NRC Research Press

ARTICLE

Robust fixed-count density estimation with virtual plots

Steen Magnussen

Abstract: Fixed-count sampling (plotless) remains attractive for forest inventories in difficult terrains and for their control of the number (k) of trees to measure. Although recent fixed-count estimators of density (PDE) are less biased than older ones, the risk of a nontrivial bias remains a deterrent. A recently published PDE based on a generic algorithm for predicting distances to the $k + m$ nearest tree ($m = 1, 2, \ldots$) has attractive properties in terms of average bias and average root mean squared errors across a wide spectrum of spatial point patterns. However, the risk of a sizeable bias remains an issue. Sensitivity to spatial patterns is seen as its main weakness. It is hypothesized that a new PDE with robust properties will mitigate the bias issue and encourage wider use. To this end, a new PDE estimator is proposed. It builds on a mixture of observed and predicted distances to a set of $k + m$ nearest trees to generate counts of actual and virtual trees inside a circle with a data-driven fixed radius. The proposed new robust fixed-count density estimator achieved an average absolute bias of 1.2% when tested across a wide range of point patterns (54 actual and four simulated). The maximum absolute bias was 4.4%, a significant reduction when compared with otherwise attractive alternative PDEs. Root mean squared errors and coverage of 95% confidence intervals were also encouraging. The deterrent bias issue in PDEs has been sharply reduced with the proposed estimator.

Key words: plotless sampling, k-tree sampling, stem density, forest inventory, point pattern, Pareto type I distribution.

Résumé : L'échantillonnage à nombre d'arbres fixe (sans placette) demeure attrayant pour les inventaires forestiers en terrain difficile et pour contrôler le nombre (k) d'arbres à mesurer. Bien que les récents estimateurs de densité par comptage fixe (EDSP : estimateur de densité sans placette) soient moins biaisés que les anciens, le risque de biais n'est pas négligeable et continue d'avoir un effet dissuasif. Un EDSP récemment publié fondé sur un algorithme générique pour prédire la distance jusqu'aux $k + m$ arbres les plus proches ($m = 1, 2, \ldots$) possède des propriétés intéressantes en termes de biais moyen et d'écart-type résiduel pour un large éventail de configurations spatiales de points. Cependant, le risque qu'un biais puisse être assez important persiste car la sensibilité de cet EDSP aux configurations spatiales est considérée comme sa principale faiblesse. On assume que le fait de doter un EDSP de propriétés robustes atténuera le problème de biais et entraînera une plus grande utilisation. Pour cela, un nouvel estimateur EDSP est proposé. Il s'appuie sur un mélange de distances observées et prédites jusqu'à un ensemble de $k + m$ arbres les plus proches pour générer des décomptes d'arbres réels et virtuels à l'intérieur d'un cercle à rayon fixe fondé sur les données. Le nouvel EDSP robuste qui est proposé a engendré un biais absolu moyen de 1,2 % lors d'un essai avec un large éventail de motifs de points (54 réels et quatre simulés). Le biais absolu maximal était de 4,4 %, une réduction significative comparativement aux autres EDSP par ailleurs attrayants. Les écart-types résiduels et les intervalles de confiance à 95 % étaient également encourageants. L'effet dissuasif des problèmes de biais des EDSP a été fortement réduit par l'estimateur proposé. [Traduit par la Rédaction]

Mots-clés : échantillonnage sans placette, échantillonnage de k arbres, densité des tiges, inventaire forestier, motif de points, distribution de Pareto (type I).

Introduction

Fixed-count sampling (viz., plotless or k-tree sampling) is attractive in terms of cost, logistics, flexibility, and control over sample yield (Coddington et al. 1991; McNeill et al. 1977; Nielson et al. 2004). However, the risk of biased estimates and a prospect of a lower precision than in comparable sampling with fixed-area plots (Kleinn and Vilčko 2006; Magnussen et al. 2008) may deter potential users (Steinke and Hennenberg 2006). Unbiased or nearly unbiased fixed-count estimators do exist, but they are either impractical (Fehrmann et al. 2012) or computational intensive (Nothdurft et al. 2010).

In fixed-count sampling, a random set of n locations is selected in a population of interest. In forestry applications, the population could be a single stand or a stratum. At each selected location (i), the distance ($r_{k,i}$, $i = 1, \ldots, n$) to the nearest kth tree is measured concurrently with other attributes of interest (Lynch and Rusydi 1999). Estimators of density derived from k-tree sampling are often referred to as plotless density estimators (PDEs) (Payandeh and Ek 1986), meaning that the actual reference area for a certain observation is missing or has to be determined by modelling. Unknown inclusion probabilities of sample trees are central to the bias problem (Fehrmann et al. 2012).

Continued interest in fixed-count sampling has fostered numerous attempts at reducing the bias problem (Magnussen et al. 2011). Successes have been limited to a subset of spatial point patterns (Magnussen et al. 2011) or, in case of composite estimators, limited to a reduction in the average bias across a large number of point patterns (Magnussen et al. 2011).

To sum up, a sizeable risk of a bias in excess of 10% remains an issue with fixed-count sampling. Unfortunately, a reliable method for predicting the performance of a fixed-count density estimator remains elusive. For this reason, the use of fixed-count estimators in ecological surveys has been discouraged (Steinke and Hennenberg 2006).

The resilience of bias suggests a common problem in model-based PDEs. They are typically designed around large-sample ex-

Received 19 July 2013. Accepted 20 December 2013.

S. Magnussen. Canadian Forest Service, Pacific Forestry Centre, 506 West Burnside Rd., Victoria, BC V8Z 1M5, Canada.
E-mail for correspondence: steen.magnussen@nrcan.gc.ca.

Published at www.nrcresearchpress.com/cjfr on 2 January 2014.

pectations and observed (fitted) trends in distances to the kth nearest tree. While a surprisingly simple recursive algorithm (Magnussen et al. 2011) can capture the expected trend in distances to the kth nearest tree (k = 3, 4, 5, …) in stationary point patterns (Illian et al. 2008), it is also clear that apparent noise and significant correlations among observed distances pose a complex obstacle towards an effective bias reduction.

The objective of this study is to improve robustness of a recently published fixed-count density estimator with virtual fixed-area plots (Magnussen et al. 2011). When tested in simulated random sampling from 54 stem-mapped actual forest stands and four artificial point patterns, the average absolute bias was only 1.6%, but in six cases, the absolute bias was between 6% and 12%. Notwithstanding, a comparison with promising alternative fixed-count estimators confirmed its attraction (Magnussen et al. 2011). When tested on the same data, the proposed new robust estimator achieved a mean absolute bias of 1.2% with a maximum of 4.4%.

Material and methods

Fixed-count estimators of density

Magnussen (2012*b*) proposed the following PDE with virtual fixed-area plots:

$$(1)\qquad \tilde{\lambda}_{\mathrm{vfix}}(k) = n^{-1} \times \left(\pi\, \hat{r}_{\mathrm{virt}}(k)^2\right)^{-1} \sum_{i=1}^{n} n_i^*$$

where n is sample size (number of randomly selected locations), k is the fixed count of distances to the nearest 1, 2,…, kth tree, $\hat{r}_{\mathrm{virt}}(k)$ is a data-dependent radius in a virtual fixed-area circular plot (see below), and n_i^* is the number of trees at the ith sample location that are located inside a circle with a radius $r_{\mathrm{virt}}(k)$ and centered on the ith sample location. If $n_i^* < k$, the count is an observation, and when $n_i^* \geq k$, the count may include predictions generated from a recursive algorithm (see below) of distances to the (k + m)th tree (m = 1, 2, …).

Computing $\tilde{\lambda}_{\mathrm{vfix}}$ involves the following steps: (*i*) choose a radius in a virtual circle centered on the sample location, as in Magnussen (2012*a*), the choice, based on trial and error, is $\hat{r}_{\mathrm{virt}} = n^2(\sum_{i=1}^{n} r_{k,i}^{-0.5})^{-2}$; (*ii*) at the ith sample location (i = 1, …, n), determine the number of trees n_i^{in} for which the distance to the jth nearest tree $r_{j,i} \leq \hat{r}_{\mathrm{virtual}}$, j = 1, …, k is true; (*iii*) if $n_i^{\mathrm{in}} < k$, set $n_i^* = n_i^{\mathrm{in}}$; (*iv*) if $n_i^{\mathrm{in}} = k$, then generate predictions of distances $\hat{r}_{k+m,i}$ to the k + 1, k + 2, …, k + m nearest tree using the generic algorithm in eq. 2 (see below) with starting values $\hat{r}_{k-1,i} = r_{k-1,i}$ and $\hat{r}_{k,i} = \hat{r}_{k-1,i} + (\bar{r}_k - \bar{r}_{k-1})$ until $\hat{r}_{k+m,i} > \hat{r}_{\mathrm{virt}}$, set $n_i^* = k + m - 1$; and (*v*) compute $\tilde{\lambda}_{\mathrm{vfix}}$ as per eq. 1. Here $\bar{r}_k$ and $\bar{r}_{k-1}$ denote the sample means of the distances to the kth and (k – 1)th nearest tree, respectively. The generic recursive algorithm proposed by Magnussen (2012*a*) is used for predicting the expected distance to the (k + m)th nearest tree:

$$(2)\qquad \hat{r}_{k+m,i} = \hat{r}_{k+m-1,i} + \frac{(\hat{r}_{k+m-1,i})^2 - \hat{r}_{k+m-1,i}\hat{r}_{k+m-2,i}}{2\hat{r}_{k+m-1,i} - \hat{r}_{k+m-2,i}},\quad m = 1, 2, \ldots;\ i = 1, \ldots, n$$

With the above choice of $\hat{r}_{\mathrm{virt}}$, k = 6, and the data used in this study, the average number of observed and predicted trees in a plot with radius $\hat{r}_{\mathrm{virt}}$ varied from 4.9 to 7.7. The fraction of virtual plots with one or more predicted tree distances varied between 0.42 and 0.55.

The estimator of variance of $\tilde{\lambda}_{\mathrm{vfix}}(k)$ follows from the random selection of sample points (Cochran 1977):

$$(3)\qquad \hat{\mathrm{V}}\mathrm{ar}(\tilde{\lambda}_{\mathrm{vfix}}(k)) = \frac{\sum_{i=1}^{n}\left(n_i^* - \bar{n}_i^*\right)^2}{n(n-1)\left(\pi\, \hat{r}_{\mathrm{virt}}^2\right)^2}$$

where $\bar{n}_i^*$ is the sample mean of the number of actual and predicted trees in the virtual plots with a radius of $\hat{r}_{\mathrm{virt}}$.

The proposed new robust fixed-count estimator of density with virtual plots $\tilde{\lambda}_{\mathrm{vfix2}}(k)$ and its estimator of variance have the same form as, respectively, the estimators in eq. 1 and eq. 3. However, counts n_i^*, for which $n_i^{\mathrm{in}} = k$, are based on a different set of predicted distances ($\tilde{r}_{k+m,\,i}$), specifically

$$(4)\qquad \tilde{r}_{k+m,i} = \tilde{r}_{k+m-1,i} \times \hat{E}\left(\frac{\tilde{r}_{k+m,i}}{\tilde{r}_{k+m-1,i}}\right) = \tilde{r}_{k+m-1,i} \times \frac{\hat{\beta}_0 + \hat{\beta}_1 \log(k+m-1)}{\hat{\beta}_0 - 1 + \hat{\beta}_1 \log(k+m-1)},\quad m = 1, 2, \ldots;\ i = 1, \ldots, n$$

where $\hat{E}(\tilde{r}_{k+m,i} \times \tilde{r}_{k+m-1,i}^{-1})$ is a model-based expectation of the distance ratio $r_{k+m,i} \times r_{k+m-1,i}^{-1}$, and $\hat{\beta}_0$ and $\hat{\beta}_1$ are nonlinear weighted least-squares regression coefficients determined from a mixture of observed and predicted distance ratios (see below). The starting value $\tilde{r}_{k,i}$ for the algorithm in eq. 4 is the observed distance $r_{k,i}$. The model-based expectation of the distance ratios in eq. 4 is that of a type I Pareto distribution with parameters 1 (the lower bound) and $\beta_0 + \beta_1 \log(k + m - 1)$ (Johnson et al. 1994, p. 574). A preliminary study indicated that for $50 > k + m \geq 3$, a type I Pareto distribution provided an excellent fit to observed distance ratios (AD test, Anderson and Darling 1952). Also, the expected value of the fitted Pareto distributions decreased almost linearly with the logarithm of the order of the tree distance in the numerator of a distance ratio. Equally important, for distances $r_{j,i}$, $j \geq 3$, pairs of sequential distance ratios appeared to be independent (Spearman's rho test, Conover 1980, p. 253).

In the current implementation of eq. 4 with k fixed at 6 (see below), the regression coefficients were estimated from the observed distance ratios $r_{3,i} \times r_{2,i}^{-1},\ldots,r_{6,i} \times r_{5,i}^{-1}$, i = 1, …, n. An additional set of M ratios, i.e., $\hat{r}_{k+1,i} \times \hat{r}_{k,i}^{-1},\ldots,\hat{r}_{k+M,i} \times \hat{r}_{k+M-1,i}^{-1}$, i = 1, 2, …, n, were then predicted via eq. 2. The number of predictions (M) should be high enough to ensure that at least one predicted distance is greater than the radius in the virtual plot. Here M was fixed at 5 k. Regression weights were the inverse of the distance order in the numerator of a ratio, i.e., an observed data ratio is given more weight than a predicted ratio. The weighting scheme also reflects the fact that the variance of a distance ratio decreased with the distance order in the numerator of a ratio. Various alternative weighting schemes were tried, but none offered improvements.

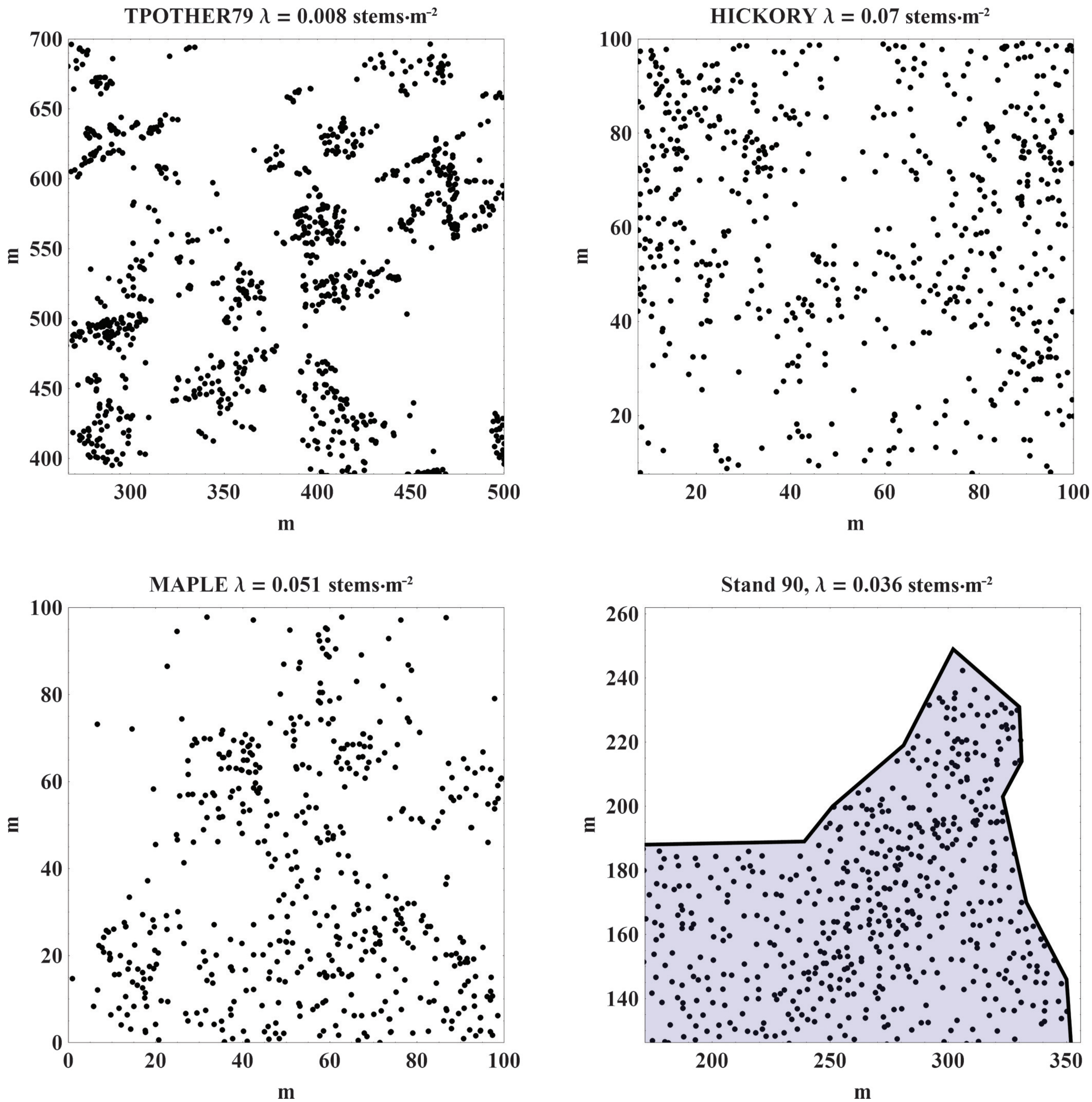
It is assumed that the combined use of observed and predicted distance ratios for estimating the mean trend in a Pareto distributed random variable (distance ratios) confers robust properties to $\tilde{\lambda}_{\mathrm{vfix2}}(k)$, robust in the sense that it is less sensitive to the spatial point pattern than $\tilde{\lambda}_{\mathrm{vfix}}(k)$.

Testing $\tilde{\lambda}_{\mathrm{vfix2}}(k)$

Testing $\tilde{\lambda}_{\mathrm{vfix2}}(k)$ was done with simulated fixed-count sampling with k = 6, sample size n = {5, 10, 15, 30}, and 18 000/n replications in each of 58 point patterns (four simulated and 54 actual). Sample locations were randomly selected. Summary statistics of the point patterns are provided in Magnussen (2012*b*). A sample (9) of contrasting point patterns are displayed in (Magnussen et al. 2011). Six additional patterns are displayed in Nothdurft et al. (2010). Figure 1 shows four of the spatial point patterns that are not completely spatially random.

Point densities vary from 0.0046 to 0.7361 m^{-2}, and the point patterns ranged from random (Poisson) to strongly clustered. A total of 35 test patterns were consistent with the hypothesis of

Fig. 1. Four nonrandom point patterns. Images have been cropped to a maximum of 600 points. Stand limits are shaded with full-line borders.



complete spatial randomness (csr; Illian et al. 2008, p. 53), and 23 indicated a significant departure from csr (Kolmogorov–Smirnov test; Conover 1980, p. 344).

To mitigate edge effects, all sample locations closer than $r_{k,i}$ to the nearest edge of a point pattern were discarded. Effective ways of minimizing edge effects in fixed-count sampling are known (Lynch 2012).

The following statistics supported inference of $\tilde{\lambda}_{\mathrm{vfix2}}$: (*i*) relative absolute bias (aBIAS) computed as the absolute of difference between the replicate mean estimate of density and the actual density divided by the actual density; (*ii*) relative root mean squared error (RRMSE) computed as the square root of the replicate variance of $\tilde{\lambda}_{\mathrm{vfix2}}$ plus the squared value of the estimate of bias divided by the actual density; (*iii*) the ratio of the absolute bias to the standard deviation of $\tilde{\lambda}_{\mathrm{vfix2}}$ in replicated sampling (Cochran 1977, p. 12); (*iv*) empirical variance of $\tilde{\lambda}_{\mathrm{vfix2}}$; and (*v*) coverage of computed nominal 95% confidence intervals for $\tilde{\lambda}_{\mathrm{vfix2}}$ (i.e., proportion of intervals covering the actual density).

All results for $\tilde{\lambda}_{\mathrm{vfix2}}$ were compared with parallel results for $\tilde{\lambda}_{\mathrm{vfix}}$, $\tilde{\lambda}_{\mathrm{AIC}}$ (Magnussen et al. 2012*a*), and $\hat{\lambda}_{\mathrm{fix}}$ (the fixed-area density estimator) and, in seven cases, with reconstruction density estimator (RDE) results by Nothdurft et al. (2010). For the comparison with RDE results, the sampling scheme of Nothdurft et al. (2010) was

Fig. 2. Smoothed probability density functions (Epanechnikov kernel) of absolute relative bias (aBIAS%) of three fixed-count (k = 6) estimators of stem density ($\tilde{\lambda}(k)$). Each estimator was tested in 58 point patterns (sample size n = 30).

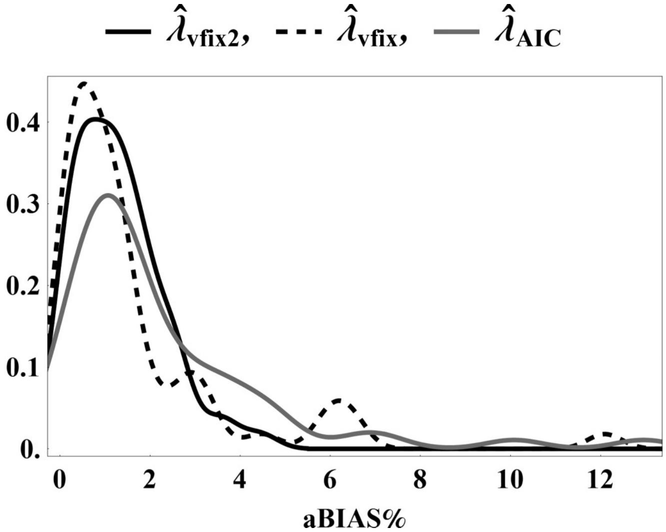
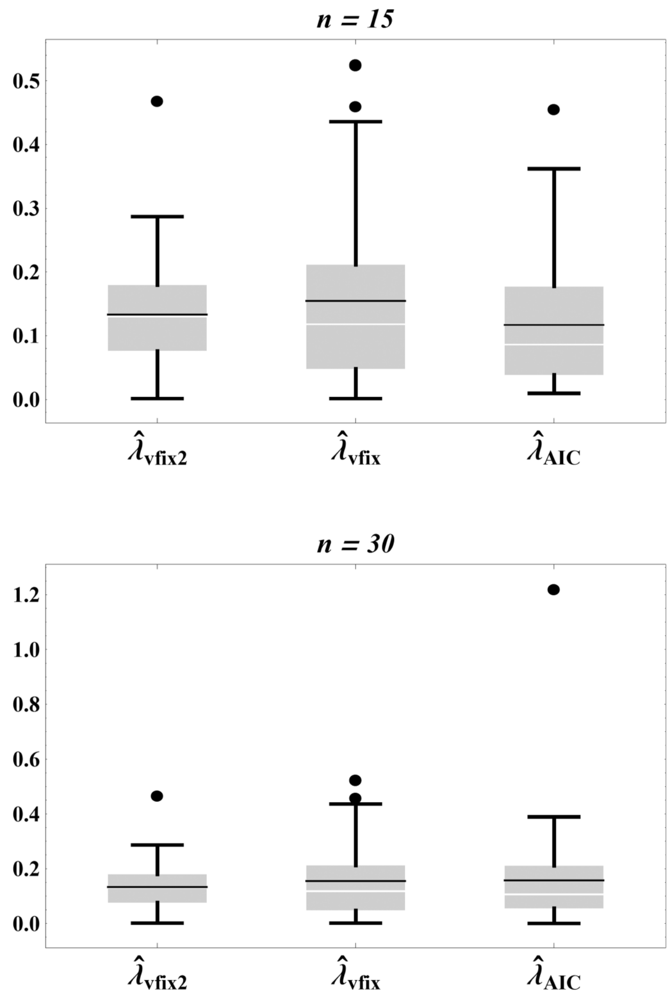


Fig. 3. Box–whisker plots of the ratio of absolute bias to precision of three fixed-count estimators of stem density ($\tilde{\lambda}(k)$) in 58 test point patterns with k = 6. Sample sizes (n) are 15 (top) and 30 (bottom). Outliers are indicated with solid circles.



adopted: one sample location per 2500 m², with sample locations located on a 50 m × 50 m grid randomly placed over a stem map. Sampling was replicated 250 times, each time on a new randomly selected grid. Results with $\hat{\lambda}_{fix}$ are unbiased and considered as benchmarks for sampling with fixed-area circular plots with a radius of $\hat{r}_{virt}$ (i.e., the number of points inside the circular area divided by the area of the circle; Illian et al. 2008, p. 83). The composite fixed-count estimator of density $\tilde{\lambda}_{AIC}$ uses Akaike's information criterion (AIC) to weight 16 model-based density estimators. Three of the 16 estimators were deemed to be the most promising point estimators for forestry applications. The remaining 13 were maximum likelihood estimators for the Poisson distribution (1), regular point patterns (3), and Poisson parameter mixture models (9). The study by Magnussen et al. (2012*a*) showed that $\tilde{\lambda}_{AIC}$ was superior (relative to three otherwise best-ranked fixed-count density estimators) in terms of performance (aBIAS, RRMSE) over the 58 point patterns in settings with $k = \{3, 6\}$, $n = \{15, 30\}$.

Results

Across the 58 test patterns, aBIAS of $\tilde{\lambda}_{vfix2}$ varied from 0.2% to 4.4%, with a mean of 1.2% and an interquartile range from 0.4% to 1.8%. On the 36 sites where the null hypothesis of csr was not rejected, the average bias was 1.2%. On 18 sites with an apparent clustered point pattern, the average bias was 1.1%. On four sites with a regular point pattern, the mean bias was 1.3%. The three means are regarded as equal (Kruskal–Wallis test, P = 0.61). For $\tilde{\lambda}_{vfix}$, aBIAS varied from 0.2% to 12.1%, with a mean of 1.6% and an interquartile range from 0.4% to 1.7%. The mean and maximum aBIAS of the composite estimator $\tilde{\lambda}_{AIC}$ were 2.3% and 12.9%, respectively. Smoothed distributions (Epanechnikov kernel smoothing; Silverman 1986, p. 42) of the 58 estimates of aBIAS from $\tilde{\lambda}_{vfix2}$, $\tilde{\lambda}_{vfix}$, and $\tilde{\lambda}_{AIC}$ are shown in Fig. 2. The hypothesis of equal distributions was rejected for all pair-wise comparisons (AD test, P < 0.01). Testing of equal locations (Kruskal–Wallis test; Conover 1980, p. 259) and equal variances (Levene's test; Shoemaker 2003) led to similar conclusions. It is clear that the advantage of $\tilde{\lambda}_{vfix2}$ is a narrower distribution of aBIAS when it is compared with those of $\tilde{\lambda}_{vfix}$ and $\tilde{\lambda}_{AIC}$. It should be noted that it is not unusual to see a bias in excess of 20% with PDEs tested on the same 58 patterns but not considered here (Magnussen 2012*b*; Magnussen et al. 2008, 2011).

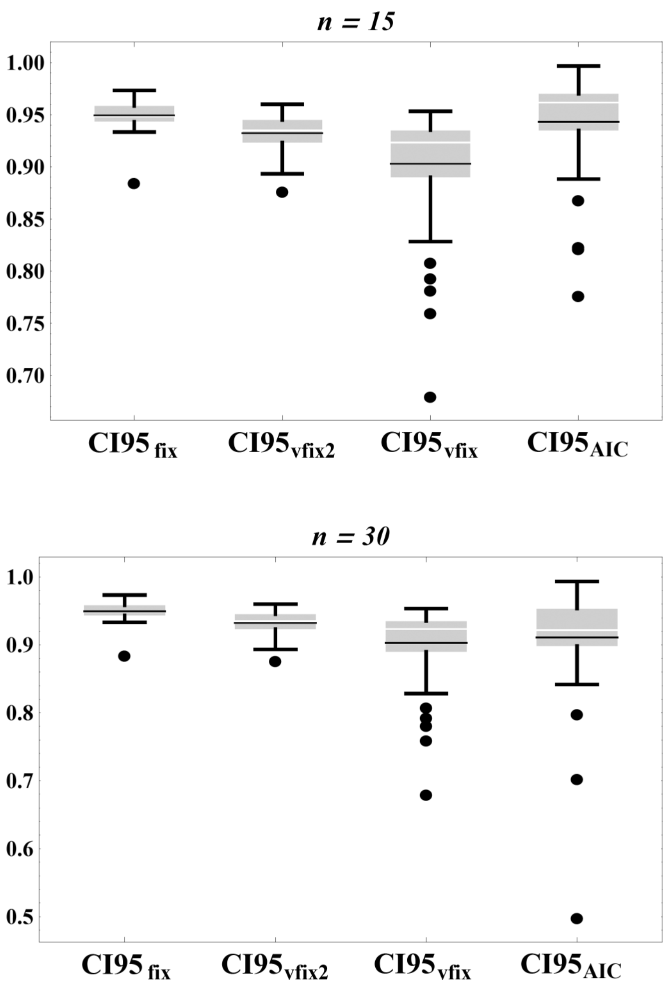
Biased estimators generate biased estimates of accuracy (Cochran 1977, p. 12). A ratio of aBIAS to the standard deviation of the estimate ($\hat{\sigma}_{\lambda}$) below 0.1 may have little effect on inference, yet a serious bias is possible for ratios greater than 0.2 (Cochran 1977, p. 13). A box–whisker plot of the ratio of aBIAS to $\hat{\sigma}_{\lambda}$ is shown in Fig. 3. The three distributions of ratios are not significantly different (AD tests, P > 0.26). The mean ratios were 0.15 for $\tilde{\lambda}_{vfix2}$, 0.13 for $\tilde{\lambda}_{vfix}$, and 0.16 for $\tilde{\lambda}_{AIC}$. A quarter of the ratios were greater than 0.2.

The empirical (replication) variance of $\tilde{\lambda}_{vfix2}$ matched, irrespective of sample size, closely (correlation coefficient > 0.99, slope ≅ 1.0) the replication average of the estimated variances obtained from eq. 3. The mean absolute relative difference (across test patterns) was 4.9% (maximum of 7.2%).

The bias reduction achieved with $\tilde{\lambda}_{vfix2}$ comes, not surprisingly, at the price of a slightly higher (9%) empirical variance than achieved with $\tilde{\lambda}_{vfix}$. The RRMSE of $\tilde{\lambda}_{vfix2}$ was, on average across point patterns and sample sizes, 2% larger than the RRMSE of $\tilde{\lambda}_{vfix}$. With n = 30, the lowest RRMSEs came from the four sites with a regular point pattern (mean: 5.7%) and the highest came from the 18 sites with a clustered point pattern (mean: 12.2%). Sites with an assumed random point pattern (36) were intermediate (mean: 8.8%). Overall, estimator differences in RRMSEs remained nearly constant across sample sizes.

Relative root mean squared error of $\tilde{\lambda}_{AIC}$ was, on average, approximately 1.5 times larger than the RRMSEs of $\tilde{\lambda}_{vfix}$ and $\tilde{\lambda}_{vfix2}$. There was apparently no effect of sample size in these trends.

Fig. 4. Box–whisker plots of coverage of nominal 95% confidence intervals (CI95) for four estimators of stem density ($\tilde{\lambda}(k)$) in 58 test point patterns with k = 6. Sample sizes (n) are 15 (top) and 30 (bottom). Outliers are indicated with solid circles.



Fixed-count sampling with virtual fixed-area plots was, as expected, less precise than simple random sampling with equal-area circular plots. Across sample sizes and test patterns, the standard deviation of $\tilde{\lambda}_{vfix2}$ was 9% larger than for $\tilde{\lambda}_{fix}$, whereas that of $\tilde{\lambda}_{vfix}$ was 8% larger. Standard deviations of $\tilde{\lambda}_{AIC}$ were approximately 24% larger than those of $\tilde{\lambda}_{fix}$.

On average, nominal 95% confidence intervals for $\tilde{\lambda}_{vfix2}$ achieved, across test patterns and sample sizes, coverage of 93% (minimum of 88%, median of 94%, and maximum of 96%). Sample size had no discernible effect on coverage. Half of the computed intervals achieved coverage of 92% to 95%. Results for $\tilde{\lambda}_{vfix}$ (mean of 90%, minimum of 68%, median of 92%, and maximum of 95%) suggested a more serious problem of undercoverage. Coverage of confidence intervals for $\tilde{\lambda}_{AIC}$ varied from a low of 88% to a high of 100%, with a mean that decreased with sample size from 0.98 at n = 5 to 0.95 at n = 30. An opposite trend in the bias of estimated variances of $\tilde{\lambda}_{AIC}$ explains these otherwise surprising results. Sampling with fixed-area plots achieved, on average, the nominal coverage (mean of 95%, minimum of 89%, median of 95%, and maximum of 97%). A box–whisker plot in Fig. 4 summarizes the coverage results.

Results from the comparison of relative bias and RRMSE obtained with RDE and $\tilde{\lambda}_{vfix2}$ are presented in Table 1. In all seven cases, the bias of $\tilde{\lambda}_{vfix2}$ is larger than the bias of the RDE estimator

Table 1. Relative bias and root mean squared error (RRMSE) of $\tilde{\lambda}_{vfix2}$ and the reconstruction density estimator (RDE) by Nothdurft et al. (2010).

		RDE		$\tilde{\lambda}_{vfix2}$	
ID[a]	No. of trees/ha	Bias%	RRMSE%	Bias%	RRMSE%
7	432	–0.7	10.0	–3.4	8.4
21	242	1.8	7.5	3.0	5.5
28	421	1.5	10.9	2.7	9.6
30	838	–0.9	11.0	3.2	7.5
61	483	0.0	9.6	–2.6	7.9
67	878	0.1	8.9	0.1	7.7
LP[b]	116	–0.2	23.5	–5.8	20.9

[a]STIPSI identifier, http://www.fva-bw.de/forschung/bui/stipsi_en.html.
[b]Tall pines (*Pinus palustris*) in Wade County (Florida), see Platt et al. (1988).

(mean absolute difference of 2.5%). Conversely, the RRMSE of $\tilde{\lambda}_{vfix2}$ was consistently smaller (mean difference of 2%).

Discussion

The most important difference between the proposed new robust fixed-count density estimator $\tilde{\lambda}_{vfix2}$ and $\tilde{\lambda}_{vfix}$ is a significant reduction in the maximum absolute bias when tested on 58 diverse point patterns. Earlier tests, with the same point patterns and a suite of otherwise promising alternative fixed-count density estimators, have shown that the risk of an absolute relative bias in excess of 10% is far from trivial (Magnussen et al. 2011).

If the maximum absolute relative bias of $\tilde{\lambda}_{vfix2}$ remains around the 4% mark in further testing and applications, the major issue of bias leveled against fixed-count density estimation (Steinke and Hennenberg 2006) becomes less important. In this favorable scenario, the well-known practical and logistical advantages of fixed-count sampling over sampling with fixed-area plots must be weighed against an anticipated loss of approximately 9% in precision. This loss can be counteracted by an increase in sample size by approximately 18% according to this study. Because local cost structures and field conditions can be highly variable, it will not be possible to draw general conclusions regarding when fixed-count sampling is more attractive than fixed-area sampling. Sampling across highly variable local densities will, everything else being equal, favor a fixed-count sampling scheme (Jonsson et al. 1992). The same goes for difficult terrain. A simple global recommendation concerning the choice of k and n is also difficult to formulate (Lessard et al. 2002). In forestry, however, it is known that low values of k (<4) produce nonrobust results irrespective of the employed PDE (Kleinn and Vilčko 2006; Lynch and Wittwer 2003; Magnussen et al. 2008), even when theory and simulation studies may suggest otherwise (for example, see Byth 1982). Sampling time does increase exponentially with k (Fehrmann et al. 2012), which dictates a k no greater than 7.

A legion of alternatives to the proposed robust fixed-count estimator of density is available, many with a reported stellar performance in a few point patterns but also a poor performance in several others (Magnussen et al. 2011). When an estimator is expected to perform well in, say, a 'Poisson forest' (Pollard 1971), a regular point pattern (Miyagawa 2009), or a quasi-regular pattern (Kleinn and Vilčko 2006), the analyst may be able to make an informed choice of estimator based on its expected performance and on inference about the spatial point pattern from data collected in a fixed-count sample. Unfortunately, sample sizes are often too small to bestow sufficient statistical power to justify the choice of one estimator over another. Besides, fixed-count sample data preclude inference about a large number of important point patterns. Robust fixed-count estimators of densities are therefore needed for protection against excessive bias (Steinke and Hennenberg 2006; White et al. 2008). Composite estimators appear promising in this regard (Magnussen et al. 2011). The remarkable RDE fixed-count estimator by Nothdurft et al. (2010) is approximately unbiased, but its RRMSE, although generally much lower than in otherwise popular PDE alternatives, can be expected to be somewhat larger than the RRMSEs of $\tilde{\lambda}_{vfix2}$ and $\tilde{\lambda}_{vfix}$. Computational issues and the lack of a variance estimator for the RDE estimator weigh in favor of the latter two.

The robust properties of $\tilde{\lambda}_{vfix2}$ come, as expected, at the price of an increase in variance and a slight increase in the expected RRMSE relative to that of $\tilde{\lambda}_{vfix}$. Yet the larger variance brings about a reduction of an otherwise important undercoverage problem with 95% confidence intervals. Vis-à-vis the greater efficiency of fixed-area plot density estimators, the increase in the variance of $\tilde{\lambda}_{vfix2}$ relative to that of $\tilde{\lambda}_{vfix}$ is less important than the improved robustness against excessive bias.

Practical implementation of $\tilde{\lambda}_{vfix2}$ is straightforward and requires no more than (*i*) the use of the recursive algorithm in eq. 2 for generating predicted distances and (*ii*) standard procedures for a weighted nonlinear least squares regression. Edge effects can be mitigated by standard procedures for fixed-area plots when fewer than *k* trees are inside the virtual plot (Gregoire and Valentine 2008, p. 223) or by a combination of methods for fixed-area and fixed-count plots (Lynch 2012) when *k* or more trees are inside the virtual plot. The associated variance estimator is also simple, apparently reliable, and in principle equal to the estimator of variance used for simple random sampling with fixed-area plots (Gregoire and Valentine 2008, p. 25). Computation with the proposed estimator is fast and programming is easy in most software packages. On a typical modern desktop computer and sample sizes of 30 or less, it takes less than 3 s to complete processing and computing of all statistics shown in this study. Simpler PDEs can be computed in less than a second. A single estimation procedure with RDE by (Nothdurft et al. 2010) can take up to 2 h.

Despite a remarkable reduction in the maximum absolute bias, the proposed robust fixed-count estimator of density remains a biased estimator. In this study, the average absolute bias was 1.2%, which sometimes is interpreted as nearly unbiased. Further improvements of the proposed estimator seem possible; yet explorations with changes to (*i*) the radius in the virtual plots, (*ii*) the number of recursively generated predictions of distances, (*iii*) the model form (trend) for the Pareto parameter, and (*iv*) the weighting scheme all suggest that further improvements by manipulating these parameters will be unimportant. An outright elimination of the bias in fixed-count density estimators requires a return to design-based estimators (Fehrmann et al. 2012). Until they become practical, the proposed robust fixed-count estimator may just be the best model-based alternative in practical applications.

Acknowledgements

Stem location data from five Brazil sites (Project: Functional Gradient, FAPESP 03/12595-7) were kindly made available by L.F. Alves. Data from 30 stands in Europe were obtained from STIPSI (http://www.fva-bw.de/forschung/bui/stipsi_en.html, accessed 12 August 2009). Data from the Smithsonian Barro Colorado Island research plot (Condit and Hubbell 1998; Hubbell et al. 1999) were downloaded from the Center for Tropical Forest Science research site (http://ctfs.si.edu, last accessed 22 May 2009). Thanks are extended to C. Kleinn for the Miombo data, I.F. Sun for the Fushan data, and W.J. Platt for the tall pines data.

References

Anderson, T.W., and Darling, D.A. 1952. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. Ann. Math. Stat. **23**(2): 193–212. doi:10.1214/aoms/1177729437.

Byth, K. 1982. On robust distance-based intensity estimators. Biometrics, **38**(1): 127–135. doi:10.2307/2530295.

Cochran, W.G. 1977. Sampling techniques. Wiley, New York.

Coddington, J.A., Griswold, C.E., Silva, D., Peqaranda, E., and Larcher, S.F. 1991. Designing and testing sampling protocols to estimate biodiversity in tropical ecosystems. *In* Proceedings of the Fourth International Congress of Systematic and Evolutionary Biology,The unity of evolutionary biology. *Edited by* E.C. Dudley. Dioscorides Press, Portland, Oregon. pp. 44–60.

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

Conover, W.J. 1980. Practical nonparametric statistics. Wiley, New York.

Fehrmann, L., Gregoire, T., and Kleinn, C. 2012. Triangulation based inclusion probabilities: a design-unbiased sampling approach. Environ. Ecol. Stat. **19**(1): 107–123. doi:10.1007/s10651-011-0177-9.

Gregoire, T.G., and Valentine, H.T. 2008. Sampling strategies for natural resources and the environment. Chapman and Hall/CRC.

Hubbell, S.P., Foster, R.B., O'Brien, S.T., Harms, K.E., Condit, R., Wechsler, B., Wright, S.J., and Lao., S.L.d. 1999. Light gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. Science, **283**: 554–557.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. 2008. Statistical analysis and modelling of spatial point patterns. Wiley, Chichester, UK.

Johnson, N.L., Kotz, S., and Balakrishnan, N. 1994. Continuous univariate distributions. 2nd ed. Wiley, New York.

Jonsson, B., Holm, S., and Kallur, H. 1992. A forest inventory method based on density-adapted circular plot size. Scand. J. For. Res. **7**(1): 405–421. doi:10.1080/02827589209382733.

Kleinn, C., and Vilčko, F. 2006. A new empirical approach for estimation in *k*-tree sampling. For. Ecol. Manage. **237**(1–3): 522–533. doi:10.1016/j.foreco.2006.09.072.

Lessard, V.C., Drummer, T.D., and Reed, D.D.A. 2002. Precision of density estimates from fixed-radius plots compared to *n*-tree distance sampling. For. Sci. **48**(1): 1–6.

Lynch, T.B. 2012. A mirage boundary correction method for distance sampling. Can. J. For. Res. **42**(2): 272–278. doi:10.1139/x11-185.

Lynch, T.B., and Rusydi, R. 1999. Distance sampling for forest inventory in Indonesian teak plantations. For. Ecol. Manage. **113**(2–3): 215–221. doi:10.1016/S0378-1127(98)00427-7.

Lynch, T.B., and Wittwer, R.F. 2003. *n*-Tree distance sampling for per-tree estimates with application to unequal-sized cluster sampling of increment core data. Can. J. For. Res. **33**(7): 1189–1195. doi:10.1139/x03-036.

Magnussen, S. 2012*a*. Fixed-count density estimation with virtual plots. Spatial Statistics, **2**: 33–46. doi:10.1016/j.spasta.2012.09.001.

Magnussen, S. 2012*b*. A new composite *k*-tree estimator of stem density. Eur. J. For. Res. **131**(5): 1513–1527. doi:10.1007/s10342-012-0619-4.

Magnussen, S., Kleinn, C., and Picard, N. 2008. Two new density estimators for distance sampling. Eur. J. For. Res. **127**(3): 213–224. doi:10.1007/s10342-007-0197-z.

Magnussen, S., Fehrman, L., and Platt, W. 2011. An adaptive composite density estimator for distance sampling. Eur. J. For. Res. **131**(2): 307–320. doi:10.1007/s10342-011-0502-8.

McNeill, L., Kelly, R.D., and Barnes, D.L. 1977. The use of quadrat and plotless methods in the analysis of the tree and shrub component of woodland vegetation. Proc. Grassld. Soc. Sth. Afr. **12**: 109–113. doi:10.1080/00725560.1977.9648817.

Miyagawa, M. 2009. Order distance in regular point patterns. Geographical Analysis, **41**: 252–262. doi:10.1111/j.1538-4632.2009.00737.x.

Nielson, R.M., Sugihara, R.T., Boardman, T.J., and Engeman, R.M. 2004. Optimization of ordered distance sampling. Environmetrics, **15**(2): 119–128. doi:10.1002/env.627.

Nothdurft, A., Saborowski, J., Nuske, R.S., and Stoyan, D. 2010. Density estimation based on *k*-tree sampling and point pattern reconstruction. Can. J. For. Res. **40**(5): 953–967. doi:10.1139/X10-046.

Payandeh, B., and Ek, A.R. 1986. Distance methods and density estimators. Can. J. For. Res. **16**(5): 918–924. doi:10.1139/x86-163.

Platt, W.J., Evans, G.W., and Rathbun, S.L. 1988. The population dynamics of a long-lived conifer (*Pinus palustris*). Am. Nat. **131**(4): 491–525. doi:10.1086/284803.

Pollard, J.H. 1971. On distance estimators of density in randomly distributed forests. Biometrics, **27**: 991–1002. doi:10.2307/2528833.

Shoemaker, L.H. 2003. Fixing the F test for equal variances. Am. Stat. **57**(2): 105–114. doi:10.1198/0003130031441.

Silverman, B.W. 1986. Density estimation for statistics and data analysis. Chapman & Hall, London.

Steinke, I., and Hennenberg, K.J. 2006. On the power of plotless density estimators for statistical comparisons of plant populations. Can. J. Bot. **84**(3): 421–433. doi:10.1139/b05-135.

White, N.A., Engeman, R.M., Sugihara, R.T., and Krupa, H.W. 2008. A comparison of plotless density estimators using Monte Carlo simulation on totally enumerated field data sets. BMC Ecol. **8**: 6. doi:10.1186/1472-6785-8-6.