = 10,000$  because the standard deviation of  $\hat{A}$  is satisfactorily low and the time required for computation is reasonable.

Number of Points	Standard Deviation of $\hat{A}$	Total Computing Time (seconds)
1000	5.76	88.35
5000	5.25	244.56
10000	5.19	615.54
15000	5.13	1164.70

Table 8: Standard Deviations of  $\hat{A}$  and Total Computing Time

## 7 Case Studies

The developments presented here were motivated by plotless data collected from pitcher plant (*Sarracenia purpurea*) populations in New England bogs. In each bog, pitcher plants were located by 1-tree sampling on a  $10 \times 10$  meter lattice with 1 meter spacing as the set of sampling sites  $S$ . We illustrate our method of estimating spatial intensity with data from four bogs having different patterns of spatial intensity.

**Twelve-mile Bog** Figure 8a shows the lattice of sample sites, the plants observed and the estimated spatial intensity of plants at Twelve-mile bog. At this site, as in virtually all of the 77 bogs we sampled, fewer than 100 plants were found because some of the plants were nearest neighbors of more than one  $s_i$ . For example, there are no red dots near the points (3,5), (3,6), and (3,7). The estimated intensity shows a corresponding dip in that region, and similarly for the points (7,4), (7,5), (8,1), and (8,2). In contrast, the estimated intensity in the upper right corner rises steeply because plants were found very close to the points at (10,8), (10,9), and (10,10). Except for the upper right corner, the intensity does not vary much and its fluctuations are very smooth. It looks much like the homogeneous bogs we simulated with  $\lambda = 4$ .

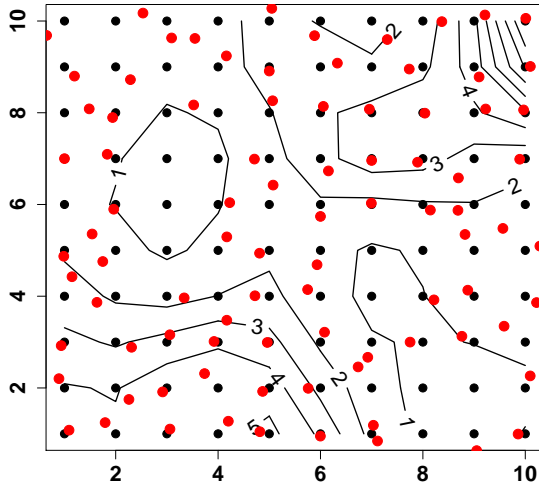
**Errol Brook Bog** This bog, Figure 8b, has a region of high intensity centered around (6,4) and low intensity elsewhere in the sample plot. It looks similar to our simulations with intensity proportional to a bivariate Normal density.

**Branch Pond Bog** The data from this lakeside bog, Figure 8c, appears to follow a cliff distribution. There are no observed plants except in the northwest corner of the grid.

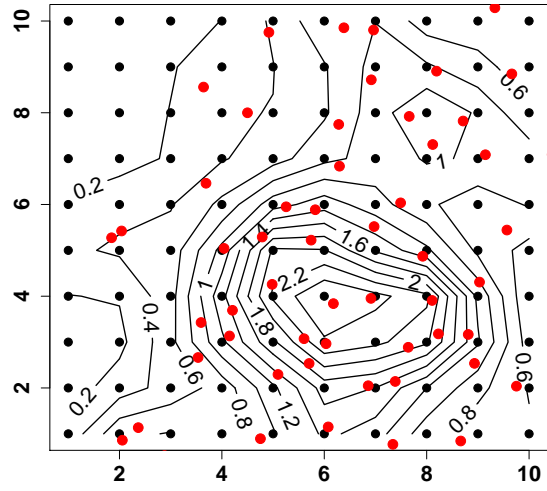
**Number Five Bog** This bog, Figure 8d, had very few *S. purpurea* plants. The kernel algorithm estimated local maxima near the few points where plants were found.

## 8 Discussion

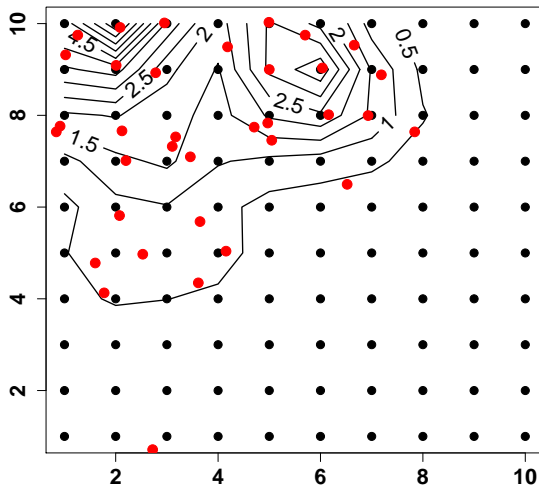
This paper is the first we know of that both derives an explicit likelihood function for  $k$ -tree sampling and addresses the common problem that pre-defined sampling points may share nearest-neighboring objects. Our derivation of the likelihood function under an inhomogeneous Poisson process model provides generality beyond the more commonly considered homogenous Poisson case (Magnussen (2012) and Magnussen et al. (2012) provide notable exceptions) or the nonparametric intensity estimators developed by, e.g., Diggle (1975,



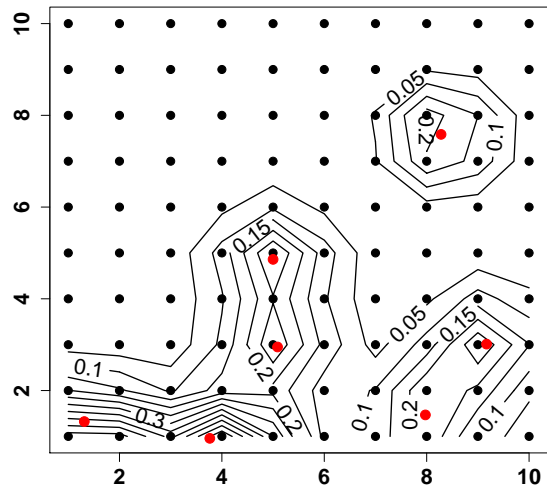
(a) Twelve-mile Bog.  $\hat{\lambda}$  estimated using  $\sigma^2 = 0.67$ .



(b) Errol Brook Bog.  $\hat{\lambda}$  estimated using  $\sigma^2 = 0.56$ .



(c) Branch Pond Bog.  $\hat{\lambda}$  estimated using  $\sigma^2 = 0.75$ .



(d) Number Five Bog.  $\hat{\lambda}$  estimated using  $\sigma^2 = 0.60$ .

Figure 8: Four case studies. Black dots:  $S$ . Red dots:  $O$ . Contours:  $\hat{\lambda}$ .

1977) for Poisson, regular, and clumped distributions. Further, we develop a kernel estimator  $\hat{\lambda}(x, y)$  that acts as a smoother and provides contours of local spatial intensities (cf. Augustin et al., 2009) and that can be integrated for an estimate  $\hat{\Lambda}$  of the total population size in a region.

Our estimators can be biased, even when the true intensity is spatially invariant, for reasons explained in Section 5. The biases can be arbitrarily large, as explained in Section 4. We have not investigated any corrective measures. It is also possible that bias may be affected by whether the sampling points are chosen randomly or deterministically. Ours were chosen deterministically, but our method applies also to random selection, and simulations to calculate bias can incorporate that randomness.

Our study was motivated by a  $k = 1$  sample of pitcher plants and we used  $k = 1$  in our simulations. However, our method applies equally well to  $k > 1$ . We have not explored the value of the information gained by setting  $k$  greater than 1, but that is a direction for future research.

Our estimates use a circularly symmetric kernel: a Normal density with covariance  $\Sigma = \sigma^2 I_2$ . In some situations — if gradients are expected to be parallel to a linear feature such as a stream, a ridge, or a forest boundary, for example — circular symmetry might not be warranted and it would be better to optimize the kernel over all three parameters of the  $N_2$  covariance matrix. We have not explored other types of estimators such as polynomials or splines for  $\lambda$ . Also, we have also not explored  $k$ -tree sampling for inference in other point process models (cf. Augustin et al., 2009) or the implications of imperfect detection, questions which may lead to fruitful research in the future.

# References

- Augustin, N. H., Musio, M., von Wilpert, K., Kublin, E., Wood, S. N., and Schumacher, M. (2009), “Modeling spatiotemporal forest health monitoring data,” *Journal of the American Statistical Association*, 104, 899–911.
- Barbour, M. C., Burk, J. H., Pitts, W. D., Gilliam, F. S., and Schwartz, M. W. (1999), *Terrestrial Plant Ecology*, Benjamin Cummings, Menlo Park, California, 3 edn.
- Byth, K. and Ripley, B. D. (1980), “On sampling spatial patterns by distance methods,” *Biometrics*, 36, 279–284.
- Damgaard, C. (2009), “On the distribution of plant abundance data,” *Ecological Informatics*, 4, 76–82.
- Diggle, P. J. (1975), “Robust density estimation using distance methods,” *Biometrika*, 62, 39–48.
- Diggle, P. J. (1977), “A note on robust density estimation for spatial point patterns,” *Biometrika*, 64, 91–95.
- Ellison, A. M. and Parker, J. N. (2002), “Seed dispersal and seedling establishment of *Sarracenia purpurea* (Sarraceniaceae),” *American Journal of Botany*, 89, 1024–1026.
- Kleinn, C. and Vilčko, F. (2006), “Distance-unbiased estimation for point-to-tree distance sampling,” *Canadian Journal of Forest Research*, 36, 1407–1414.
- Mackenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., and Hines, J. E. (2006), *Occupancy Modeling and Estimation*, Academic Press, San Diego.
- Magnussen, S. (2012), “A new composite  $k$ -tree estimator of stem density,” *European Journal of Forest Research*, 131, 1513–1527.
- Magnussen, S., Fehrmann, L., and Platt, W. J. (2012), “An adaptive composite density estimator for  $k$ -tree sampling,” *European Journal of Forest Research*, 131, 307–320.

- Murdoch, W. W. (1994), "Population regulation in theory and practice," *Ecology*, 75, 271–287.
- Nothdurft, A., Saborowski, J., Nuske, R. S., and Stoyan, D. (2010), "Density estimation based on  $k$ -tree sampling and point pattern reconstruction," *Canadian Journal of Forest Research*, 40, 953–967.
- Patil, S. A., Burnham, K. P., and Kovner, J. L. (1979), "Nonparametric estimation of plant density by the distance method," *Biometrics*, 35, 597–604.
- Pyle, G. H. and Ehrlich, P. R. (2010), "Biological collections and ecological/environmental research: a review, some observations and a look to the future," *Biological Reviews*, 85, 247–266.
- Seber, G. A. F. and Salehi, M. M. (2012), *Adaptive sampling designs: inference for sparse and clustered populations*, Springer, New York.