Ecology, 101(10), 2020, e03143

Distance-based methods for estimating density of nonrandomly distributed populations

GUOCHUN SHEN,[1,2] XIHUA WANG,[1,2,4] AND FANGLIANG HE[1,3]

[1]*ECNU-Alberta Joint Lab for Biodiversity Study, Zhejiang Tiantong Forest Ecosystem National Observation and Research Station, School of Ecology and Environmental Science, East China Normal University, Shanghai 200241 China*
[2]*Shanghai Institute of Pollution Control and Ecological Security, 1515 North Zhongshan Road (No. 2), Shanghai 200092 China*
[3]*Department of Renewable Resources, University of Alberta, Edmonton, Alberta T6G 2H1 Canada*

Citation: Shen, G., X. Wang, and F. He. 2020. Distance-based methods for estimating density of nonrandomly distributed populations. Ecology 101(10):e03143. 10.1002/ecy.3143

Abstract. Population density is the most basic ecological parameter for understanding population dynamics and biological conservation. Distance-based methods (or plotless methods) are considered as a more efficient but less robust approach than quadrat-based counting methods in estimating plant population density. The low robustness of distance-based methods mainly arises from the oversimplistic assumption of completely spatially random (CSR) distribution of a population in the conventional distance-based methods for estimating density of non-CSR populations in natural communities. In this study we derived two methods to improve on density estimation for plant populations of non-CSR distribution. The first method modified an existing composite estimator to correct for the long-recognized bias associated with that estimator. The second method was derived from the negative binomial distribution (NBD) that directly deals with aggregation in the distribution of a species. The performance of these estimators was tested and compared against various distance-based estimators by both simulation and empirical data of three large-scale stem-mapped forests. Results showed that the NBD point-to-tree distance estimator has the best and most consistent performance across populations with vastly different spatial distributions. This estimator offers a simple, efficient and robust method for estimating density for empirical populations of plant species.

Key words: distance-based methods; nearest-neighbor distance; negative binomial distribution; plotless method; population density estimator; spatial distribution of species.

INTRODUCTION

Population density, defined as the average number of individuals per unit area, is the most basic ecological parameter for understanding the dynamics of populations, biological resource management, monitoring, and conservation. Although enormous effort has been paid to develop and test methods for estimating population size (Eberhardt 1967, Seber 1982, Greig-Smith 1983, Buckland et al. 1993, Krebs 1998, Mitchell 2007), practical yet robust methods applicable to real populations remain in demand. The predicament is not only found in estimating the size of populations of high mobility, such as avian species; it is equally challenging to estimate the size of sessile populations such as plant species (Bouldin 2008). For plants, there are two major estimation methods. One is based on quadrat counts (Gleason 1920, Mitchell 2007) and the other is based on tree-to-tree (or called event-to-event) or point-to-tree distances (point means an empty location; see Cogbill et al. 2018 for a recent review). The distance-based methods, also called plotless methods, are considered to be more efficient and less labor intensive than quadrat-based methods because they do not require extra effort of establishing quadrats but directly measure distances from randomly selected focal trees or points to their neighboring trees (Cottam and Curtis 1956). Furthermore, the distance-based methods are less affected by terrain variation and easier to implement in the field (Mitchell 2007).

A common assumption for the distance-based methods is that distributions of populations must be completely spatially random (CSR; Engeman et al. 1994). However, individuals of a plant species are often aggregated distributed (Greig-Smith 1983, He et al. 1997, Condit 1998). For such species the estimation of population density by many existing methods is biased, sometimes seriously (Diggle 1975, Buckland et al. 1993). For example, the widely used estimator of Cottam et al. (1953) based on tree-to-tree distances always overestimates the density of aggregated populations, whereas those based on point-to-tree distances underestimate the density. Attempts have been made to address biases for

---

Manuscript received 4 December 2019; revised 13 March 2020; accepted 9 June 2020. Corresponding Editor: Karen C. Abbott.

[4] E-mail: xhwang@des.ecnu.edu.cn

distance-based methods (Diggle 1975, Lewis 1975, Cox and Lewis 1976, Clayton and Cox 1986, Magnussen et al. 2007, Kronenfeld 2009, Magnussen et al. 2012), but the unrealistic assumption of CSR distribution persists and is required at neighborhood scales even for those estimators aiming to deal with non-CSR populations (Morisita 1957). To avoid this assumption, there have been attempts to develop nonparametric estimators using order statistics that ignore the spatial distribution of populations (Patil et al. 1979, Patil et al. 1982, Engeman et al. 1994), but that simultaneously denies the possibility to improve the robustness of estimators using spatial information (Picard and Bar-Hen 2007). To our knowledge no analytical estimators are available to consider the degree of spatial aggregation in estimating density of non-CSR populations explicitly. This study aims to fill in this methodological gap.

Over time, there have been at least 20 distance-based methods being proposed in the literature to improve the performance of density estimation or to reduce sampling effort (Cottam et al. 1953, Cottam and Curtis 1956, Pollard 1971, Besag and Gleaves 1973, Diggle 1975, Byth and Ripley 1980, Byth 1982, Buckland et al. 1993). These conventional methods are developed from various sampling schemes, notably from tree-to-tree or from point-to-tree distances or from different orders of nearest neighbors (e.g., distances to the first, second, and *i*th neighbors). Although many simulation studies have been conducted to compare the performance of these various methods, little consensus has been reached (Engeman et al. 1994, Liu 2001, Bakus et al. 2007, Magnussen et al. 2007, White et al. 2008, Khan et al. 2016). Most of these studies have not fully realized that many of those compared methods are actually related and can be commonly derived from a general form first proposed by Morisita (1957) until the recent work of Cogbill et al. (2018). Those ad hoc comparisons do not offer much statistical guideline for helping choose proper estimators in practice.

In this study, we proposed new density estimators to deal with the non-CSR nature of population distribution. The first is a modified estimator to correct for the bias associated with the original composite estimator (Diggle 1975). We then relaxed the CSR assumption to derive the second estimator that considers a wide range of spatial distributions. This estimator is very general and includes CSR distribution as a special case from which many of the existing distance-based methods are derived. We tested the performance of the estimator and other existing estimators by simulation and empirical tree populations from three forests (one tropical rain forest and two subtropical forests) in which the real size of each tree population is known. The results showed that our non-CSR distribution estimator substantially improved the density estimation over the existing methods. Our estimator offers a simple and useful distance-based method of high accuracy for practical applications.

THE THEORETICAL BACKGROUND OF EXISTING GENERAL DENSITY ESTIMATORS

To estimate unknown population density λ, n focal trees or points are sampled independently with equal probability. Centered on each focal tree/point, q equal angle sectors are divided. Then a distance r_{lij} is measured from the ith $(i=1,2,\ldots,n)$ focal tree/point to its lth $(l=1,2,\ldots)$ closest neighbor in the jth $(j=1,2,\ldots,q)$ sector. This sampling regime results in the total sampled distances $\{r_{lij}: i=1,2,\ldots,n; j=1,2,\ldots,q\}$ and a total sample size nq. For simplicity, we sometimes write r_{lij} by $\{r_m: m=1,2,...,nq\}$. For a special case of the general sampling scheme, indices i and j are changed from 1 to n and from 1 to q, respectively, whereas l is a fixed value. Figure 1 shows two cases of the general sampling scheme where $j = 1$ and $l = 2$ in the first case and $j = 1, 2, 3, 4$ and $l = 1$ in the second case.

Given a population with CSR distribution, the unbiased simple population density estimator is (Morisita 1957: Eq. 15):

$$\hat{\lambda}_s = \hat{\lambda}\frac{nql-1}{nql} = \frac{q(nql-1)}{\pi\sum_{m=1}^{nq} r_m^2}. \qquad (1)$$

A nice feature of this general estimator is that it has the same form whether the sample is tree-to-tree distances (r_{ttot}) or point-to-tree distances (r_{ptot}) for CSR populations. Many other distance-based density estimators are special cases of Eq. 1 (see details in Appendix S1: Table S1).

In an attempt to derive estimators for non-CSR populations, Morisita (1957) assumed that population distribution in a small area (e.g., within a sector of q equal angle-divided sectors for a circle or a rectangle area; see Fig. 1) follows a CSR distribution even though the population in the overall area is aggregated. Following this assumption, he proposed two estimators based on point-to-tree distances, r_{ptot}, for aggregated populations (Morisita 1957: Eqs. 26 and 31):

$$\hat{\lambda}_{m1\_\text{ptot}} = \frac{l-1}{\pi n}\sum_{m=1}^{nq}\frac{1}{r^2_{m\_\text{ptot}}} \qquad (2)$$

$$\hat{\lambda}_{m2\_\text{ptot}} = \frac{lq-1}{\pi n}\sum_{i=1}^{n}\frac{q}{\sum_j^q r^2_{ij\_\text{ptot}}}. \qquad (3)$$

Cogbill et al. (2018) tested and compared the performance of these two estimators against other existing plotless methods currently available and found Eqs. 2 and 3 behave best for non-CSR populations.

NEW DENSITY ESTIMATORS FOR NON-CSR POPULATIONS

The unbiased composite population density estimator

Like any of the existing density estimators, the general estimator (Eq. 1) and Morisita's estimators (Eqs. 2 and

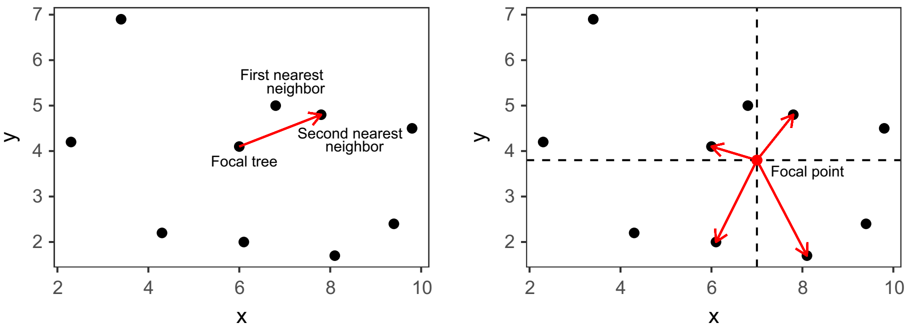

FIG. 1. Illustrating the general distance sampling scheme. On the left panel, distance is sampled from a randomly selected tree (i.e., a tree represented by a black dot) to its second nearest neighbor. On the right panel, distances are sampled from a random point (i.e., an empty location represented by a red dot) to its first nearest-neighbor tree in each of the four equal angle sectors. In the first case, $q = 1$ and $l = 2$; in the second case, $q = 4$ and $l = 1$. [Color figure can be viewed at wileyonlinelibrary.com]

3) may have low performance for non-CSR populations because both were derived from the basic assumption of CSR distribution. We now propose new density estimators to deal with the non-CSR nature of population distribution. The first is a modified estimator to correct for the bias associated with the original composite estimator of Diggle (1975).

To reduce the bias for non-CSR populations, Diggle (1975) proposed a composite estimator by combining tree-to-tree distances (r_{ttot}) and point-to-tree distances (r_{ptot}). That composite population density estimator is defined as a geometric mean of $\hat{\lambda}_{s\_\mathrm{ttot}}$ and $\hat{\lambda}_{s\_\mathrm{ptot}}$: $\sqrt{\hat{\lambda}_{s\_\mathrm{ttot}}\hat{\lambda}_{s\_\mathrm{ptot}}}$, where $\hat{\lambda}_{s\_\mathrm{ttot}}$ and $\hat{\lambda}_{s\_\mathrm{ptot}}$ are density estimators based on tree-to-tree distances and point-to-tree distances defined by Eq. 1, respectively. This general composite estimator is often underestimated for non-CSR population and in fact it is also biased even for CSR populations (see Appendix S1 for a proof).

To correct this bias, we proposed an unbiased general composite population density estimator (see Appendix S1):

$$\hat{\lambda}_c = \frac{q\Gamma(n_1 ql)\Gamma(n_2 ql)}{\pi\Gamma(n_1 ql - 1/2)\Gamma(n_2 ql - 1/2)\sqrt{\sum_{m=1}^{n_1} r^2_{m\_\mathrm{ttot}} \sum_{m=1}^{n_2} r^2_{m\_\mathrm{ptot}}}} \qquad (4)$$

where r_{m_ptot} and r_{m_ttot} are the $\{r_m : m = 1, 2, \ldots, nq\}$ distances from point-to-tree and tree-to-tree, respectively. Γ is a gamma function. n_1 and n_2 are the sample size of tree-to-tree distances and point-to-tree distances, respectively. Note Eq. 4 is the geometric mean of $\hat{\lambda}_{s\_\mathrm{ttot}}$ and $\hat{\lambda}_{s\_\mathrm{ptot}}$ defined by Eq. 1 under the assumption of CSR distribution; thus $\hat{\lambda}_{s\_\mathrm{ttot}}$ and $\hat{\lambda}_{s\_\mathrm{ptot}}$ may overestimate or underestimate the density of highly aggregated populations (Engeman et al. 1994). These opposite biases may cancel out in Eq. 4 and lead to a more robust density estimator. We also generalized Diggle's (1975) arithmetic mean composite estimator in Appendix S1. But it is not interesting enough to be included here, because its performance is theoretically not as good as the above geometric mean composite estimator $\hat{\lambda}_c$ (see Appendix S1).

The general density estimators based on negative binomial distribution

The second estimators we derive in this study directly incorporate non-CSR in the estimators. Here, we relax the assumption of CSR distribution by assuming counts of individuals in each sector of q equal angle sectors follow a negative binomial distribution (NBD). NBD is one of the most widely used models to describe spatial distributions of empirical populations ranging from CSR to aggregated distributions (Bliss and Fisher 1953, Boswell and Patil 1970, Krebs 1998, He and Gaston 2003). Following the NBD, Gao (2013) derives the pdfs of r_{lij_ttot} and r_{lij_ptot} (the distances of a tree and a point to the lth nearest neighbor in the jth equal angle sector, respectively). If we assume the sampled distances from focal trees/points are completely independent, the non-CSR population density estimator for tree-to-tree distances ($\hat{\lambda}_{n\_\mathrm{ttot}}$) and point-to-tree distances ($\hat{\lambda}_{n\_\mathrm{ptot}}$) can be found by the moment estimation method as (Appendix S1):

$$\hat{\lambda}_{n\_\mathrm{ttot}} = \frac{nq^2 l}{\pi\sum_{m=1}^{nq} r^2_{m\_\mathrm{ttot}}} \qquad (5)$$

$$\hat{\lambda}_{n\_\mathrm{ptot}} = \frac{q(2l-1)\sum_{m=1}^{nq} r^{-1}_{m\_\mathrm{ptot}}}{\pi\sum_{m=1}^{nq} r_{m\_\mathrm{ptot}}} - \frac{nq^2 l}{\pi\sum_{m=1}^{nq} r^2_{m\_\mathrm{ptot}}}. \qquad (6)$$

To our knowledge, Eqs. 5 and 6 are the first population density estimators that explicitly consider non-CSR distribution of populations. It is important to note that Eqs. 5 and 6 do not actually require knowledge about the NBD aggregation parameter although the estimators are derived from the NBD assumption. To recognize this connection, we call these estimators as NBD estimators.

Testing Performance of the Estimators

The performance of the five general estimators (Eqs. 2–6) for non-CSR populations was evaluated by both simulated and empirical populations. For simulation, all of the hypothetical populations were generated by a homogeneous Thomas point process (Thomas 1949) in which the density of cluster centers is 0.01 and the mean number of points per cluster is 10. Therefore, the expected number of points is 0.1 per unit area. The degree of aggregation for each simulated population is determined by the aggregation parameter σ in the Thomas process, which is the standard deviation of random displacement of a point from its cluster center. It ranges from 1 to 5.5 in our simulation. For example, when $\sigma = 1$, the spatial distribution of simulated population is highly aggregated (Appendix S2: Fig. S1). As the value of σ increases, the distribution of the simulated populations becomes close to CSR (Appendix S2: Fig. S1).

The empirical data are tree communities in three large-scale fully stem-mapped forest plots. They are a 50-ha (1,000 × 500 m) tropical rain forest plot on Barro Colorado Island, Panama (Hubbell and Foster 1986), a 50-ha (1,000 × 500 m) and a 20-ha (500 × 400 m) subtropical forest plots in Heishiding Nature Reserve (Yin and He 2014) and Tiantongshan Nature Reserve (Yang et al. 2011) in China, respectively. In each plot, the true population density and the exact spatial location of every individual stem with diameter at breast height (DBH) $\geq$1 cm are known. This detailed spatial information allows us to examine the performance of the density estimators for real populations rigorously. In this study, species with at least 5 individuals in each plot were used. The spatial aggregation of each species at neighborhood scale was quantified by the Hopkins and Skellam dispersion index (Hopkins and Skellam 1954). Appendix S2: Fig. S1 showed that there is a large difference in the extent of aggregation among the real tree species distributions.

To assess the performance of the estimators, both accuracy and precision were calculated. Accuracy is expressed as the relative bias $\left(\lambda-\bar{\hat{\lambda}}\right)/\lambda$, where λ is the true density and $\bar{\hat{\lambda}}$ is the averaged estimate of population density. Precision is quantified by the coefficient of variation, $s_{\hat{\lambda}}/\bar{\hat{\lambda}}$, where $s_{\hat{\lambda}}$ is the standard error of the estimate. To combine the information of the accuracy and precision, the relative root-mean-squared error (RRMSE) was used. It is defined as $\text{RRMSE}=\sqrt{\left(\sum_{i=1}^{N}(\lambda-\hat{\lambda})^2\right)/N\lambda^2}$, where N is the number of replications of the density estimation. A good estimator has a low RRMSE value.

The analytical expressions of accuracy, precision and RRMSE for Eqs. 2–6 are not available for non-CSR populations. Thus, for each simulated or real tree population, a mean accuracy, precision, and RRMSE of $\{q=1,2,3,4;l=2,3,4;n=480/q\}$ for an estimator was calculated. Here both point-to-tree distances and tree-to-tree distances in each case $\{r_{lj}:q=1,2,3,4;l=2,3,4\}$ were sampled. The sample size of neighborhood distances equals 480 in each case for abundant species, but equals the total population size for species with abundance less than 480. Relationships between the distance sampling parameters (e.g., l and q) and the ln(RRMSE) of the estimators were quantified by simple linear regression. Sensitivity of each estimator to sample size (np) was evaluated by the coefficient of variation of $\hat{\lambda}$ across sample size.

We also compared our best-performing estimator (that is, Eq. 6) with several other distance-based estimators that were designed to deal with non-CSR populations but could not fit into the general distance sampling framework on which this study is based. These other estimators include the order statistics-based estimator proposed by Persson (1964), the KV estimator of Kleinn and Vilčko (2006), the best adaptive composite estimator proposed by Magnussen et al. (2012), and the Picard estimator (Picard and Bar-Hen 2007). Their performance was compared following the same methods described above, but only carried out at $q = 1$ level because these estimators were only derived under $q = 1$.

All the analyses in this study were implemented by R codes (codes are publicly available; see Data Availability), and the calculations were in R software.[5] The new estimators can be easily calculated by the R package *sce* (also publicly available; see Data Availability).

RESULTS

Performance of the estimators for simulated non-CSR populations

The performances of the five general estimators for simulated populations are shown in the first two rows of Table 1. The RRMSEs of the estimators were negatively correlated with the order of neighbors (l). Thus, increasing the order of neighbors generally increased the performance of the estimators (the first row in Table 1). In contrast, increasing the number of equiangular sectors (q) had a negative effect on the performance of the estimators except the λ_{n_ttot} estimators based on tree-to-tree distances (the second row in Table 1).

For the simulated populations, the NBD estimator based on point-to-tree distances (Eq. 6), $\hat{\lambda}_{n\_\text{ptot}}$, had the highest accuracy and precision among the five general estimators (Fig. 2a, b) and also the lowest sensitivity (Fig. 2c) to the sample size, although the performance of all the estimators decreased with increasing population aggregation. The advantage of the $\hat{\lambda}_{n\_\text{ptot}}$ is especially apparent for highly aggregated populations (e.g., small σ values in Fig. 2). In contrast, the NBD tree-to-tree estimator, $\hat{\lambda}_{n\_\text{ttot}}$, behaved worst (which is expected because the sampled distances from focal trees

---

[5] http://www.r-project.org/.

TABLE 1. Coefficients of linear regression between distance sampling parameters (l orders of neighbors and q numbers of equiangular sectors) and the performance of the five general estimators (Eqs. 2–6), as measured by the negative log-transformed relative root-mean-squared error (RRMSE). The test data are the simulated tree populations and real tree populations from three forest plots.

	Morisita's estimator		Simple composite estimator	NBD estimator	
	λ_{m1_ptot}	λ_{m2_ptot}	λ_c	λ_{n_ptot}	λ_{n_ttot}
Simulated populations					
l	0.024	0.004	0.018	0.034	0.015
q	−0.007	−0.037	−0.098	−0.003	0.084
Barro Colorado Island plot					
l	0.183	0.142	0.020	0.06	0.317
q	−0.117	0.165	−0.289	−0.028	−1.027
Heishiding plot					
l	0.329	0.185	0.218	0.054	0.518
q	−0.218	0.185	−3.18	−0.028	−0.995
Tiantongshan plot					
l	0.586	0.117	0.051	0.067	0.226
q	−0.059	0.143	−0.348	−0.037	−0.18

are not completely independent of each other and thus violates the independent assumption needed for deriving $\hat{\lambda}_{n\_ttot}$).

Estimating density for real tree populations

Consistent with results of simulated populations, performance of the estimators generally increased with the order of neighbors (l) and decreased with the number of sectors (q) in real tree species, except that the performance of $\hat{\lambda}_{m2\_ptot}$ increased with q (Table 1).

The NBD point-to-tree estimator, λ_{n_ptot} performed best among all the estimators for the real tree populations of the three plots. It accurately estimated the abundances of most species in each plot (top panels, Fig. 3). The achieved accuracy can be partly explained by its low sensitivity to spatial aggregation as evidenced by the lowest mean RRMSE for the different aggregated groups of species (bottom panels, Fig. 3). The λ_{n_ptot} also performed better than other existing estimators aiming to deal with non-CSR populations (top panels, Fig. 4), and its superior performance is insensitive to the increased population aggregation (bottom panels, Fig. 4).

DISCUSSION

There has been a long history of interest in searching for effective methods for estimating density of tree populations (Judd 1913). Much of this interest has been paid in developing robust distance-based estimators for real populations of non-SCR distribution (Morisita 1957, Persson 1964, Patil et al. 1979, 1982, Engeman et al. 1994, Kleinn and Vilčko 2006, Picard and Bar-Hen 2007, Magnussen et al. 2012, Cogbill et al. 2018). This effort has proven to be challenging, as real species can be distributed in so many different ways. In this study we proposed two new general estimators by a widely proven observation that distributions of the majority of empirical species can be approximately described by the

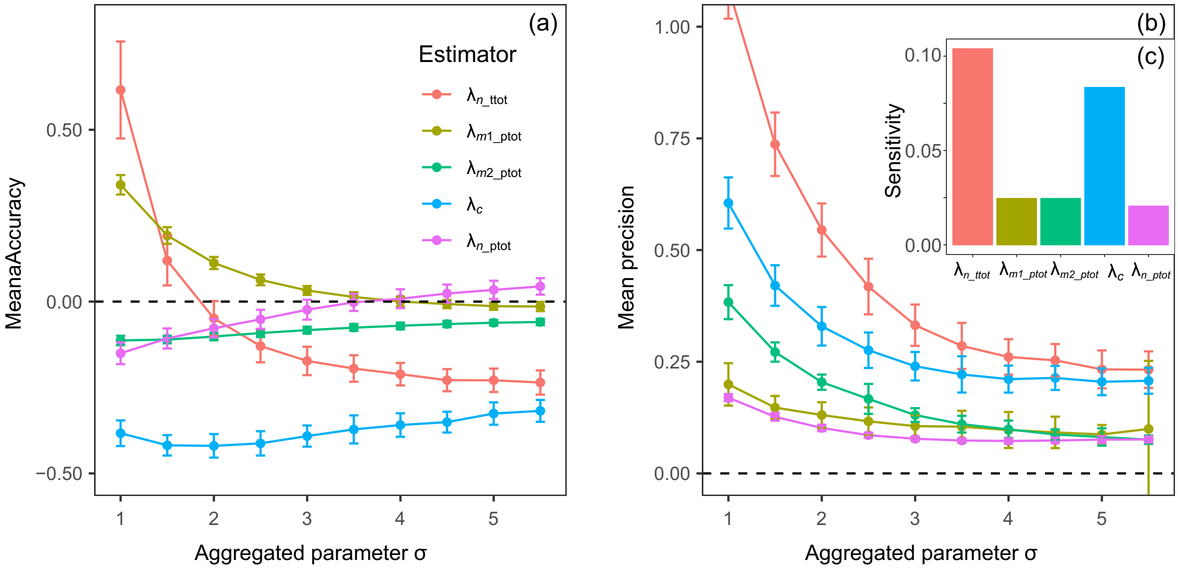

FIG. 2. Performance of the five population density estimators, λ_{m1_ptot}, λ_{m2_ptot}, λ_c, λ_{n_ptot} and λ_{n_ttot} (see Table 1 for the notation), for simulated populations varying from high (small σ) to low (large σ) aggregation. Performance of each estimator was presented by plotting (a) the averaged accuracy, (b) precision against aggregation parameter σ, and (c) sensitivity across sample size over 99 repeated simulated populations. Vertical bar at each solid point is one standard deviation. Note Eqs. 2 and 3 are indefinite at $l = 1$ and $lq = 1$, respectively; the performance shown here was based on distances at least to the second nearest tree ($l \geq 2$). Performance of all estimators, except λ_{m1_ptot}, for distances ($lq \geq 2$) is given in Appendix S4. [Color figure can be viewed at wileyonlinelibrary.com]

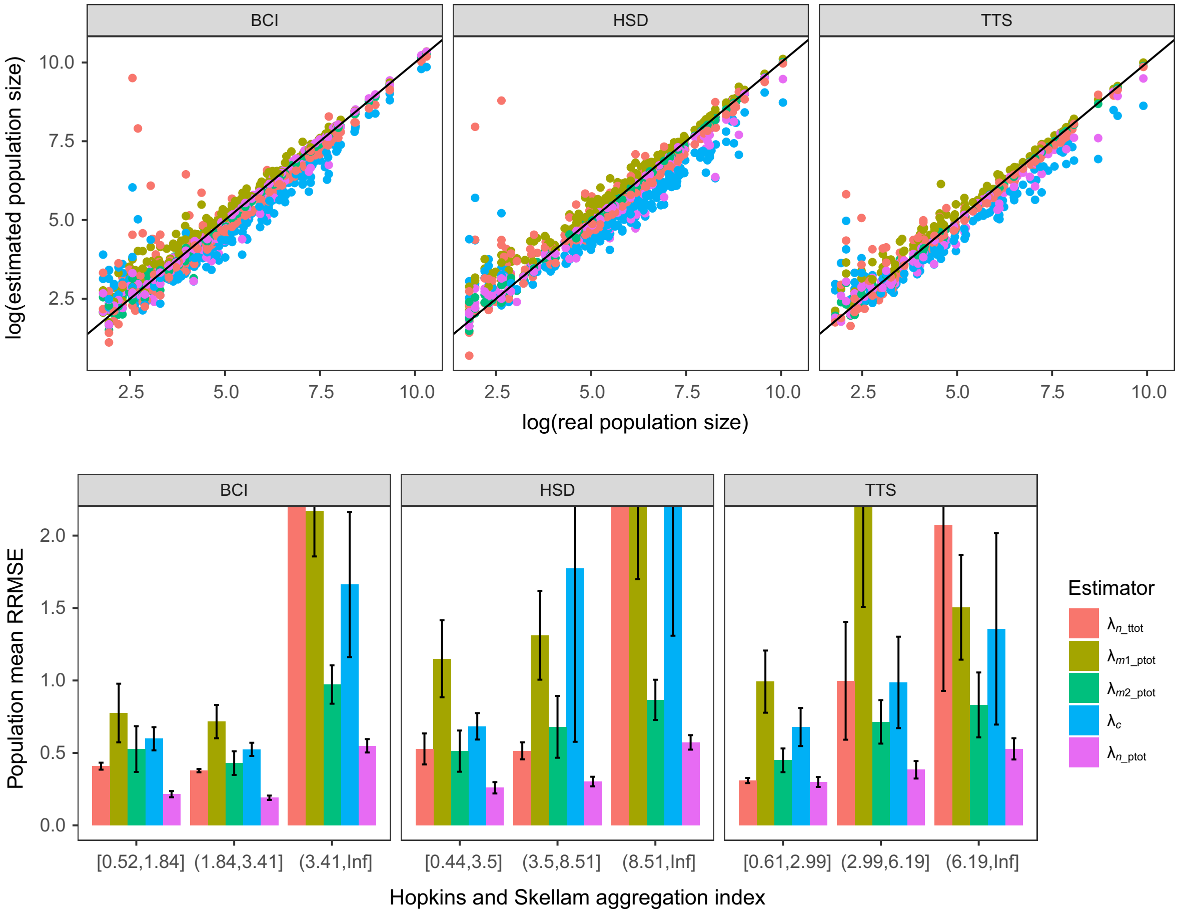

FIG. 3. Performance of the five estimators (λ_{m1_ptot}, λ_{m2_ptot}, λ_c, λ_{n_ptot} and λ_{n_ttot}; see Table 1 for notation) for real tree populations in the Barro Colorado Island (BCI), Heishiding (HSD), and Tiantongshan (TTS) plots. The top-row panels show the real population densities against the estimated population density. The bottom-row panels show the group mean RRMSEs of the estimators with one standard error bars against the groups of the Hopkins and Skellam (1954) aggregation index for the species examined. The aggregation of species distribution increases with the group values from left to right. [Color figure can be viewed at wileyonlinelibrary.com]

negative binomial distribution (Bliss and Fisher 1953, Boswell and Patil 1970, Krebs 1998, He and Gaston 2003, Zillio and He 2010). Our simulation and empirical tests showed that the NBD point-to-tree distance estimator (Eq. 6) has the lowest bias and highest precision. It works best among all the existing estimators.

Our NBD estimators (Eqs. 5 and 6) are the first closed-form estimators that explicitly deal with non-CSR distribution, but they have contrast performance (Fig. 2). All the distance-based estimators require a simple random sample of distances (e.g., r_{lij} in our study). This requirement is met for a CSR population but not so for non-CSR populations. For species that follow NBD, the focal trees that are sampled are of course not independently distributed. As such, the NBD tree-to-tree estimator (Eq. 5) fails to produce useful results. This result is very general and applies to non-CSR populations across a large range of population density (Appendix S3: Fig. S1). Opposite to the tree-to-tree sampling, in the point-to-tree estimator (Eq. 6) the points are randomly and independently determined. It is worth mention that the performance of our NBD estimators does depend on how well real populations follow the NBD distribution. Tree species often aggregate at several spatial scales (Wiegand et al. 2007) and no simple probability distribution, including the NBD, can fully capture such complexity of aggregation. Nonetheless, the NBD, $\hat{\lambda}_{n\text{-}ptot}$, seems to behave more robustly than any other estimator, although there remains a noticeable consistent underestimation (Figs. 3 and 4) for highly aggregated populations.

The performance tests provide us with some practical guidelines for the use of the non-CSR estimators as described below. First, inconsistent with previous findings (e.g., Engeman et al. 1994; but see Cogbill et al. 2018), increasing the number of sectors q will not always increase the performance of distance-based density estimators for non-CSR populations (Table 1), although sampling efficiency may increase if several (e.g., four) distances are sampled from a common focal tree/point.

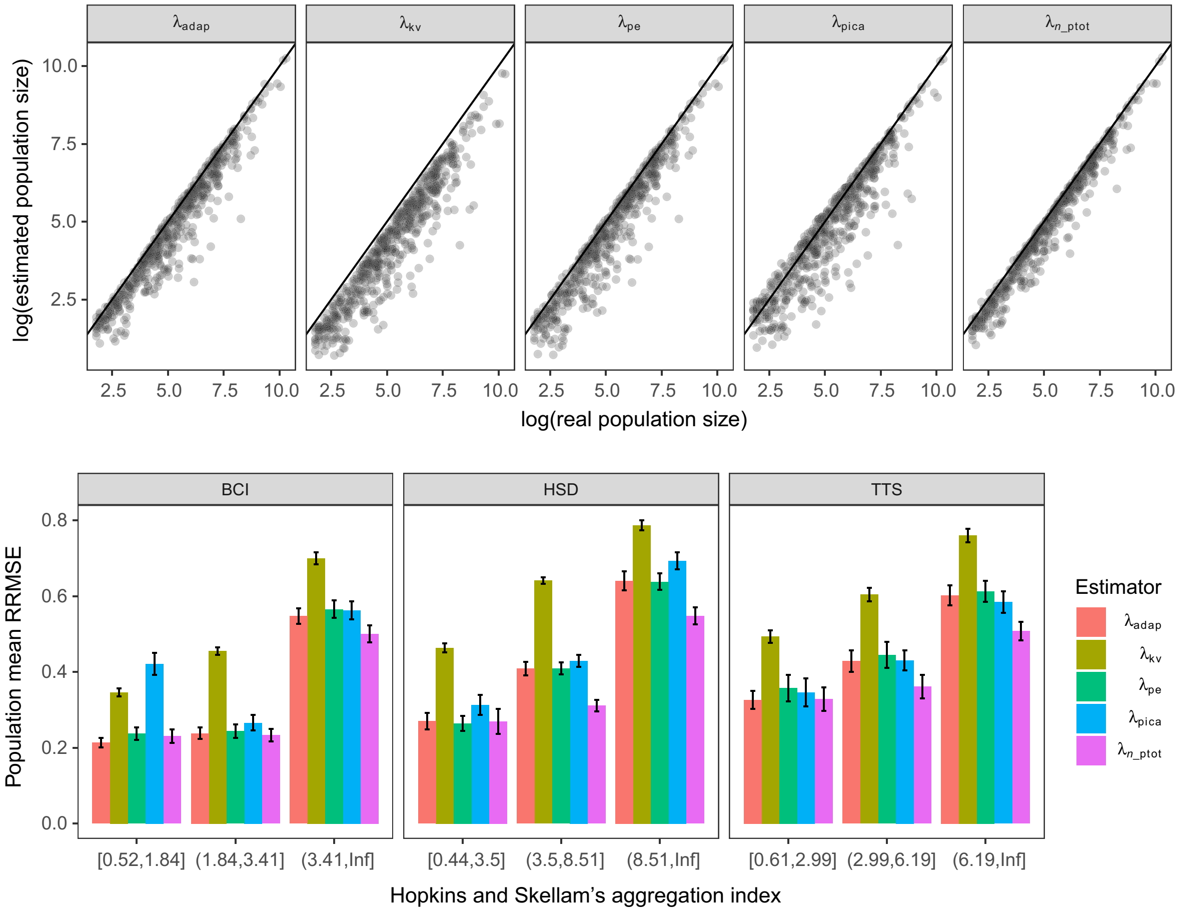

FIG. 4. Performance of our best-performing negative binomial distribution (NBD) estimator (λ_{n_ptot}) and other existing estimators for non-CSR population: the λ_{pe} estimator (Persson 1964), the λ_{kv} estimator (Kleinn and Vilčko 2006), the best λ_{adap} estimator (Magnussen et al. 2012), and the λ_{pica} estimator (Picard and Bar-Hen 2007) for real tree populations in the three forest plots (Barro Colorado Island [BCI], Heishiding [HSD], and Tiantongshan [TTS]. The top-row panels show the real population densities against the estimated population density for the five estimators. The bottom-row panels show the group mean RRMSEs of the estimators with one standard error bar against the groups of the Hopkins and Skellam (1954) aggregation index of the species distribution. [Color figure can be viewed at wileyonlinelibrary.com]

This inconsistency could partially be understood from the variance of λ_{m1_ptot} and λ_{m2_ptot} derived under CSR distribution. If the total distance sample size (nq) is fixed, increasing the number of sectors q will reduce the number of focal points (n). That will decrease the variance of λ_{m2_ptot} ($= \lambda^2/nql - 2n$; see Appendix S1), leading to a positive effect of increasing q on the performance of λ_{m2_ptot}. But that will not change the variance ($\lambda^2/nq(l-2)$; Appendix S1) and thus not the performance of λ_{m1_ptot}, if the total distance sample size (nq) is fixed. A more subtle issue is that increasing q will decrease the number of focal points for a fixed sample size (nq). In this case, distances (to a tree in each of the q sectors) measured from the same focal point may contain less information than the same number of distances measured from multiple focal points. These explain why the performance of our best estimator λ_{n_ptot} as well as λ_{m1_ptot} decreases with increasing q for aggregated populations. For the similar reason as explained by the explicit forms of variance above, increasing the order of neighbors (l) can improve the robustness of most estimators, including our best estimator $\hat{\lambda}_{n\_ptot}$ (Table 1). However, the neighborhood distance increases with the order of neighbors l; for example, the expected neighborhood distance for a CSR population is $\Gamma(l+1/2)\sqrt{q}\left(\Gamma(l)\sqrt{(\pi\lambda)}\right)^{-1}$ (see Appendix S1). As such, although the performance of $\hat{\lambda}_{n\_ptot}$ may be enhanced by using higher orders of l, it does increase the labor in the field in searching for and measuring distances of higher-order neighbors.

The second practical concern is the minimum sample size of distances required to ensure a robust estimation of density for a population. As is shown in Appendix S1, for CSR populations, the sample size nq should be larger than $428.2/l$ and $1623/l$, if one wants to respectively achieve 90% and 95% confidence intervals for density estimate $\hat{\lambda}$. For example, if distances are measured from focal trees to their second nearest neighbors ($l = 2$) in each of four equal angle sectors ($q = 4$), then 54 and 203 focal trees are needed to ensure 90% and 95% accuracy. For aggregated populations, it is very difficult to make a general suggestion on sample size. Because the RRMSE for the non-CSR estimators does not only depend on the sampling scheme (e.g., different combinations of l and n), but also relates to the specific spatial distribution of the population, thus the minimum distance sample size varies with different sampling schemes and the different distributions of populations. Simulation results (Appendix S1: Fig. S1) show that the minimum sample size of our best estimator, $\hat{\lambda}_{n\_\mathrm{ptot}}$, should be larger for highly aggregated (small σ) than less aggregated (large σ) species. If one wants to reach 90% accuracy for highly aggregated species ($\sigma \leq 3$), around 300 distances should be sampled from random points to their forth nearest neighbor ($l = 4$).

To summarize, in this study we developed two distance-based methods for estimating population density for non-CSR species. The first is a modified composite estimator to correct for the bias associated with that estimator, although independent sampling of tree-to-tree distances for non-CSR trees in the field could be challenging. The second method was derived from the negative binomial distribution to directly deal with non-CSR distribution of species. We showed, by both simulation and empirical data, that the NBD point-to-tree distance estimator has the best performance (most robust and least bias) over all existing non-CSR distance methods. This study offers a practically useful method for estimating density of empirical plant populations.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) (31870404) and the National Key Research and Development Program (NKRDP) (2016YFC0503102) to GCS, NSFC (31210103920) to XHW, the Natural Sciences and Engineering Research Council of Canada (NSERC Canada) to FH and ECNU Multifunctional Platform for Innovation (008). We thank the CTFS-ForestGEO network for making the Barro Colorado Island forest dynamics plot available for use.

Literature Cited

Bakus, G. J., G. Nishiyama, E. Hajdu, H. Mehta, M. Mohammad, U. D. S. Pinheiro, and E. Hanan. 2007. A comparison of some population density sampling techniques for biodiversity, conservation, and environmental impact studies. Biodiversity and Conservation 16:2445–2455.

Besag, J. E., and J. T. Gleaves. 1973. On the detection of spatial pattern in plant communities. Bulletin of the International Statistical Institute 45:153–158.

Bliss, C. I., and R. A. Fisher. 1953. Fitting the negative binomial distribution to biological data. Biometrics 9:176–200.

Boswell, M. T., and G. P. Patil. 1970. Chance mechanisms generating the negative binomial distribution. Pages 3–22 *in* G. P. Patil, editor. Random counts in models and structures. Pennsylvania State University Press, University Park, Pennsylvania, USA.

Bouldin, J. 2008. Some problems and solutions in density estimation from bearing tree data: A review and synthesis. Journal of Biogeography 35:2000–2011.

Buckland, S. T., D. R. Anderson, K. P. Burnham, and J. L. Laake. 1993. Distance sampling: estimating abundance of biological populations. Springer, Dordrecht, The Netherlands.

Byth, K. 1982. On robust distance-based intensity estimators. Biometrics 38:127–135.

Byth, K., and B. Ripley. 1980. On sampling spatial patterns by distance methods. Biometrics 36:279–284.

Clayton, G., and T. F. Cox. 1986. Some robust density estimators for spatial point processes. Biometrics 42:753–767.

Cogbill, C. V., A. L. Thurman, J. W. Williams, J. Zhu, D. J. Mladenoff, and S. J. Goring. 2018. A retrospective on the accuracy and precision of plotless forest density estimators in ecological studies. Ecosphere 9:e02187.

Condit, R. 1998. Tropical forest census plots: methods and results from Barro Colorado Island, Panama and a comparison with other plots. Springer, Berlin, Germany.

Cottam, G., and J. T. Curtis. 1956. The use of distance measures in phytosociological sampling. Ecology 37:451–460.

Cottam, G., J. T. Curtis, and B. W. Hale. 1953. Some sampling characteristics of a population of randomly dispersed individuals. Ecology 34:741–757.

Cox, T. F., and T. Lewis. 1976. A conditioned distance ratio method for analyzing spatial patterns. Biometrika 63:483–491.

Diggle, P. J. 1975. Robust density estimation using distance methods. Biometrika 62:39–48.

Eberhardt, L. 1967. Some developments in ’distance sampling’. Biometrics 23:207–216.

Engeman, R. M., R. T. Sugihara, L. F. Pank, and W. E. Dusenberry. 1994. A comparison of plotless density estimators using Monte Carlo simulation. Ecology 75:1769–1779.

Gao, M. 2013. Detecting spatial aggregation from distance sampling: a probability distribution model of nearest neighbor distance. Ecological Research 28:397–405.

Gleason, H. A. 1920. Some applications of the quadrat method. Bulletin of the Torrey Botanical Club 47:21–33.

Greig-Smith, P. 1983. Quantitative plant ecology. Third edition. Blackwell, Oxford, UK.

He, F., and K. J. Gaston. 2003. Occupancy, spatial variance, and the abundance of species. American Naturalist 162:366–375.

He, F., P. Legendre, and J. V. LaFrankie. 1997. Distribution patterns of tree species in a Malaysian tropical rain forest. Journal of Vegetation Science 8:105–114.

Hopkins, B., and J. G. Skellam. 1954. A new method for determining the type of distribution of plant individuals. Annals of Botany 18:213–227.

Hubbell, S. P., and R. B. Foster. 1986. Commonness and rarity in a neotropical rain forest: implications for tropical tree conservation. Pages 205–231 *in* M. E. Soule, editor. Conservation biology: the science of scarcity and diversity. Sinauer Associates Inc., Sunderland, Massachusetts, USA.

Judd, C. S. 1913. A short-cut method of cruising timber. Forestry Quarterly 11:405–421.

Khan, M. N. I. et al. 2016. An evaluation of the plant density estimator the point-centred quarter method (PCQM) using monte carlo simulation. PLoS One 11:e0157985.

Kleinn, C., and F. Vilčko. 2006. A new empirical approach for estimation in k-tree sampling. Forest Ecology and Management 237:522–533.

Krebs, C. J. 1998. Ecological methodology. Second edition. Addison-Wesley Educational Publishers Inc, Boston, Massachusetts, USA.

Kronenfeld, B. J. 2009. A plotless density estimator based on the asymptotic limit of ordered distance estimation values. Forest Sciences 55:283–292.

Lewis, S. 1975. Robust estimation of density for a two-dimensional point process. Biometrika 62:519–521.

Liu, C. 2001. A comparison of five distance-based methods for spatial pattern analysis. Journal of Vegetation Science 12:411–416.

Magnussen, S., L. Fehrman, and W. J. Platt. 2012. An adaptive composite density estimator for k-tree sampling. European Journal of Forest Research 131:307–320.

Magnussen, S., C. Kleinn, and N. Picard. 2007. Two new density estimators for distance sampling. European Journal of Forest Research 127:213–224.

Mitchell, K. 2007. Quantitative analysis by the point-centered quarter method. Hobart and William Smith Colleges, Geneva, New York, USA.

Morisita, M. 1957. A new method for the estimation of density by spacing method applicable to nonrandomly distributed populations. Seiri Seitai 7:134–144.

Patil, S. A., K. P. Burnham, and J. L. Kovner. 1979. Nonparametric estimation of plant density by the distance method. Biometrics 35:597–604.

Patil, S. A., J. L. Kovner, and K. P. Burnham. 1982. Optimum nonparametric estimation of population density based on ordered distances. Biometrics 38:243–248.

Persson, O. 1964. Distance methods: the use of distance measurements in the estimation of seeding density and open space frequency. Skogshögskolan, Studia Forestalia Suecica 15:68.

Picard, N., and A. Bar-Hen. 2007. Estimation of the density of a clustered point pattern using a distance method. Environmental and Ecological Statistics 14:341–353.

Pollard, J. H. 1971. On distance estimators of density in randomly distributed forests. Biometrics 27:991–1002.

Seber, G. A. F. 1982. The estimation of animal abundance and related parameters. Second edition. Edward Arnold, London, UK.

Thomas, M. 1949. A generalization of Poisson’s binomial limit for use in ecology. Biometrika 36:18–25.

White, N. A., R. M. Engeman, R. T. Sugihara, and H. W. Krupa. 2008. A comparison of plotless density estimators using Monte Carlo simulation on totally enumerated field data sets. BMC Ecology 8:6.

Wiegand, T., S. Gunatilleke, N. Gunatilleke, and T. Okuda. 2007. Analyzing the spatial structure of a Sri Lankan tree species with multiple scales of clustering. Ecology 88:3088–3102.

Yang, Q. S., Z. P. Ma, Y. B. Xie, Z. G. Zhang, Z. H. Wang, H. M. Liu, and X. H. Wang. 2011. Community structure and species composition of an evergreen broad leaved forest in Tiantong’s 20 ha dynamic plot, Zhejiang province, eastern China. Biodiversity Science 19:215–223.

Yin, D. Y., and F. He. 2014. A simple method for estimating species abundance from occurrence maps. Methods in Ecology and Evolution 5:336–343.

Zillio, T., and F. He. 2010. Modeling spatial aggregation of finite populations. Ecology 91:3698–3706.

Supporting Information

Additional supporting information may be found in the online version of this article at http://onlinelibrary.wiley.com/doi/10.1002/ecy.3143/suppinfo

Data Availability Statement

Associated R codes are available on Zenodo. https://doi.org/10.5281/zenodo.3887317 and https://doi.org/10.5281/zenodo.3887311.