Ecological Modelling 147 (2002) 1–9

ECOLOGICAL MODELLING

www.elsevier.com/locate/ecolmodel

Estimating tree diameter class frequencies

Edwin J. Green [a,*], Michael Clutter [b]

[a] *Department of Ecology, Evolution and Natural Resources, Rutgers University, 14 College Farm Road, New Brunswick, NJ 08901-8551, USA*

[b] *The Timber Company, 100 Peachtree Street NW, Ste. 2650, Atlanta, GA 30303, USA*

Received 1 September 2000; received in revised form 18 April 2001; accepted 7 May 2001

Abstract

The problem of estimating stand tables in stands with few sample points is considered. The usual point sampling estimate of trees per hectare by diameter class is examined, along with two alternative estimators: a precision-weighted composite estimator and a pseudo-Bayes estimator. Stand tables are estimated for a subject stand with each of the three estimators in a simulation experiment. Both the composite and pseudo-Bayes estimator appear superior (in terms of average absolute error and mean squared error) to the usual estimator, although they do introduce a slight bias. The pseudo-Bayes estimator appears to perform the best. This estimator is also easier to use than the composite estimator because it does not require variance estimates.

Keywords: Composite estimator; Bayesian inference; Forest inventory; Loblolly pine

1. Introduction

One common method of summarizing forest inventory data is the construction of stand tables. These are estimates of the number of trees per unit area by diameter class in a particular forest stand or homogenous (in some sense) group of trees. In the US, diameter classes are usually 1 inch (2.54 cm) wide, although this is subject to variation. Stand tables are commercially important because the product potential of a tree is heavily dependent on its diameter. In addition, stand tables are important from an ecological perspective because tree diameter is highly related to a number of important attributes, such as biomass, sequestered carbon, mast and forage production, wildlife habitat value, etc.

Stand tables are ordinarily constructed from data collected exclusively in the stand of interest. Individual trees are not ordinarily selected in a simple random sample. Instead, groups of trees are usually measured. This is deemed more economical, since a large portion of the cost of a forest inventory is accounted for by travel time. Hence, once a particular point in the stand is reached, it would seem wasteful to measure only one tree. The groups of trees are usually selected in one of two ways: fixed area plots or horizontal points (e.g. see Avery and Burkhart, 1994). Fixed

* Corresponding author. Fax: +1-732-932-8746.
E-mail addresses: green@crssa.rutgers.edu (E.J. Green), mike_clutter@ttcmail.com (M. Clutter).

PII: S0304-3800(01)00402-1

area plots are self-explanatory. In this case, the estimate of the number of trees per unit area in a diameter class is simply the average number of trees per unit area in that diameter class over all the plots visited. However, fixed area plot sampling is not as common as point sampling. In the latter, a point is established on the ground and trees are selected by comparing their diameters to an angle projected horizontally from the point. It can be shown that this results in selecting trees with probability proportional to their cross-sectional area. Thus, larger trees have a higher probability of being selected and the sample is concentrated on larger trees. Since these trees tend to be the most valuable, this method of sampling is often preferred to fixed-area plot sampling. Point sampling complicates the estimation of trees per unit area in a diameter class slightly, but the usual unbiased estimator is well known and easily applied. In this study, we will assume that all data arise from a point sample.

The number of points installed in a stand is generally proportional to the size of the stand. It follows that stand table estimates should be acceptable in terms of accuracy and/or precision in large stands because many sample points are observed. However, in small stands that receive only a few points, the estimates may be highly imprecise. Recognition of this motivated us to consider alternative methods for estimating stand tables. Our objective was to use information from other similar stands to help improve the estimated stand table for a particular subject stand. In this study, we compare two alternative estimators to the usual stand table estimator in a simulation experiment.

2. Estimators

2.1. Usual estimator

Suppose we are concerned with p stands and that n points are randomly installed in the subject stand. The usual estimate of the number of trees per hectare in diameter class k ($k = 1, 2, \ldots, c$) at point i ($i = 1, 2, \ldots, n$) is:

$$\hat{N}_{ki} = \sum_{j=1}^{m_i} \frac{\text{BAF}}{b_{ij}} I(d_{ij}),$$

$$I(d_{ij}) = \begin{cases} 1 & \text{if } d_{ij} \text{ is in class } k, \\ 0 & \text{if not,} \end{cases} \tag{1}$$

where BAF is the constant basal area factor, i.e. the basal area (or cross-sectional area measured at 1.3 m above ground) in m^2/ha represented by each sample tree; m_i, the number of sample trees at point i in subject stand; c, the number of diameter classes in stand table; d_{ij}, the diameter (cm) of tree j on point i in subject stand, $j = 1, 2, \ldots, m_i$; and b_{ij}, the basal area (m^2) of tree j on point i in subject stand, $j = 1, 2, \ldots, m_i$.

For more details on this estimator see any standard forest measurements text, such as Avery and Burkhart (1994). The usual estimate for the number of trees per hectare in diameter class k is:

$$\hat{N}_k = \frac{\sum_{i=1}^{n} \hat{N}_{ki}}{n}. \tag{2}$$

2.2. Precision-weighted composite estimator

Assume there is an alternative estimate (e.g. from a similar stand) available of the stand table for the subject stand. Further, assume that the sample variance is available for the number of trees per hectare in each diameter class for the alternative estimate. Then, the precision-weighted composite estimator (hereafter referred to as simply the composite estimator) is

$$\bar{N}_k = \frac{\hat{N}_k {s_{\hat{N}_k}}^{-2} + \hat{T}_k {s_{\hat{T}_k}}^{-2}}{{s_{\hat{N}_k}}^{-2} + {s_{\hat{T}_k}}^{-2}} \tag{3}$$

where $\hat{T}_k$ is the alternative estimator of trees per hectare in diameter class k in subject stand; $s^2_{\hat{N}_k}$, is the sample variance for $\hat{N}_k$

$$s^2_{\hat{N}_k} = \frac{\sum_{i=1}^{n} (\hat{N}_{ik} - \hat{N}_k)^2}{n(n-1)} \tag{4}$$

and $s^2_{\hat{T}_k}$ is the sample variance for $\hat{T}_k$ (calculated in the same manner as $s^2_{\hat{N}_k}$).

2.3. Pseudo-Bayes estimator

In the pseudo-Bayes estimator, we predict the number of trees per hectare in stand s and the probability that a given tree is in diameter class k. The product of these terms yields an estimate of the number of trees per hectare in the diameter class. We modify the common Bayesian Dirichlet-multinomial model to estimate the diameter class probabilities.

2.3.1. Diameter class probabilities

Suppose $Y = (Y_1, Y_2, \ldots, Y_c)'$ follows a multinomial distribution, i.e. $Y \sim M_c(w, p_1, p_2, \ldots, p_c)$, where Y_i is the number of observations in class i, $i = 1, 2, \ldots, c$, w is the sample size and p_i is the probability of an observation falling into class i. Assume the conjugate Dirichlet prior distribution for $p = (p_1, p_2, \ldots, p_c)'$, i.e. $p \sim D_c(\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_c)$, where $\alpha_0 = \Sigma_j \alpha_j$. Under this model, the prior expectation for p_k is (α_k/α_0) and the variance of p_k is $(\alpha_0 - \alpha_k)\alpha_k/\alpha_0^2(\alpha_0 + 1)$. As α_0 increases, the variance of p_k decreases and α_0 can be regarded as an analog of sample size. In this sense, we see that α_0 measures the 'strength' of the prior belief. Now, let y_k be the number of observations with value k, $k = 1, 2, \ldots, c$. The posterior distribution for p is well known (e.g. see O'Hagan, 1994) to be the Dirichlet distribution:

$$p \setminus y, \alpha \sim D_c(\alpha_0 + w,\ \alpha_1 + y_1,\ \alpha_2 + y_2, \ldots, \alpha_c + y_c), \tag{5}$$

where $y = (y_1, y_2, \ldots, y_c)'$, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_c)'$, $w = \Sigma_j y_j$ and the posterior expectation for p_k is $(y_k + \alpha_k)/(w + \alpha_0)$.

Unfortunately, the Dirichlet-multinomial model seems inappropriate for point sampling data. In the model, the sample size (w) is the total number of sample units observed and classified, i.e. the number of trees. In the present case, the proper sample size is the number of points visited not the total number of trees measured. Hence, to more accurately reflect the sampling conditions, we make the following ad-hoc modification to the standard Bayesian estimator: define a new variable z_k such that

$$z_k = \frac{\hat{N}_k}{\sum_{j=1}^{c} \hat{N}_j} n, \qquad k = 1, 2, \ldots, c \tag{8}$$

This definition ensures that $n = \Sigma_j z_j$, i.e. the sample size is equal to the sum of the observations, just as in the multinomial model. However, z_k is not constrained to assume integer values and hence, it cannot follow the multinomial distribution.

We invoke the usual Dirichlet prior on p, i.e. $p \sim D_c(\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_c)$, where

$$\alpha_k = \frac{\hat{T}_k}{\sum_{j=1}^{c} \hat{T}_j} \alpha_0, \qquad k = 1, 2, \ldots, c, \tag{9}$$

i.e. our prior guess for p_k is $(\hat{T}_k/\Sigma_j \hat{T}_j)$. Substituting z_k for y_k in the posterior expectation for p_k yields the following estimate of p_k:

$$\hat{p}_k = \frac{z_k + \alpha_k}{n + \alpha_0}. \tag{10}$$

Hence, we have arrived at a pseudo-Bayesian estimator for the probability that a tree on a randomly chosen point falls in diameter class k. The choice of α_0 will be discussed in Section 2.4, as will the suitability of the Dirichlet prior distribution.

2.3.2. Total number of trees per hectare

To complete our estimate of the number of trees per hectare in diameter class k, we require an estimate of the total number of trees per hectare (N). One reasonable estimate for this parameter is:

$$\hat{N} = \frac{n \sum_{j=1}^{c} \hat{N}_j + \alpha_0 \sum_{j=1}^{c} \hat{T}_j}{n + \alpha_0}. \tag{11}$$

One could, of course, replace Eq. (11) with a precision-weighted estimate of N. We prefer Eq. (11) for reasons that will be discussed later.

2.3.3. Stand table estimator

The pseudo-Bayesian estimator for the number of trees per hectare in diameter class k is thus

$$\hat{N}''_k = \hat{N}\left(\frac{z_k + \alpha_k}{n + \alpha_0}\right). \tag{12}$$

2.4. Simulation experiment

In order to evaluate the performance of $\hat{N}_k$, $\hat{N}'_k$ and $\hat{N}''_k$, we performed a simulation study. To evaluate the performance of a forest inventory estimator, one needs large stands in which the size and location of each individual tree is known. Unfortunately, such stands are rare. Hence, we used simulated stands, generated from PTAEDA2, a commonly used loblolly pine growth and yield simulation model (Farrar et al., 1987). The output from PTAEDA2 includes the diameter and location of every live tree in the stand.

We assumed that our subject stand was 27 years old, planted at a density of 1482 trees per hectare (600 trees per acre) on land with a site index of 19.8 m (65 ft) (site index is the average height of the tallest trees in a stand at a reference age and is commonly used to assess the productive capacity of a site). The generated stand was ≈ 40.5 ha (100 ac) in size. We generated numerous alternative stands and used these to compute the composite and pseudo-Bayes estimates of the stand table in the subject stand. The attributes of the alternative stands are presented in Table 1. Each alternative stand was also ≈ 40.5 ha in size.

Table 1
Attributes of alternative stands

Age (years)	Trees planted per hectare	Site index (m)
22	1482	19.8
24	1482	19.8
25	1482	19.8
27	1482	19.8
29	1482	19.8
30	1482	19.8
32	1482	19.8
27	1235	19.8
27	1359	19.8
27	1482	19.8
27	1606	19.8
27	1730	19.8
27	1482	16.8
27	1482	18.3
27	1482	18.9
27	1482	19.5
27	1482	19.8
27	1482	20.1
27	1482	20.7
27	1482	21.3
27	1482	22.9

The simulation experiment proceeded as follows: on each iteration, five sample points were randomly located in the subject stand, using a basal area factor (BAF) of 10 (see, for example Avery and Burkhart, 1994 for a discussion of the BAF). We chose this BAF because it is the standard for use in Eastern and Southern US. Experience has shown that a BAF 10 point sample will usually yield around ten to 12 trees per point and this is deemed adequate. Smaller BAFs would include too many sample trees, while larger BAFs would include too few. The usual stand table estimate was computed using the data obtained from the subject stand. Then, a number of sample points (5, 10, 25 or 50) were randomly located in one of the alternative stands and the data obtained were used to compute the composite and pseudo-Bayes estimates of the stand table in the subject stand.

The parameter α_0 in the Dirichlet prior distribution for p can be viewed as a measure of the strength of the prior belief. We set this equal to the number of points installed in the alternative stand (n_a). Each simulated stand table was compared to the known stand table for the subject stand. The simulation was run for 1000 iterations. Diameter classes were 1 inch (2.54 cm) in width. On each iteration of the simulation, the largest sample tree found in either the subject stand or the alternative stand was determined. The maximum size class (c) was taken to be the size class into which the largest tree fell.

In order to check whether the Dirichlet prior used in the pseudo-Bayes estimator was reasonable, we randomly located 100 BAF 10 sample points in the subject stand and calculated the percentage of trees per hectare in each diameter class. If the Dirichlet prior was reasonable, then the marginal distributions of these percentages should have followed a beta distribution. In Fig. 1, we display the scaled relative frequencies for four representative diameter classes (jagged lines).

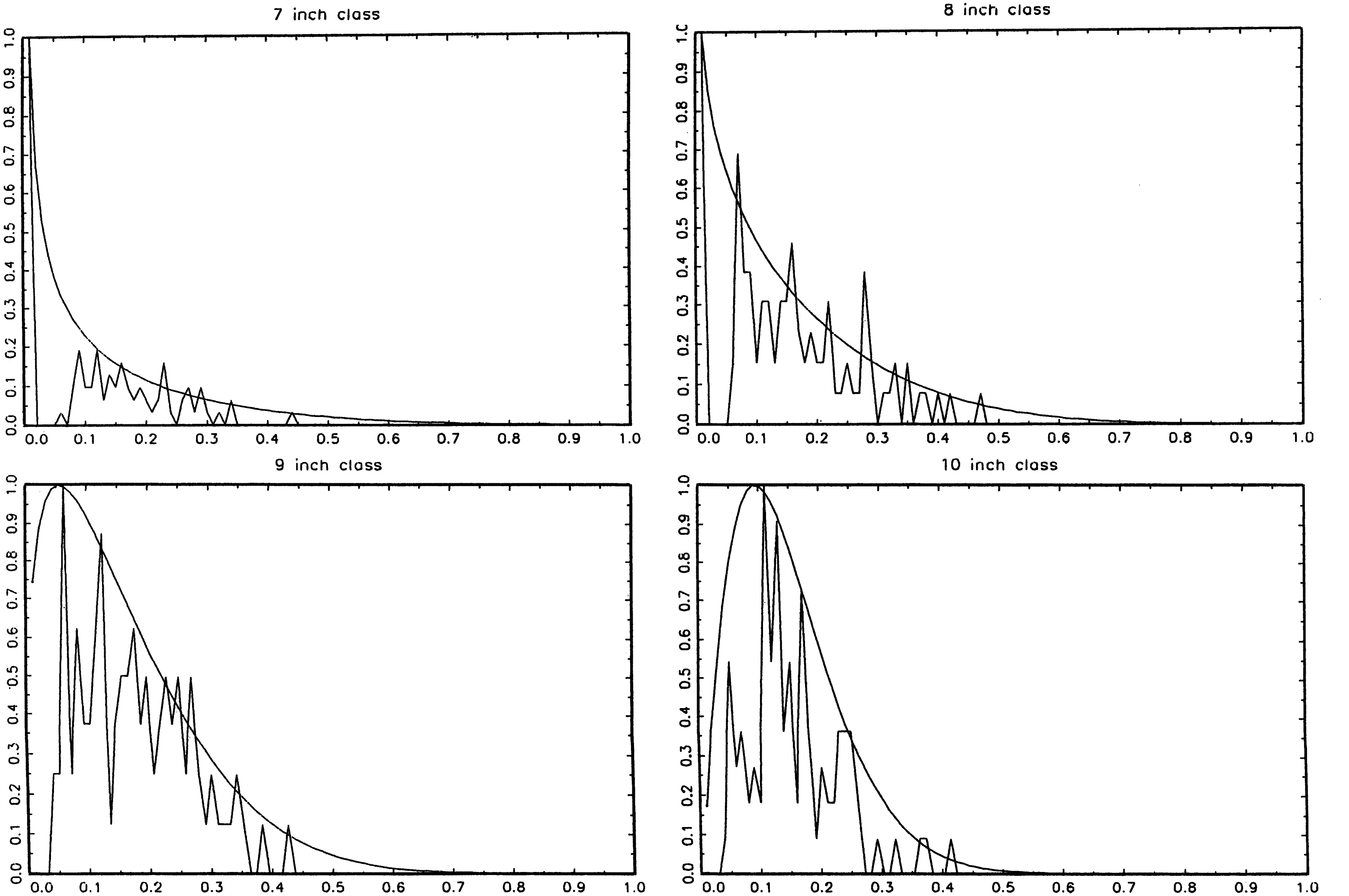

Fig. 1. Scaled relative frequencies (jagged lines) and fitted beta distributions for four representative diameter classes.

Table 2
Mean difference, mean absolute difference and mean squared difference over all alternative stands by estimator and number of sample points in alternative stand (n_a)

Estimator	Mean difference[a]	Mean absolute difference[a]	Mean squared difference[a]
$n_a = 5$			
Usual	−0.009	16.0	813.6
Composite	−0.008	14.3	654.7
Pseudo-Bayes	−0.040	12.1	442.2
$n_a = 10$			
Usual	−0.024	16.3	850.7
Composite	−0.328	12.2	467.7
Pseudo-Bayes	−0.073	10.6	333.0
$n_a = 25$			
Usual	−0.026	16.0	818.7
Composite	−0.446	9.5	272.2
Pseudo-Bayes	−0.084	8.8	223.1
$n_a = 50$			
Usual	−0.138	16.0	825.5
Composite	−0.390	8.2	205.6
Pseudo-Bayes	−0.085	7.8	175.8

[a] Trees/ha per day-class.

Superimposed on these are suitably scaled beta distributions, fitted by maximum likelihood to the diameter class percentage data (smooth lines). The results indicate that the beta distribution was a reasonable prior distribution for the percentage of trees in a diameter class.

3. Results

The average error, average absolute error and mean squared error (averaged over stands and simulations) are presented for each estimator in Table 2. Several trends are apparent from Table 2. First, the usual estimator is generally superior to the other two in terms of mean difference. Hence, if an unbiased estimator is strictly required, then neither $\hat{N}'_k$ nor $\hat{N}''_k$ should be used. However, if biased estimators are permitted, it appears that both $\hat{N}'_k$ and $\hat{N}''_k$ are substantially superior to $\hat{N}_k$. Furthermore, $\hat{N}''_k$ was superior to $\hat{N}'_k$ under mean absolute error and mean squared error for all simulations.

In order to examine the performance of $\hat{N}'_k$ and $\hat{N}''_k$ in relation to the attributes of the alternative stand, mean squared error for each estimator is plotted against age, density and site quality of the alternative stand in Fig. 2. The results are plotted separately for $n_a = 5$, 10, 25 and 50. As expected, for a given n_a, the performance of both $\hat{N}'_k$ and $\hat{N}''_k$ improves as the attributes of the alternative stand are closer to those of the subject stand. The difference between $\hat{N}'_k$ and $\hat{N}''_k$ decreases as n_a increases, but $\hat{N}''_k$ is always superior to $\hat{N}'_k$.

4. Discussion

The fact that $\hat{N}'_k$ and $\hat{N}''_k$ seemed to perform better than $\hat{N}_k$ in terms of squared and absolute error was expected. Many comparisons of standard estimators with composite estimators and/or Bayesian estimators have found the same result (e.g. see Burk et al., 1981; Green and Strawderman, 1986, 1992 for similar results in the forestry literature). However, the superiority of $\hat{N}''_k$ over $\hat{N}'_k$ was somewhat surprising. We attribute this result to two factors: first, the variance of the number of trees per hectare from point samples can be very large, especially in the smaller diame-

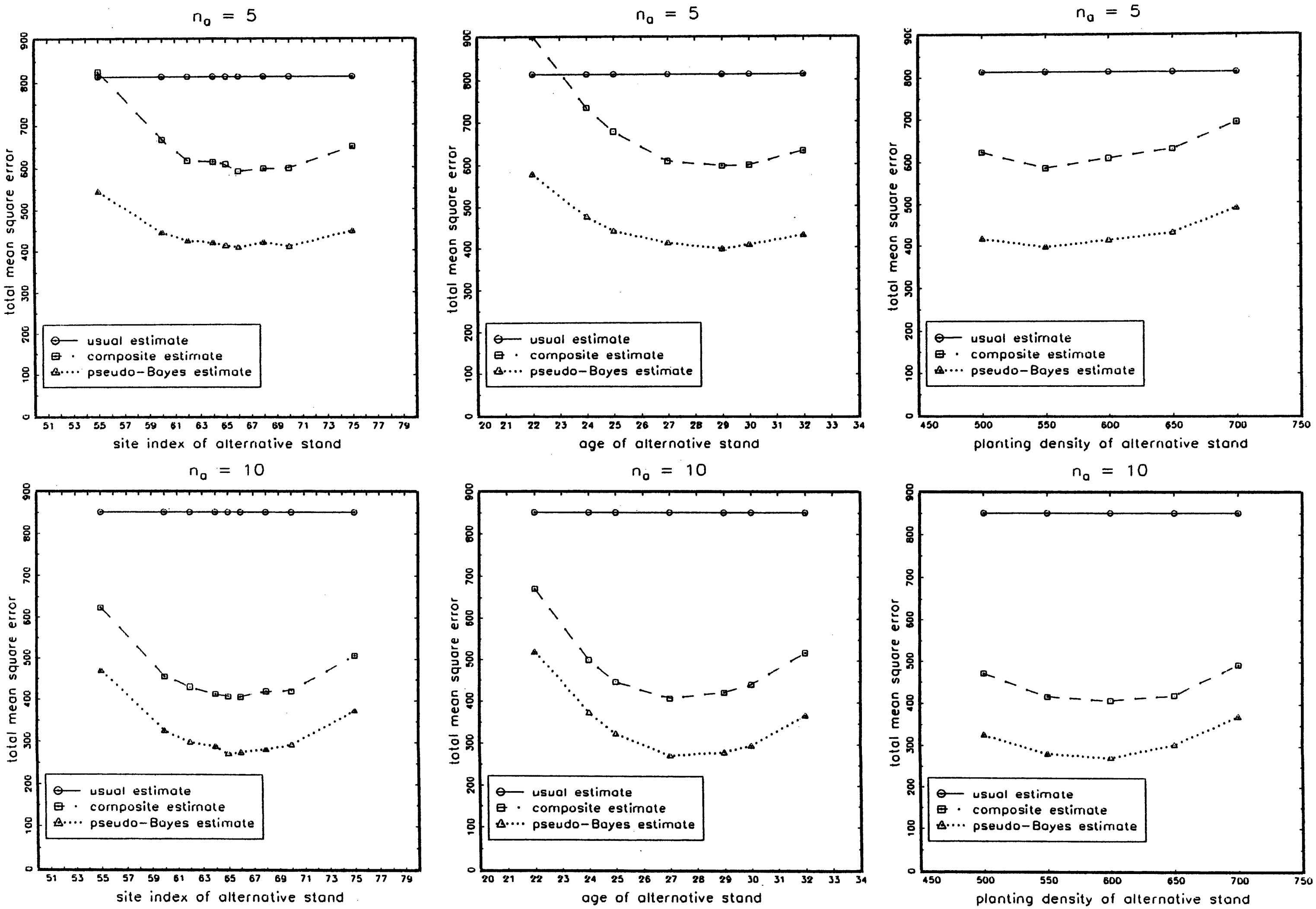

Fig. 2. Mean squared error over all diameter classes versus alternative stand attributes for each estimator, by alternative stand sample size (n_a).

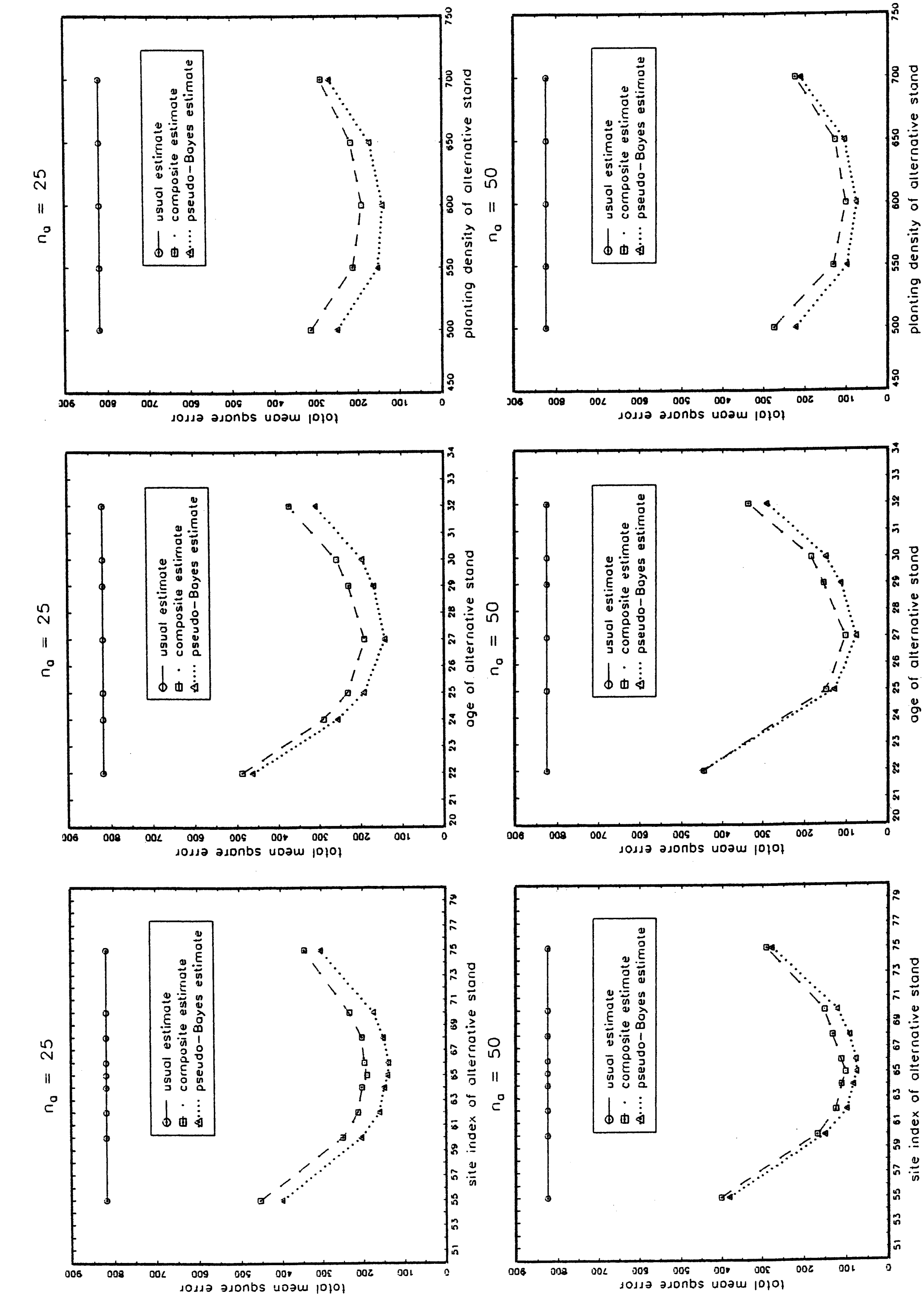

Fig. 2. (*Continued*)

ter classes, where each sample tree represents many trees per hectare. Large variances for $\hat{T}'_k$ (as one would expect with small n_a) force $\hat{N}'_k$ to approach $\hat{N}_k$. This does not occur as dramatically with $\hat{N}''_k$ because with the latter estimator, weights are determined based only on number of points. Second, the pseudo-Bayesian version of the standard multinomial-Dirichlet model appears superior to the composite estimator model. In the latter, the number of trees per diameter class is estimated independently for each class. In the former, the probabilities of class membership are estimated simultaneously (and constrained to equal 1). These probabilities are then multiplied by an estimate of the total number of trees per hectare. It seems intuitive that the total number of trees per hectare should be a more stable estimate than the number of trees per hectare by diameter class. This stability results in enhanced performance of $\hat{N}''_k$.

Instead of using the ad-hoc pseudo-Bayes estimator presented here, it might be advisable to derive a proper Bayesian estimator by specifying a different likelihood function. For instance, since the z_k's are percentages, a natural choice might be a Dirichlet likelihood, resulting in a Dirichlet–Dirichlet model. However, solving that model for the posterior expectation of p would be challenging and probably would entail the use of Markov Chain Monte Carlo approaches. This would not be acceptable, as this estimation procedure is one that must be performed very often, in real time, by non-statisticians.

5. Conclusions

If one is constrained to use only unbiased estimators, then the usual estimator of trees per hectare by diameter class is recommended. However, if one relaxes this restriction, it is possible to achieve dramatic savings in squared or absolute error. The pseudo-Bayes estimator seems to perform better than the composite estimator and thus, would be the recommended choice. The former has the additional benefit that it does not require estimation of the variances of the stand table estimates (this is why we prefer Eq. (11) to a precision-weighted estimate of N). This is important because many potential users may not have the variances available on a routine basis. Hence, use of the composite estimator might require revision of databases, while use of the pseudo-Bayes estimator would not.

References

Avery, T.E., Burkhart, H.E., 1994. Forest Measurements, fourth ed. McGraw-Hill, New York, p. 408.

Burk, T.E., Hansen, M.H., Ek, A.R., 1981. Combining sources of information for improved in-place inventory statistics. In: Brann, T.B., House, L.O. IV, Lund, H.G. (Eds.), In-Place Resource Inventories: Principles and Practices—Proceedings of a National Workshop, pp. 413–420 Soc. Am. For. Publ. No. 82-02.

Farrar, K.D., Amateis, R.L., Burkhart, H.E., Daniels, R.F., 1987. PTAEDA2, Version 1: Simulation of Individual Tree Growth and Stand Development in Loblolly Pine Plantations on Cutover, Site Prepared Areas. School of Forestry and Wildlife Resources, VPI and SU, Blacksburg, VA.

Green, E.J., Strawderman, W.E., 1986. Reducing sample size through the use of a composite estimator: an application to timber volume estimation. Can. J. For. Res. 16, 1116–1118.

Green, E.J., Strawderman, W.E., 1992. A comparison of hierarchical Bayes and empirical Bayes methods. For. Sci. 38, 350–366.

O'Hagan, A., 1994. Kendall's advanced theory of statistics. In: Bayesian Inference, vol. 2B. Edward Arnold, London, p. 330.