Accounting for surveyor inconsistency and bias in estimation of tree density from presettlement land survey records

Barry J. Kronenfeld and Yi-Chen Wang

Abstract: Presettlement land survey records provide baseline data on forest characteristics prior to major European settlement, but questions regarding surveyor bias and methodological consistency have limited confidence in quantitative analyses of this important data source. We propose new correction factors, calculated from bearings, distances, and species of bearing trees, to account for the effects of (*i*) inconsistency in quadrant configuration, (*ii*) bearing angle bias, and (*iii*) species bias on forest density and species composition estimated from presettlement land survey records. Computer simulations confirmed accuracy in random and nonuniform density forests, with moderate bias in very clustered and dispersed forests. A case study of township and quarter-section corners surveyed by the Holland Land Company in western New York demonstrates the potential magnitude of errors caused by surveyor inconsistency/bias in estimation of density and relative species frequency. The influence of nonuniform density, clustering, and dispersal on plotless density estimators remains an important obstacle to quantitative analysis of Presettlement land survey records. However, by accounting for uncertainties regarding surveyor methodology, the proposed correction factors add confidence to conclusions made regarding presettlement forest structure and composition.

Résumé : Les relevés d'arpentage antérieurs à la colonisation fournissent des données de base sur les caractéristiques de la forêt qui existait avant l'arrivée massive des colons européens. Mais les doutes concernant le biais et la consistance méthodologique des arpenteurs ont limité la confiance dans les analyses quantitatives de cette importante source de données. À partir de l'azimut, de la distance et de l'espèce des arbres témoins, nous proposons de nouveaux facteurs de correction pour tenir compte des effets (*i*) de l'inconsistance dans la configuration du quadrant, (*ii*) du biais relié à l'angle de relevé et (*iii*) du biais relié à l'espèce sur l'estimation de la densité de la forêt et de la composition en espèces à partir des relevés d'arpentage antérieurs à la colonisation. Des simulations sur ordinateur confirment la justesse de l'approche pour les forêts dont la densité est aléatoire et hétérogène et révèlent des biais modérés pour les forêts dispersées et fortement agrégées. Une étude de cas, où les angles de township et de quarts de sections ont été relevés par la compagnie Holland Land dans l'ouest de l'État de New York, démontre l'ampleur potentielle des erreurs causées par le biais et l'inconsistance des arpenteurs dans l'estimation de la densité et de la fréquence relative des espèces. L'influence de l'hétérogénéité de la densité, de l'agrégation et de la dispersion sur les estimateurs de densité en l'absence de placettes demeure un obstacle important pour l'analyse quantitative des relevés d'arpentage antérieurs à la colonisation. Cependant, en tenant compte des incertitudes concernant la méthodologie des arpenteurs, les facteurs de correction proposés augmentent la confiance dans les conclusions qu'on peut tirer au sujet de la structure et de la composition des forêts qui existaient avant la colonisation.

[Traduit par la Rédaction]

Introduction

Presettlement land survey records (PLSRs) provide significant historical ecological data that can be used to reconstruct forest structure and composition prior to large-scale disturbance by European settlement in North America. Early surveyors blazed marks on between two and four trees near posts marking survey boundaries to facilitate their rediscovery. In written notes, they provided information on each tree including species, diameter, and distance and bearing relative to the post. The term "witness tree" is used generally to denote these blazed trees, while "bearing tree" refers to any witness tree whose bearing and distance from the post were recorded (Whitney and DeCant 2001).

PLSRs are often used as a baseline to compare with data on modern forest conditions obtained from field surveys (Janke et al. 1978; Palik and Pregitzer 1992; Cowell 1995) and identify changes in forest structure and composition that have occurred since major European settlement (Jackson et al. 2000; Dyer 2001). Large-scale forest inventories are allowing such comparisons to be made over larger areas than was previously feasible (Frelich 1995; Radeloff et al. 1999; Friedman and Reich 2005). Comparisons are usually quantitative in nature, with summary tables pre-

Received 10 January 2007. Accepted 27 March 2007. Published on the NRC Research Press Web site at cjfr.nrc.ca on 28 November 2007.

B.J. Kronenfeld.[1] Department of Geography, George Mason University, 4400 University Drive, MS 1E2, Fairfax, VA 22030, USA.
Y.-C. Wang. Department of Geography, National University of Singapore, 1 Arts Link, 117570, Singapore.

[1] Corresponding author (e-mail: bkronenf@gmu.edu).

doi:10.1139/X07-068

sented comparing percentages of species present at different time periods (Siccama 1971) and maps illustrating spatial changes in species abundance (Friedman and Reich 2005) and distributions of vegetation communities (White and Mladenoff 1994).

Quantitative comparison is hampered by uncertainty regarding how surveyors selected bearing trees, however, and this presents one of the biggest obstacles to the use of PLSRs for ecological reconstruction (Wang 2005). The problem may be most pronounced in the early metes-and-bounds and private land surveys conducted in the eastern United States in which surveyors' instructions were often unavailable (Wang 2004), witness tree densities varied across different physiographic regions and landforms (Black and Abrams 2001), and observations of tree diameters were not made (Siccama 1971). Standardization of survey methods began prior to the establishment of the General Land Office (GLO) in 1812 but was far from comprehensive. GLO rules changed often (Bourdo 1956), and in any case were not specific enough to dictate the precise tree(s) that should be selected around a given post. Further compounding this uncertainty is the belief that early surveyors may have been biased towards specific species or diameter classes of trees or in selecting trees in certain geometric arrangements in relation to the survey post (Grimm 1984).

Uncertainty regarding the bearing tree selection process has resulted in caution and a tendency toward qualitative, rather than quantitative, interpretation of PLSRs. Researchers have suggested that comparisons with modern surveys should be made only at the landscape scale (Manies and Mladenoff 2000; Wang 2005) and warned against direct comparison between bearing tree diameters and tree diameters in modern forests (Almindinger 1997). Grimm (1984) suggested that incomplete knowledge of surveyor methods precluded quantitative comparison with modern inventories. Lack of confidence in quantitative analyses is regrettable because PLSRs arguably contain the most comprehensive data available regarding ecological conditions prior to major European settlement.

Where written documentation of surveyor methods is incomplete or inconsistent, the data themselves may provide information that, when properly analyzed, will reveal the idiosyncrasies and biases of the surveyors. Rules about dividing the area around the post into quadrants or other sections, habits concerning the angular location of bearing trees, and bias towards or against particular species will all leave their trace in the statistical properties of the bearing tree data. If detected, this information can be used to reconstruct surveyor methods and, by extension, the presettlement forest in which they worked. The idea is not a new one. Bourdo (1956) provided basic methods for determining the sampling scheme used in the Michigan survey, Manies et al. (2001) searched for and detected biases of individual surveyors in Wisconsin, and Mladenoff et al. (2002) used logistic regression to probabilistically classify ambiguously identified trees to species level. The concept has not been fully exploited, however, to account for such biases in estimating important forest parameters.

We set out to develop statistical techniques to account for surveyor bias in estimating tree density and species composition. These techniques differ from previous methods that use χ^2 analysis (Bourdo 1956) and ANOVA (Delcourt and Delcourt 1974) to look for statistically significant biases but do not provide means for correcting for bias. Tests of statistical significance are conservative in nature and may fail to find real biases that influenced bearing tree selection, even if only slightly. An alternative approach is to quantify the maximal effect of bias on estimation of important forest characteristics and thereby place bounds on the likely values of these parameters.

Bearing tree selection

PLSRs can be categorized into metes-and-bounds, private land surveys such as the Holland Land Company surveys in western New York, and public GLO surveys, each with different data characteristics (Wang 2005). Distances and bearings were not generally recorded in the early metes-and-bounds surveys, which were also irregularly shaped; for these, contingency table methods can be used to establish significant differences in relative species frequencies across physiographic or other units (e.g., Abrams and McCay 1996), but absolute frequencies cannot be estimated. In the present study, we restrict our attention to private land surveys and public GLO surveys, the PLSRs in which distances and bearings were recorded and the division of the land was regular. In these surveys, land was typically divided into townships of 6 × 6 miles (1 mile = 1.6 km) and then subdivided into sections or lots of various sizes, of which 1 × 1 mile was most commonly used. Posts were erected at survey corners, including township corners, section corners (the intersection of section lines), and quarter corners (the midpoint between section corners). At these survey corners, bearing trees were blazed and their distances and bearings from the designated posts were recorded.

Density can be estimated from these PLSRs by applying statistically derived formulas (Cottam and Curtis 1956; Pollard 1971) to the measured distances from survey corners to bearing trees. These formulas presume an intentional sampling method scheme used by surveyors, but in reality, PLSRs represent an unintentional vegetation survey (Black and Abrams 2001). Extant written instructions provide some clues regarding the bearing tree selection process but often are not conclusive. GLO manuals specify two to four trees to be selected, depending on the corner type and year of publication (Bourdo 1956). Whenever four trees are to be marked, the instructions specify selection of one tree in each of four adjacent sections, consistent with the point-quarter sampling of Cottam and Curtis (1956). The existence of a rule, however, does not imply its consistent application. Manies et al. (2001) tested for individual surveyor biases towards or against specific quadrants in northern Wisconsin, suggesting that not all surveyors conformed to the one-tree-per-quadrant rule. In data from the Holland Land Company in western New York, we found that bearing trees deviated from the rule at 6% of township corners.

In some GLO section corners and most quarter corners, only two trees were marked as bearing trees. Possible configurations for two trees include opposite quadrants, adjacent quadrants, and opposite sides of the survey line, all of which appear in GLO instructions at various times (Bourdo 1956). It is often difficult to trace down the rule specified for a particular survey, however, and rules were often changed or tinkered with in the field. For this reason, Bourdo emphasized the importance of statistical analysis and provided simple methods to test for common configurations.

In addition to inconsistencies in quadrant configuration rules, density estimation will also be affected if surveyors favored any subset of trees over any other. Surveyor bias implies that the selected bearing trees were not always the closest to the corner, with the result that recorded distances will be inflated. Several possible types of surveyor bias have been identified, including bearing angle (Manies et al. 2001), species (Almindinger 1997), and diameter (Bourdo 1956).

Preference in the angular placement of bearing trees in relation to survey posts was identified in northern Wisconsin by Manies et al. (2001), who found significantly fewer trees with bearings near to the cardinal directions. Preference for particular bearings may be the result of explicit instructions: the GLO manuals for 1815 and 1855 instructed surveyors to select two trees "in opposite directions, as nearly as may be" (Bourdo 1956, p. 757). Alternatively, it may be that surveyors, when presented with instructions to select trees by quadrant, would naturally favor the interior angles and avoid trees near the survey lines.

Species bias has been widely speculated upon in PLSR research (Almindinger 1997). Although no written instructions have been found instructing surveyors to preferentially select or avoid specific species, the 1843 GLO instructions do allude to the possibility, instructing surveyors to "...select for bearing trees those which are the soundest and most thrifty in appearance, and of the size and kinds of trees experience teaches will be the most permanent and lasting..." (White 1984, p. 333). Bourdo (1956) suggested that surveyors would have preferred species with thin barks that were easy to blaze and inscribe as well as uncommon species that would be more easily spotted during resurvey. Grimm (1984) also suggested that species longevity and conspicuousness in the stand would have been important.

Previous studies have attempted to identify species bias by calculating the average distance from post to bearing trees of each species over a large area (Almindinger 1997). It is widely recognized that nonuniform density will invalidate such analysis (Grimm 1984); to mitigate this problem, the study area is often stratified into physiographic regions. Although several studies have failed to find statistically significant species bias (Bourdo 1956; Delcourt and Delcourt 1974), small sample sizes in these studies might preclude the detection of slight biases. Another reason why species biases may be difficult to detect is that they may be inconsistent. Manies et al. (2001) found significant bias for individual surveyors, but each surveyor preferred different species.

Surveyors were likely to avoid very small tree species, which have high mortality rates and for which blazing is more likely to cause death (Nelson 1997). The GLO instructions in 1846 and 1851 required that the trees be not less than 5 inches in diameter (Bourdo 1956). However, such instructions do not imply that all trees greater than the specified cutoff would have been considered equal candidates for selection. Manies et al. (2001) found that individual surveyors consistently favored either smaller or larger diameters and that these preferences held irrespective of species.

Thus, historical evidence and recent studies suggest at least three aspects of bearing tree selection that might significantly impact density estimation from PLSRs: (*i*) quadrant configuration variability, (*ii*) bearing angle bias, and (*iii*) species bias. The latter is also important in its own right, as estimates of species composition are commonly used to describe the presettlement forest. These three sources of inconsistency/bias add to the uncertainties that already exist in density estimation from point-to-point distances due to the effects of nonuniform density (Pollard 1971) and patterns of clustering and dispersal (Pielou 1959; Steinke and Hennenberg 2006). The next section reviews the most relevant density estimation formulas.

Density estimation

Density (λ) is the number of individuals per unit area. Estimating λ from fixed-area plots is unproblematic, but it can also be estimated from measurements of distances either from one tree to another or from a random point in the forest to a set of nearby trees (Cottam and Curtis 1956; Barbour et al. 1999). The latter concept is applicable to PLSRs, with the implicit assumption that survey post locations were random with respect to the nearby trees. Distance-based density estimators assume uniform density and random locations of individuals without clustering or dispersion (i.e., regular spacing); given these assumptions, density can be derived theoretically. Let $R = \{r_1, r_2, ..., r_n\}$ be a set of distances to n trees from their corresponding corners. One may conceptualize the area occupied by each tree as a function of the square of the inverse of its radial distance; the mean of this area for all trees will be inversely proportional to density. This is the mean area (μ_R) of Cottam and Curtis (1956), which we define formally as

$$[1] \qquad \mu_R = \pi \sum_{i=1}^{n} r_i^2/n$$

Suppose that at each of c corners the surrounding forest was divided into b equiangular sections and the nearest tree chosen in each section (so that $bc = n$). For $b = 4$, the sampling scheme is point-quarter, and Cottam and Curtis (1956) proposed using the inverse of μ to estimate density. This estimator is biased, and although the bias is small for large values of n, it is easily corrected. Pollard (1971) gave the following unbiased estimator for point-quarter sampling:

$$[2] \qquad E(\lambda) = (n-1)/c\mu$$

The estimator has a known variance:

$$[3] \qquad \mathrm{var}(E(\lambda)) = \lambda^2/(n-2)$$

One may wish to estimate density using trees other than the nearest. Further denote k_i is the distance rank of the ith tree. Pollard (1971) provided the following density estimator estimating density when each tree is the kth nearest to a random point:

$$[4] \qquad E(\lambda) = \left(\sum_{i=1}^{n} k_i - 1\right)/n\mu$$

which has a variance of

$$\text{[5]} \qquad \mathrm{var}(E(\lambda)) = \lambda^2 / \left(\left(\sum_{i=1}^{n} k_i \right) - 2 \right)$$

Equations 2–5 assume density to be uniform across the study area and further that the spatial pattern is random (i.e., not clustered or dispersed). Violation of either of these conditions will bias the results, and this bias can be quite large. The effects of nonuniform density can be quantified if the densities and areal proportions of each subregion are known (Pollard 1971; Jost 1993). Jost (1993) provided an unbiased estimator for point-quarter sampling that is valid in conditions of nonuniform density but gives no variance for this estimator, nor a solution to the more general cases of b equiangular divisions.

The influence of clustering and dispersal on density estimation is harder to quantify, since patterning can occur at different scales. In general, clustered populations will result in larger point-to-plant distances, and thus, density will be underestimated; the converse is true for dispersed populations. This effect is well known, and indeed, Pielou (1959) used the relationship between density and mean area as an index of clustering/dispersal. The difficulty lies in determining both values simultaneously from distance measurements. Batcheler (1971) reviewed several attempts to do this, including his own, but none have proven robust nor are they applicable to the point-quarter sampling method, which is predominant in PLSRs.

Despite the well-known problems noted above, we use Pollard's (1971) estimators throughout because they are well understood, general, and unbiased in random forests. Although the formula of Jost (1993) shows promise, to date no estimator has been found to be robust to all types of spatial patterning. The methods developed here to correct for surveyor inconsistency and bias could in principle be modified to relate to any such estimator if it exists.

Correction factors for surveyor inconsistency and bias

Although bias due to spatial patterning of individual plants is an important problem, in the case of PLSRs, this is further compounded by uncertainty regarding surveyor methodology. We set out to develop statistical methods to detect and correct for error due to the three types of inconsistency and bias identified above. For each source of error, a correction factor is presented that can be computed from distances, bearings, and species of bearing trees. Each correction factor takes the form

$$\frac{E'(\lambda)}{E(\lambda)}$$

where $E(\lambda)$ is an appropriate density estimate for point-quarter sampling and $E'(\lambda)$ is an adjusted estimate after accounting for surveyor inconsistency or bias. The following variables are defined: c is the number of corners, b is the number of bearing trees recorded at each corner (= $c \times b$), n is the total number of bearing tress, and μ is the mean area of trees at all corners. Equation 2 is used as the base estimator $E(\lambda)$ throughout.

Quadrant configuration inconsistency

Suppose that surveyors normally selected one bearing tree in each of four quadrants but occasionally lapsed and simply selected the four nearest trees to a corner regardless of quadrant. The proportion of corners at which bearing trees conform to point-quarter sampling can be determined from the data. These will be referred to as "conforming" corners. Denote p as the proportion of corners conforming to point-quarter sampling (i.e., one tree per quadrant), μ_q as the mean area of trees at conforming corners, and ϕ as the probability of four nearest trees occurring in four quadrants due to chance. The parameter ϕ is used to determine the number of "false conformities", i.e., corners at which surveyors did not follow the point-quarter sampling method, but bearing trees are located in different quadrants due to chance. In a random forest, ϕ = 3/32. To derive the correction factor, a weighted average is taken of the density estimate based on point-quarter sampling and that of nearest k tree sampling. Combining eqs. 2 and 4, the following correction factor can be derived algebraically:

$$\text{[6]} \qquad \frac{E'(\lambda)}{E(\lambda)} = \frac{(p-\phi)^2}{\left(p\frac{\mu_q}{\mu} - \phi\right)(1-\phi)} + \frac{c(b+1)}{2(cb-1)} \frac{(1-p)^2}{\left(1 - p\frac{\mu_q}{\mu}\right)(1-\phi)}$$

A detailed derivation is provided in Appendix A. Since the correction factor is derived by averaging density estimates from conforming and nonconforming corners, it is robust to systematic variations in density between these sets of corners. That is, it will remain accurate even if surveyors abandoned the point-quarter scheme more often in sparse (or dense) forests.

Incidentally, the correction factor may be used with any b-tree sampling scheme consisting of selecting the nearest tree in each of b equiangular sections. Specifically, it may be also applied to the case of selection of two trees on opposite sides of the survey line. In this case, the appropriate value of ϕ in a random forest is 0.5, since that is the probability that the two nearest trees will be located opposite each other.

Bearing angle bias

Suppose that surveyors consistently avoided certain bearing angles, for example by avoiding trees near the survey line. If the probability of avoidance is related only to bearing angle, the effect on density is easily determined. Consider the maximum angular section in which trees were not avoided; the relative proportion of bearing trees occurring in this section will be greater than for any other section. Denote α as the angular proportion of section with highest proportion of bearing trees and p as the proportion of bearing trees in above section. The set of trees contained in the angular section α can be conceptualized as an unbiased sample of a proportion α of the entire forest. This leads intuitively to a correction factor of

$$\text{[7]} \qquad \frac{E'(\lambda)}{E(\lambda)} = \frac{p}{\alpha}$$

Further details are provided in Appendix A. The correction factor will always be greater than 1. Some caution is required to avoid too great of a correction, since one can always find an arbitrarily narrow section that contains an abnormally large number of bearing trees. To be prudent, the circle may be divided into a large number of small angular sections. The sections with the smallest ratio of actual to expected bearing tree counts are eliminated sequentially until χ^2 analysis shows the count distribution of the remaining sections not to differ significantly from random.

Species bias

A model of species bias can be created by defining the "perceived distance" of each candidate tree of species j as a function of the true distance multiplied by a species bias parameter. Define s as the number of distinct species, p_j as the observed survey frequency of species j:

$$\sum_{j=1}^{s} p_j = 1$$

ϕ_j as the actual frequency of species j:

$$\sum_{j=1}^{s} \phi_j = 1$$

and μ_j as the observed mean area of bearing trees of species j. If surveyors select bearing trees with the lowest perceived distance as defined above, the mean area of each species will be overstated to the same degree that its relative frequency is understated (or vice versa). Given mean areas and relative frequencies of each species in the survey data, the mean areas and relative frequencies that would result from unbiased sampling can be estimated, resulting in the following correction factor:

$$[8] \qquad \frac{E'(\lambda)}{E(\lambda)} = \mu \sum_{j=1}^{s} \left(\frac{p_j}{\mu_j} \right)$$

A detailed derivation is provided in Appendix A. Interestingly, the solution also yields estimates of the actual relative frequencies of each species:

$$[9] \qquad \phi_j = p_j/\mu_j \sum_{j=1}^{s} \left(p_j/\mu_j \right)$$

Two words of caution are in order. First, the use of tree distances and mean areas to account for species bias requires the assumption that each species is found in forests of the same overall density. If some species occur naturally in denser forests than do other species, differences in mean areas will be observed even if there is no surveyor bias.

A test for species bias that is robust to nonuniform forest density can be developed by comparing average ranks, rather than mean areas, of each species; the null hypothesis is that the average ranks will not differ between species. Rank comparison provides a more robust test for the presence of species bias than distance comparison but cannot be used to quantify its effects. However, comparing average rank with average distance for each species can provide information on the general validity of the species bias correction factor for a particular study region.

Evaluation methods

Randomized forest simulation was used to assess the validity of each of the above correction factors as well as their robustness to nonuniform density and nonrandom spatial patterns. We then applied the factors to PLSR data for western New York for further validation and to illustrate appropriate usage and interpretation.

Simulation trials

Forest simulation was performed in Microsoft Visual Basic 6.0 using a custom program created by the authors (available upon request). At each simulated survey corner, a predetermined number of trees (usually 100) were placed on a square region with the survey corner located in the center and the distance, bearing, and species of the nearest tree in each quadrant were identified and recorded. These represented the ideal bearing trees that would be selected absent of surveyor inconsistency/bias. The trees that would be selected by surveyors under various scenarios of surveyor inconsistency and bias were also identified and their distances, bearings, and species recorded. At many simulated corners, the ideal bearing trees coincided with those selected by the surveyors, but at others, they did not. For all trials, distances, bearings, and species of both ideal and selected trees were recorded in a database and an initial density estimate was computed using eq. 2. For the surveyor-selected bearing trees, an adjusted estimate was derived by calculating the appropriate correction factor using eq. 6, 8, or 11 and multiplying this by the initial estimate.

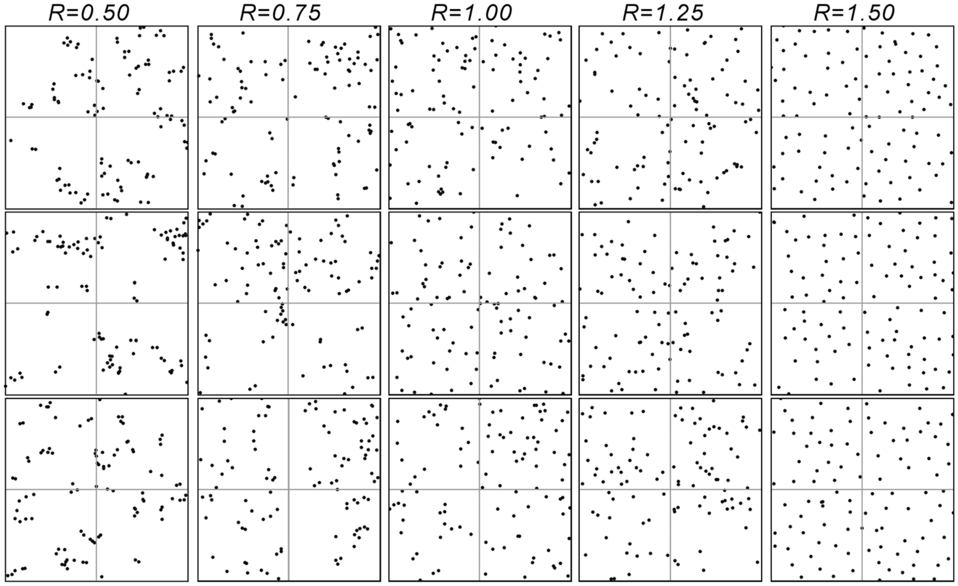
Three groups of simulations were performed to assess (*i*) validity of the correction factors under the assumption of complete spatial randomness, (*ii*) effects of nonuniform density, and (*iii*) effects of clustering and dispersal.

Each trial in the first group consisted of 30 sets of 100 simulated corners. The number of trees placed at a given corner was drawn from a Poisson distribution with mean 100, and each tree was placed by selecting random coordinates in the X and Y directions. Density estimates and correction factors were calculated individually for each set of 100 corners, and means and standard deviations were then computed.

For the second and third simulation groups, each trial consisted of a single set of 10 000 simulated corners. In the second group, nonuniform density was simulated by distinguishing two types of forest with different densities. Four density ratios ($\lambda 1{:}\lambda 2$) were used: 1:1.5, 1:2, 1:2.5, and 1:3. In each case, 100 trees were placed around corners in the first type of forest so that 150–300 trees were placed around corners in the second type. The forest types were assumed to be equally distributed so that 5000 corners were located in each forest type.

In the third simulation group, clustered and dispersed forests were simulated via probabilistic replacement of random points to achieve a target nearest-neighbor distance, measured using the nearest-neighbor statistic (R) of Clark and Evans (1954). To avoid edge effects, distances were measured on a torus during point creation, and Donnelly's (1978) correction was applied to the computed value of R. The theoretical range of R is from 0 (maximally clustered) to 2.13 (maximally dispersed); simulated values ranged

Fig. 1. Simulated clustered, random, and dispersed distributions with corresponding values of the nearest-neighbor statistic (*R*).



from 0.5 to 1.5 in increments of 0.25. Figure 1 shows three typical examples each of tree distributions generated by the probabilistic replacement algorithm for each target *R* value.

Within each of the above simulation groups, four sets of trials were conducted: one set of trials without bias to confirm simulation methodology and three sets of trials to simulate the effects of quadrant configuration inconsistency, bearing bias, and species bias.

To simulate the effects of quadrant configuration inconsistency, five trials were conducted with surveyor consistency (i.e., adherence to the point-quarter method) ranging from 90% to 50%. At each plot, a random number was used to determine if the point-quarter method was applied; otherwise, the four nearest trees were selected regardless of quadrant. To test the bearing correction factor, five trials were conducted with a minimum angular tolerance parameter $\alpha_{\min}$ ranging from 5° to 25°. Trees with a bearing angle α less than $\alpha_{\min}$ were rejected with probability $p = \alpha/\alpha_{\min}$. To test the species correction factor, trees were assigned to two hypothetical species with equal probability, and surveyor selection was determined by multiplying distance to trees of the second species by a factor ranging from 1.5 to 3 in four sets of trials.

For ease of reporting, all results were scaled to a fixed "true" density of 100 trees/ha. Since the area implicitly defined by the square region used in simulation is arbitrary, this rescaling does not affect the validity of the results.

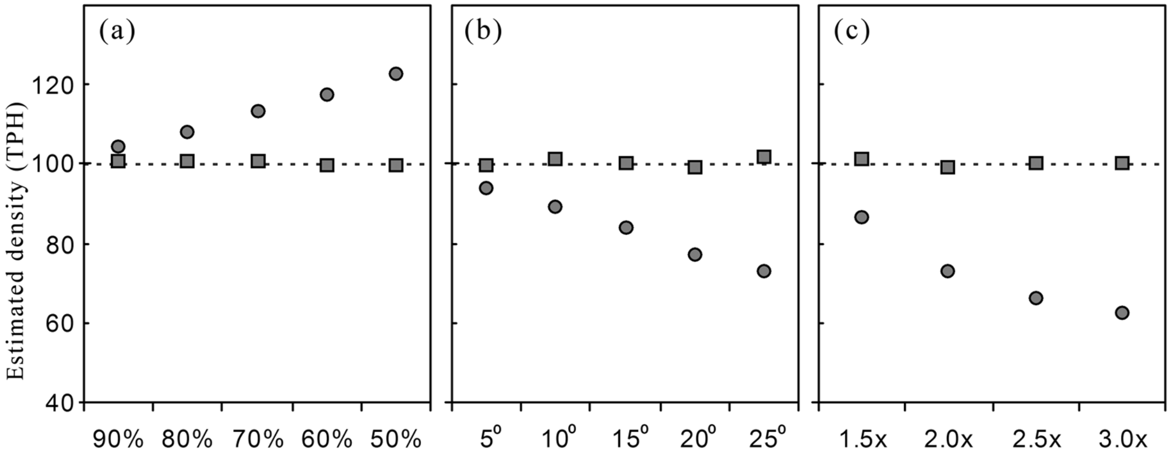
Application to PLSR data

We also applied the correction factors to estimation of forest density in presettlement western New York from the township survey records of the Holland Land Company (HLC). A map of the study region is given in Wang (2005). The HLC township survey was conducted between 1797 and 1799 during the transition from metes-and-bounds to public GLO surveys. As such, it is broadly representative both of early surveys conducted in the eastern United States in which bearing tree diameters were not recorded and of the public GLO surveys conducted from Ohio to California in which a rectangular grid of the township and range system was used. Joseph Ellicott, the chief surveyor, is well respected and known to have enforced a high degree of quality control on his surveyors (Wyckoff 1988). The survey area encompasses 14 250 km^2 and spans the eastern broadleaf and Laurentian mixed-forest provinces (Bailey 1995).

Within the HLC survey, two types of corners were analyzed. At township corners, generally spaced at 6-mile intervals, four bearing trees were selected. We did not analyze section corners, which were located at 1-mile intervals between township corners because selection of three bearing trees at these corners is rare in other PLSRs. Instead, we used data from quarter corners located halfway between section corners at which two bearing trees were selected. Corners located along the survey boundary, including the Pennsylvania border and Lakes Erie and Ontario, as well as those along the edges of interior Indian reservations, were excluded to avoid edge effects. We further excluded five corners with unusually large bearing tree distances that were located near to survey line descriptions noting the presence of "plains" and "swamps". The final data sets consisted of 135 township corners with 540 bearing trees and 1458 quarter corners with 2916 bearing trees. In testing for species bias, only species recorded 25 or more times in the bearing tree data were analyzed individually to eliminate random effects of small sample sizes; the remaining species were combined and treated as a single group.

The purpose of applying the proposed correction factors to the HLC data was twofold. First, by examining two data sets (i.e., township and quarter corners) from the same survey, some measure of validation is possible. Since the data cover the same geographical region at the same period of time, they should produce similar estimates of overall tree density. Each data set, however, is likely to contain unique biases. It was hypothesized that at the more important township corners, surveyors would have been more diligent, con-

Fig. 2. Simulated effects of surveyor inconsistency/bias on density estimation in random forests: (*a*) inconsistency in point-quarter sampling, (*b*) bearing bias, and (*c*) species bias. Estimates shown before (circles) and after (squares) applying correction factors. The broken line is actual density (100 trees/ha).



forming to the prescribed quadrant arrangement and avoiding trees near the survey lines. The effect on species bias was deemed unpredictable. On the one hand, surveyors may have been more likely to search for sturdy, long-lived species at the township corners. On the other hand, the idiosyncratic preferences of individual surveyors may have flourished at the less prominent quarter corners. Because precise matching to modern-day species nomenclature is sometimes difficult to determine, species' names are given as recorded by the surveyors.

The second purpose of applying the proposed correction factors to the HLC data is to illustrate their proper use. One critical aspect of this is the implementation of methods for testing the underlying assumptions of each factor and subsequent interpretation.

Results

Simulation trials

Complete spatial randomness

Under complete spatial randomness, the correction factors accurately corrected for the effects of quadrant configuration inconsistency, bearing bias, and species bias (Fig. 2). The actual density of 100 trees/ha is represented by the broken line in the figure. Over the 30 simulation trials in each group, inconsistent application of the point-quarter sampling method resulted in density overestimation by 5, 8, 13, 17, and 23 trees/ha for consistency rates of 90%, 80%, 70%, 60%, and 50%, respectively (Fig. 2*a*). Avoidance of trees near the survey line resulted in density underestimation by 6, 11, 16, 2,3 and 27 trees/ha for $\alpha_{\min}$ of 5, 10, 15, 20, and 25, respectively (Fig. 2*b*). Preferential selection of species resulted in density underestimation by 13, 27, 34, and 38 trees/ha for relative bias (i.e., preference of one species over the other) of 1.5×, 2.0×, 2.5×, and 3.0×, respectively (Fig. 2*c*). Standard deviations within each group of 30 trials were between 3 and 8 trees/ha and did not show any trend with the degree of inconsistency/bias.

After applying the corresponding correction factors, adjusted density estimates averaged between 99 and 102 trees/ha for all simulation groups. The correction factor for quadrant configuration inconsistency ranged from 0.97 to 0.81 (±0.01–0.03), bearing bias corrections ranged from 1.06 to 1.38 (±0.02–0.08), and species bias corrections ranged from 1.17 to 1.61 (±0.05–0.14) (results not shown). Although overall accuracy was confirmed, the high standard deviations of the latter two correction factors led to increased estimation variability. When Pollard's estimator (eq. 2) was applied to the bearing trees that would have been selected absent of surveyor inconsistency/bias, the standard deviation of the estimate ranged from 4 to 6 trees/ha. In comparison, standard deviations of the adjusted estimates ranged from 4 to 5 trees/ha for quadrant configuration inconsistency but from 5 to 10 trees/ha for bearing bias and from 7 to 12 trees/ha for species bias.

Table 1. Simulated observed and adjusted relative frequencies of nonpreferred species with species bias under complete spatial randomness (actual relative frequency = 50%).

Species bias	Observed frequency (%)	Adjusted frequency (%)
1.5×	30.7	50.8
2×	20.0	49.7
2.5×	13.9	49.3
3×	10.1	48.7

Table 1 shows simulated observed and adjusted relative frequencies (from eq. 9) for the nonpreferred species (frequencies for preferred and nonpreferred species summed to 100%). As would be expected, simulated species bias resulted in overrepresentation of the preferred species and underrepresentation of the nonpreferred species in surveyor-selected bearing trees. Adjusted estimates of relative frequencies were close to the correct value of 50%.

Nonuniform density forests

Without surveyor inconsistency/bias, nonuniform density caused significant underestimation of density (Fig. 3). Estimated density decreased as nonuniformity increased, declining to 75.2 trees/ha when one forest was three times as dense as the other (λ_1:λ_2 = 1:3). The simulated ratio between estimated and actual density corresponded to theoretical expectations, and sample standard deviations were consistent with standard errors computed from eq. 3.

Performance of the correction factors for selected parameter values (80% quadrant consistency, 15° minimum angular tolerance, 2x× bias for preferred species) is shown for nonuniform density simulations in Fig. 4. In all cases, initial estimates reflected errors from both nonuniform density and surveyor inconsistency/bias. The correction factors accurately captured the effects of surveyor inconsistency and bias but not nonuniform density. The resulting adjusted density estimates were always within 1.1 trees/ha of the results of trials without bias (Fig. 3) (shown as a gray line in Fig. 4). Results were similar for other parameter values. In correcting for species bias, adjusted estimates of the relative frequencies of the preferred and nonpreferred species were generally within 2% of the correct values of 50% (results not shown).

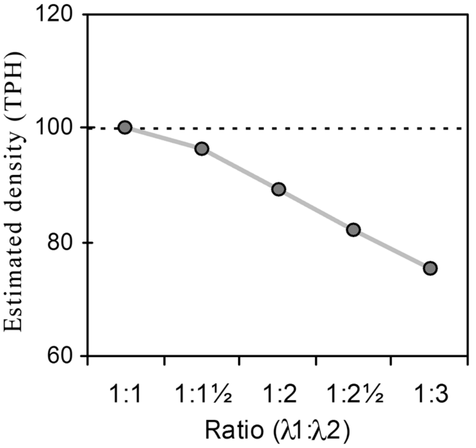
Clustered and dispersed forests

Without surveyor inconsistency/bias, clustering caused significant underestimation and dispersal caused significant overestimation of density (Fig. 5). Density estimates ranged from 49.8 to 156.7 trees/ha in a nearly linear fashion with respect to the nearest-neighbor statistic (R from 0.5 to 1.5).

Performance of the correction factors for selected parameter values (80% quadrant consistency, 15° minimum angular tolerance, 2x× bias for preferred species) is shown for clustered and dispersed spatial patterns in Fig. 6. Again, initial estimates reflected errors from both clustering/dispersal and surveyor inconsistency/bias. Application of the correction factors resulted in adjusted estimates that were similar to the results of trials without bias (Fig. 5) (shown as a gray line in Fig. 6). However, a small but significant interaction effect was evident. For quadrant configuration inconsistency and species bias, clustered patterns resulted in adjusted estimates that were slightly higher, and dispersed patterns resulted in adjusted estimates that were slightly lower, than expected. The reverse was true for bearing bias. The magnitude of the interaction was highest for bearing bias in dispersed patterns and relatively high for quadrant configuration inconsistency and bearing bias in clustered patterns as well as species bias in dispersed patterns.

In correcting for species bias, adjusted estimates of the relative frequencies of the preferred and nonpreferred species were close to the correct values of 50% (Table 2). However, an interaction effect was evident. Clustering enhanced the effect of species bias on simulated surveyor selection, making the preferred species even more likely to be selected, while dispersal had the opposite effect. The errors in adjusted frequency estimates were modest (≤3.2%) for dispersed (R = 1.5, 1.25) and slightly clustered (R = 0.75) spatial patterns but noticeably higher (≤8.8%) for the most clustered simulation group (R = 0.5).

Fig. 3. Simulated density estimates in nonuniform density forests without surveyor bias/inconsistency. The ratio signifies relative densities of two equally distributed forest types. The broken line is actual overall density (100 trees/ha); the solid line is for reference in Fig. 4.



Application to PLSR data

Preserved HLC records contain instructions for surveyors to select one bearing tree in each of the four quadrants at township corners. Of the 135 township corners, 127 (94.1%) conformed to this pattern (p < 0.01). No instructions were found regarding the arrangement of trees at quarter corners, but data analysis revealed that one bearing tree was located on each side of the survey line at 1280 (87.8%) out of 1458 corners. This also is significantly greater than would be expected due to chance (p < 0.01). Bearing trees were located on opposite sides of the perpendicular (i.e., forwards and backwards of the post) only 51.8% of the time (p = 0.08), so it is unlikely that surveyors favored trees in opposite quadrants. It was therefore assumed that surveyors followed a prescribed rule of selecting one bearing tree in each adjacent township.

At township corners, mean area (eq. 1) was 244.9 m^2 for all corners and 245.6 m^2 for corners conforming to the point-quadrat sampling configuration. For quarter corners, mean area was 90.8 m^2 for all corners and 91.7 for corners with one bearing tree on each side of the survey line. Equation 2 yielded estimated densities of 163.0 trees/ha at township corners and 220.2 trees/ha at quarter corners, with standard errors of 0.8 and 0.5 trees/ha, respectively. From eq. 6, the correction factors at township corners was computed as

$$\frac{E'(\lambda)}{E(\lambda)} = \frac{(0.941 - 0.09375)^2}{\left(0.941\,\frac{245.6}{244.9} - 0.09375\right)(1 - 0.09375)} + \frac{5}{8}\,\frac{(1 - 0.941)^2}{\left(1 - 0.941\,\frac{245.6}{244.9}\right)(1 - 0.09375)} \approx 0.975$$

and at quarter corners as

$$\frac{E'(\lambda)}{E(\lambda)} = \frac{(0.878 - 0.5)^2}{\left(0.878\,\frac{91.7}{90.8} - 0.5\right)(1 - 0.5)} + \frac{3}{4}\,\frac{(1 - 0.878)^2}{\left(1 - 0.878\,\frac{91.7}{90.8}\right)(1 - 0.5)} \approx 0.936$$

Greater adjustment was required for quarter corners than for township corners owing to the fact that surveyors deviated more often from the prescribed configuration rule of selecting one tree on each side of the survey line.

Figure 7 shows bearing tree counts by angular section for

Fig. 4. Simulated density estimates in nonuniform density forests with surveyor bias/inconsistency: (*a*) quadrant configuration inconsistency, (*b*) bearing bias, and (*c*) species bias. Estimates shown before (circles) and after (squares) applying correction factors. The broken line is actual density (100 trees/ha); the solid line recreates no-bias estimates from Fig. 3.

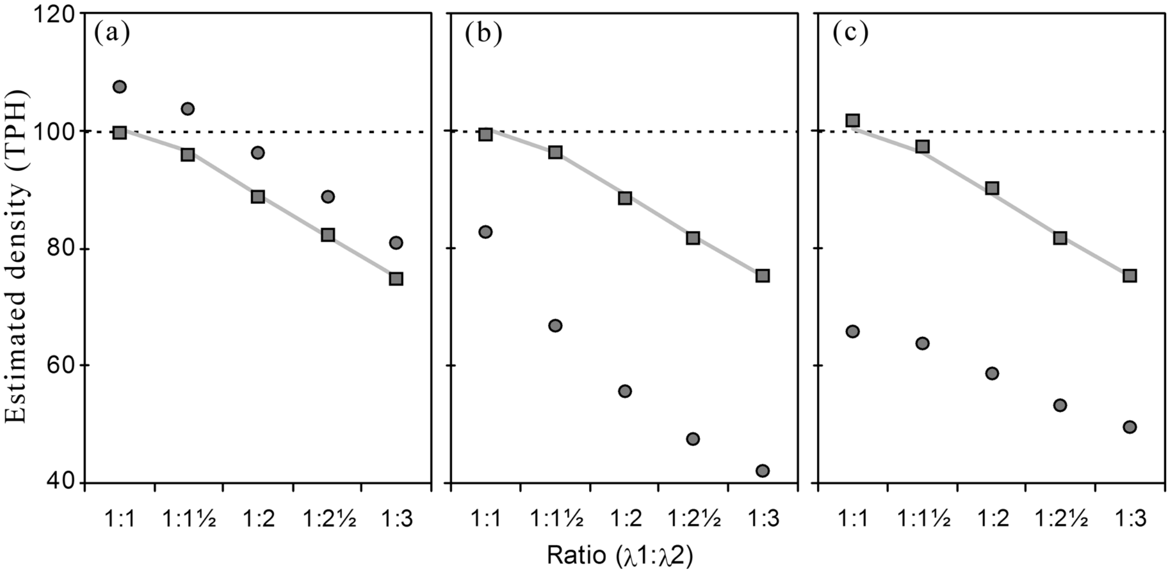
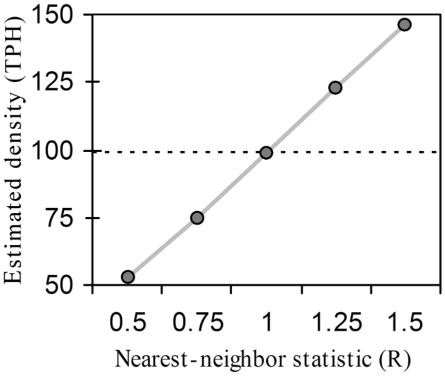


Fig. 5. Simulated density estimates in clustered, random, and dispersed forests without bias. The broken line is actual density (100 trees/ha); the solid line is for reference in Fig. 6.



township and quarter corners. Bearing trees at township corners were grouped into nine angular sections of 5° each, ranging from 0 to 45° from the nearest of the two intersecting survey lines (Fig. 7*a*). After sequential elimination, the count distribution in the five sections ranging from 21 to 45° was determined not to differ significantly from random (Table 3). These sections encompass 54.6% of the circle but contained 74.4% of the bearing trees with a mean area of 241.8 m^2. At quarter corners, trees were grouped into nine angular sections of 10° each according to angle from the survey line (Fig. 7*b*). After sequential elimination, the count distribution of the range 21–60° did not differ significantly ($p > 0.05$) from random (Table 4). This encompassed 44.4% of the circle but contained 53.9% of the bearing trees with a mean area of 90.3 m^2. From eq. 7, the correction factor at township corners was computed as

$$\frac{E'(\lambda)}{E(\lambda)} = \frac{0.744}{0.546} = 1.367$$

and at quarter corners as

$$\frac{E'(\lambda)}{E(\lambda)} = \frac{0.539}{0.444} = 1.213$$

The surveyors avoided the cardinal directions at both corners but to a higher degree at the township corners resulting in a higher correction factor.

After aggregating species represented by fewer than 25 bearing trees, mean areas of four individual species were analyzed at township corners (Table 5) and of 13 species at quarter corners (Table 6). To assess the degree to which differences in species' mean areas represented species bias by surveyors as opposed to occurrence in different-density forests, the mean areas of bearing trees of different species were compared with their mean distance ranks (Fig. 8). These correlated reasonably well (r^2 = 0.54 and 0.27 for township and quarter corners, respectively), indicating that at least some of the differences in mean areas were the result of species bias and not variations in tree density. For township corners, the species bias correction factor was calculated from eq. 8 as

$$\frac{E'(\lambda)}{E(\lambda)} = 244.9\left(\frac{0.409}{230.7} + \frac{0.196}{328.8} + \frac{0.100}{231.6} + \frac{0.057}{171.1} + \frac{0.237}{223.3}\right) \approx 1.028$$

and for quarter corners as 1.034 using the same method. Adjusted estimates of relative frequencies of individual species were calculated using eq. 9; for example, the frequency of American beech (*Fagus grandifolia* Ehrh.) at township corners was estimated as

$$0.409/230.7 \times \left(\frac{0.409}{230.7} + \frac{0.196}{328.8} + \frac{0.100}{231.6} + \frac{0.057}{171.1} + \frac{0.237}{223.3}\right) \approx 0.422$$

representing an adjustment of +1.3% over the relative fre-

Fig. 6. Simulated density estimates in clustered, random, and dispersed forests with surveyor bias/inconsistency: (*a*) quadrant configuration inconsistency, (*b*) bearing bias, and (*c*) species bias. Estimates shown before (circles) and after (squares) applying correction factors. The broken line is actual density (100 trees/ha); the solid line recreates no-bias estimates from Fig. 4.

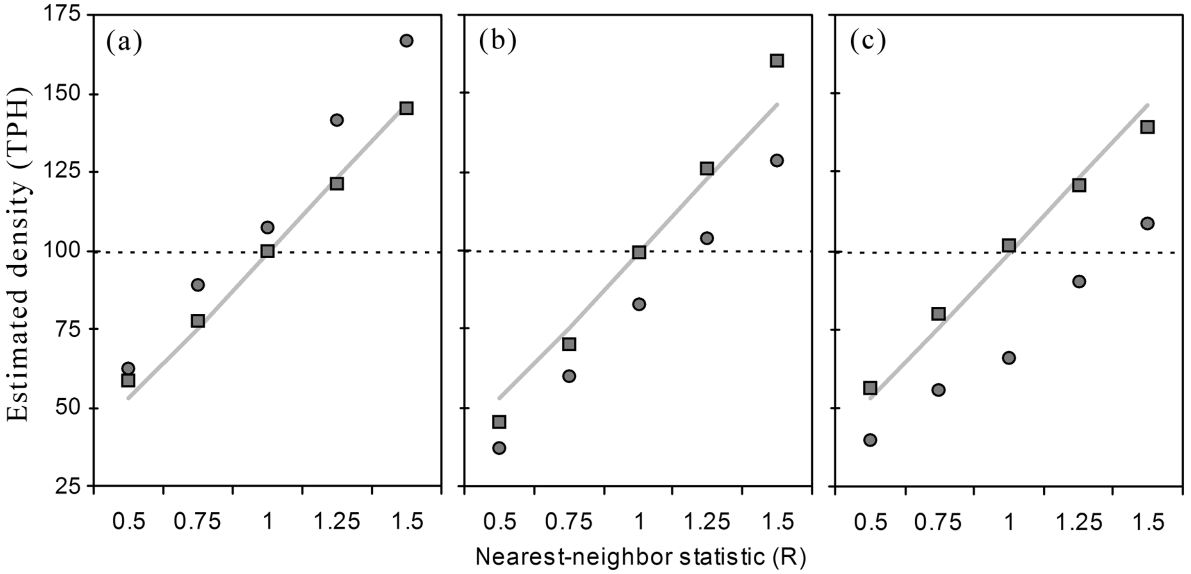
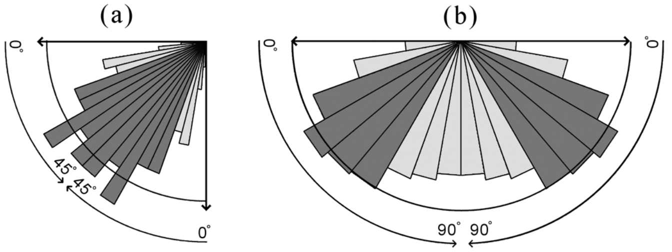


Fig. 7. Angular arrangement of bearing trees at (*a*) township corners and (*b*) quarter corners along township perimeters. The distance of bars is proportional to the number of bearing trees located in each angular section.



quency in the survey records. Of the four species (beech, sugar maple (*Acer saccharum* Marsh.), eastern hemlock (*Tsuga canadensis* (L.) Carrière), and elm (primarily American elm (*Ulmus americana* L.))) common to both analyses, there was agreement in the direction of bias for three species. Considerable biases against beech and for sugar maple were apparent in both data sets. A small bias against hemlock was also detected, primarily at the quarter corners. For elm trees, township corners showed evidence of negative bias but quarter corners showed evidence of positive bias. In the quarter corner data, the analysis also suggested that surveyors favored white oak (*Quercus alba* L.) and avoided basswood (*Tilia americana* L.) and ironwood (*Ostrya virginia* (Mill.) K. Koch).

The correction factors for the HLC data are summarized in Table 7. Different biases were clearly evident in the two data sets, with bearing angle bias stronger at the township corners and quadrant configuration inconsistency stronger at the quarter corners; species bias was approximately the same in both data sets. These results conform to our initial hypotheses that greater surveyor diligence would result in higher quadrant consistency and greater avoidance of the survey lines at township corners. Furthermore, application of the correction factors resulted in convergence of the density estimates. The final estimate range of 223.1–258.0 trees/ha was nearly 40% higher than the initial estimate obtained from the township corner data without correcting for surveyor bias and inconsistency and higher also than that obtained from the quarter corner data.

Discussion

Although PLSRs provide one of the most comprehensive sources of data on the ecology of presettlement forests, uncertainty regarding the methods used by surveyors has led to much skepticism regarding the quality of information gleaned from them. Several previous studies have addressed potential surveyor bias in bearing or witness tree selection in PLSRs (Bourdo 1956; Hushen et al. 1966; Delcourt and Delcourt 1974; Almindinger 1997; Black and Abrams 2001; Manies et al. 2001). These studies all emphasize that while bias does not render data useless, a proper understanding of the effects of bias is necessary for meaningful ecological interpretation.

We have exploited the traces of bias and inconsistency contained in the survey data to quantify their effects on estimation of presettlement forest structure and composition. Correction factors were proposed to account for errors introduced from inconsistency in bearing tree configuration, bearing angle bias, and species bias. The correction factors are applicable to all PLSRs in which distances and bearings of witness trees were recorded, including private land surveys and public GLO surveys.

Validity of the correction factors was confirmed in randomized simulation trials and further supported in a qualitative assessment using two different data sets taken from the same forested region of western New York. In randomized simulation trials, the correction factors resulted in accurate adjusted density estimates for all three types of surveyor inconsistency and bias in random and nonuniform density forests. Although they were moderately biased in very clus-

Table 2. Simulated observed and adjusted frequencies of nonpreferred species clustered and dispersed forests with species bias (actual relative frequency = 50%).

	Observed frequency (%)				Adjusted frequency (%)			
Species bias	R = 0.50	R = 0.75	R = 1.25	R = 1.50	R = 0.50	R = 0.75	R = 1.25	R = 1.50
1.5×	17.4	24.8	34.8	38.7	43.8	49.0	51.7	51.7
2×	10.9	15.5	23.2	26.5	43.5	48.4	52.2	53.2
2.5×	7.7	11.3	15.6	17.8	42.0	49.8	51.9	52.8
3×	6.0	8.3	11.2	12.4	41.2	49.8	51.2	51.9

Table 3. Bearing tree counts by angular section for HLC township corners.

Bearing range	Count	Expected count[a]	Order of removal	p[b]	Mean area (m^2)
0–5	14	66	1	0.000	71.3
6–10	24	60	2	0.000	65.8
11–15	54	60	3	0.000	80.4
16–20	46	60	4	0.002	86.6
21–25	74	60	(5)[c]	0.355	79.1
26–30	73	60			66.2
31–35	95	60			89.4
36–40	80	60			71.2
41–45	80	54			78.9

[a] First and last sections differ because bearing angles were recorded as whole integers.

[b] χ^2 statistic for all remaining ranges prior to removal of given range.

[c] Section tested for removal but not removed.

Table 4. Bearing tree counts by angular section for HLC quarter corners.

Bearing range	Count	Expected count[a]	Order of removal	p[b]	Mean area (m^2)
0–10	136	341.4	1	0.000	38.1
11–20	249	325.1	2	0.000	29.6
21–30	362	325.1	(6)[c]	0.223	29.5
31–40	420	325.1			30.5
41–50	399	325.1			32.5
51–60	396	325.1			29.2
61–70	337	325.1	5	0.023	34.7
71–80	331	325.1	4	0.003	31.0
81–90	296	308.9	3	0.000	27.8

[a] First and last ranges differ because bearing angles were recorded as whole integers.

[b] χ^2 statistic for all remaining ranges prior to removal of given range.

[c] Range tested for removal but not removed due to high p value.

tered and dispersed forests, only small biases were evident in moderately clustered and dispersed forests. Estimation variability also increased somewhat due to bearing and species bias. Calculations for the HLC survey are consistent with the hypotheses that greater surveyor diligence at township corners would have resulted in greater conformity to the prescribed quadrant arrangement rule but also avoidance of trees along survey lines. Convergence in the overall range of estimates, from a difference of 57.1 trees/ha initially to 34.9 trees/ha after application of the correction factors, supports the validity of the calculations, and a density range of 223.1–258.0 trees/ha is consistent with both data sets.

The proposed correction factors are based on a minimal set of underlying assumptions, and as with all statistical methods, careful assessment of these assumptions is necessary for proper interpretation. Two types of assumptions are relevant and should be given due consideration. First, it is assumed that the spatial pattern of the forest is uniform random. Nonuniform density, as exists in savannas and other landscapes, will generally result in density underestimation as shown in Fig. 3 and other studies (e.g., Pollard 1971). Clustered and dispersed forest patterns will result, respectively, in density underestimation and overestimation as shown in Fig. 5 and other studies (e.g., Steinke and Hennenberg 2006). Extreme clustering and dispersal will further result in moderate biases in the proposed correction factors as shown in Fig. 6. Thus, significant patterns of nonuniform density, clustering, or dispersal should be ruled out prior to any quantitative analysis, or else their effects should be thoroughly assessed.

Second, each individual correction factor is based on a model of surveyor inconsistency or bias in random forests, and the assumptions of these individual models should be carefully analyzed. In particular, the correction factor for species bias assumes that (*i*) each species is found in forests of equal density so that any differences in average distance from the survey corner are the result of surveyor bias and (*ii*) surveyor bias is consistent within the study area. The former can be validated by comparing species distances with distance ranks, as in Fig. 8; a strong correlation will support the validity of the correction factor. The latter assumption is especially important in light of the results of Manies et al. (2001) in which species bias was found to vary from one surveyor to another. The effects of this variability were not evaluated in the present study owing to the difficulty in determining the surveyor responsible for individual corners in the HLC region. However, when analyzing GLO records where such information is readily available, we recommend separate calculation of the species bias correction for each individual surveyor.

It is hoped that the correction factors proposed here will prove useful in facilitating several types of quantitative analysis of PLSR data. First, in regions where original survey records contain bearing tree diameters as well as distances and bearings, reliability of direct comparisons between the density of presettlement and modern-day forests will be enhanced. Nevertheless, proper caution should be applied, as there is uncertainty associated with the correction factors as well as uncorrected density estimates. We suggest that a range of estimates be computed by determining the probable bounds of each correction factor, likely effects of nonuniform density, clustering,and dispersal, and the variance of the base estimator (eq. 3).

Second, in private land company surveys where diameters were often not recorded, improved density estimation from

Table 5. Estimated relative species frequencies at township corners in the HLC survey region before and after bias correction.

Species	Mean area (m^2)	Initial %	Adjusted %
American beech (*Fagus grandifolia* Ehrh.)	230.7	40.9	42.2
Sugar maple (*Acer saccharum* Marsh.)	328.8	19.6	14.2
Eastern hemlock (*Tsuga canadensis* (L.) Carrière)	231.6	10.0	10.3
American elm (*Ulmus americana* L.)	171.1	5.7	8.0
Other	223.3	23.7	25.3

Table 6. Estimated relative species frequencies at section corners in the HLC survey region before and after bias correction.

Species	Mean area (m^2)	Initial %	Adjusted %
American beech (*Fagus grandifolia* Ehrh.)	77.6	36.8	42.1
Sugar maple (*Acer saccharum* Marsh.)	112.0	22.8	18.0
Eastern hemlock (*Tsuga canadensis* (L.) Carrière)	78.8	8.3	9.4
Basswood (*Tilia americana* L.)	100.3	5.8	5.1
American elm (*Ulmus americana* L.)	122.4	4.8	3.5
Black ash (*Fraxinus nigra* Marsh.)	75.9	3.4	3.9
Birch (*Betula alleghaniensis* Britt., *Betula lenta* L.)	94.7	2.7	2.5
White ash (*Fraxinus americana* L.)	104.1	2.1	1.8
Maple (*Acer rubrum* L., *Acer saccharum* Marsh.)	76.8	2.0	2.3
White oak (*Quercus alba* L.)	130.7	2.0	1.4
White pine (*Pinus strobus* L.)	84.3	1.5	1.6
American chestnut (*Castanea dentata* (Marsh.) Borkh.)	99.9	1.4	1.3
Ironwood (*Ostrya virginia* (Mill.) K. Koch)	54.2	1.1	1.9
Other	88.5	5.4	5.4

Fig. 8. Relationship between mean area and mean distance rank of bearing trees by species at (*a*) township corners and (*b*) quarter corners.

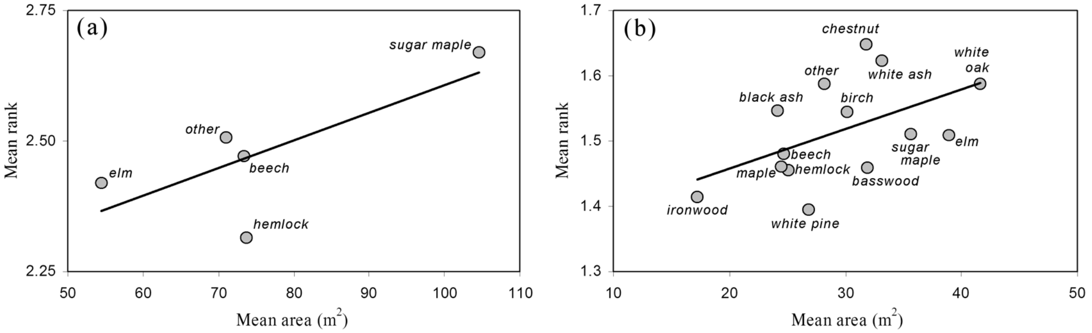


Table 7. Summary of correction factors and initial and adjusted estimates of tree density in the HLC survey region.

		Correction factor			
Corner type	Initial estimate (trees/ha)	Quadrant configuration inconsistency	Bearing bias	Species bias	Adjusted estimate (trees/ha)
Township	162.8	0.975	1.367	1.028	223.1
Quarter	219.7	0.936	1.213	1.034	258.0

survey records can enable appropriate comparison with modern surveys on equivalent bases. This is important because diameters were not recorded in all surveys, and surveys using different diameter cutoffs will result in different densities especially for small understory species. For example, we estimated a density of between 223 and 258 trees/ha (Table 7) from the HLC bearing trees. In mature modern-day forests in this region, this density corresponds to trees >9 in. diameter at breast height, which we hypothesize as the cutoff used by HLC surveyors. Studies in other regions have assumed a cutoff of 4–5 in. when comparing PLSRs with modern forest inventories (Fralish et al. 1991; Palik and Pregitzer 1992; Friedman and Reich 2005), but our results correspond better to Dyer's (2001) analysis of forest change and vegetation site relationships in Ohio, which used a 9-in. cutoff.

Third, the correction factor for species bias provides valuable information on the probable magnitude and direction of bias for particular tree species. For example, Bourdo (1956) suggested that surveyors might favor trees with thin or highly visible bark, leaving open the possibility that the dominance of beech trees in PLSR data from New York (Seischab 1992) and New England (Cogbill et al. 2002) is due at least in part to surveyor bias. However, our analysis showed a significant bias *against* beech in the HLC survey data at both the township and quarter corners.

Although the computed values for the correction factors will vary from one region to another, it will be interesting to see if general tendencies exist, as broader examination of such tendencies will enhance our general understanding and interpretation of the PLSRs. As summarized in Table 7, for the HLC survey in western New York, we found the correction required for bearing angle bias to be greatest, affecting density estimation by 36.7% in the case of township corners. At quarter corners, the correction required for quadrant configuration was appreciable. As for species bias, although it has received the most attention in the literature, our computations showed its effect to be small for the HLC data. All of these tendencies are consistent with GLO instructions and the results of Manies et al. (2001), and we suspect them to be generally representative for GLO surveys.

The methods presented here can provide important information regarding surveyor tendencies and add confidence to conclusions made regarding presettlement forest structure and composition. Even if surveyor methods can be perfectly understood, however, the effects of natural patterns of clustering and dispersal on distance-based density estimators remain large and are perhaps the biggest single impediment to quantitative analysis of PLSRs. Simulations can provide information on the magnitude of errors introduced into density estimators for various degrees of clustering/dispersal. Further work is needed, however, to develop methods to estimate spatial patterning from point-to-tree distance data.

Acknowledgements

We wish to thank Chris Larsen for his discussion in the formative stage of this research and for the helpful suggestions of two anonymous reviewers. Funding was providing by the Academic Research Fund (R-109-000-060-112 and R-109-000-060-133) from the National University of Singapore.

References

Abrams, M.D., and McCay, D.M. 1996. Vegetation–site relationships of witness trees (1780–1856) in the presettlement forests of eastern West Virginia. Can. J. For. Res. **26**: 217–224.

Almindinger, J. 1997. Minnesota's bearing tree database. Biol. Rep. 56. Minnesota Department of Natural Resources, St. Paul, Minn.

Bailey, R.G. 1995. Description of the ecoregions of the United States. 2nd ed. Misc. Publ. No. 1391. Map scale 1 : 7 500 000. USDA Forest Service, Washington, D.C.

Barbour, M.G., Burk, J.H., Pitts, W.D., Gilliam, F.S., and Schwartz, M.W. 1999. Terrestrial plant ecology. 3rd ed. Addison Wesley Longman Inc., Menlo Park, Calif.

Batcheler, C.L. 1971. Estimation of density from a sample of joint point and nearest-neighbor distances. Ecology, **52**: 703–709. doi:10.2307/1934161.

Black, B.A., and Abrams, M.D. 2001. Influences of Native Americans and surveyor biases on metes and bounds witness-tree distribution. Ecology, **82**: 2574–2586.

Bourdo, E.A. 1956. A review of the general land office survey and of its use in quantitative studies of former forest. Ecology, **37**: 754–768. doi:10.2307/1933067.

Clark, P.J., and Evans, F.C. 1954. Distance to nearest neighbor as a measure of spatial relationships in populations. Ecology, **35**: 445–453. doi:10.2307/1931034.

Cogbill, C.V., Burk, J., and Motzkin, G. 2002. The forests of presettlement New England, USA: spatial and compositional patterns based on town proprietor surveys. J. Biogeogr. **29**: 1279–1304. doi:10.1046/j.1365-2699.2002.00757.x.

Cottam, G., and Curtis, J.T. 1956. The use of distance measures in phytosociological sampling. Ecology, **37**: 451–460. doi:10.2307/1930167.

Cowell, C.M. 1995. Presettlement piedmont forests: patterns of composition and disturbance in central Georgia. Ann. Assoc. Am. Geogr. **85**: 65–83.

Delcourt, H.R., and Delcourt, P.A. 1974. Primeval magnolia–holly–beech climax in Louisiana. Ecology, **55**: 638–644. doi:10.2307/1935154.

Donnelly, K. 1978. Simulations to determine the variance and edge-effect of total nearest neighbor distance. *In* Simulation methods in archaeology. *Edited by* I. Hodder. Cambridge University Press, London, U.K. pp. 91–95.

Dyer, J.M. 2001. Using witness trees to assess forest change in southeastern Ohio. Can. J. For. Res. **31**: 1708–1718. doi:10.1139/cjfr-31-10-1708.

Fralish, J.S., Crooks, F.B., Chambers, J.L., and Harty, F.M. 1991. Comparison of presettlement, second-growth and old-growth forest on six site types in the Illinois Shawnee Hills. Am. Midl. Nat. **125**: 294–309. doi:10.2307/2426234.

Frelich, L.E. 1995. Old forest in the Lake States today and before European settlement. Nat. Areas J. **15**: 157–167.

Friedman, S.K., and Reich, P.B. 2005. Regional legacies of logging: departure from presettlement forest conditions in northern Minnesota. Ecol. Appl. **15**: 726–744. doi:10.1890/04-0748.

Grimm, G.C. 1984. Fire and other factor controlling the big woods vegetation of Minnesota in the mid-nineteenth century. Ecol. Monogr. **54**: 291–311. doi:10.2307/1942499.

Hushen, T.W., Kapp, R.E., Bogue, R.D., and Worthington, J.T. 1966. Presettlement forest patterns in Montcalm County, Michigan. Mich. Bot. **5**: 192–211.

Jackson, S.M., Pinto, F., Malcolm, J.R., and Wilson, E.R. 2000. A comparison of pre-European settlement (1857) and current (1981–1995) forest composition in central Ontario. Can. J. For. Res. **30**: 605–612. doi:10.1139/cjfr-30-4-605.

Janke, R.A., McKaig, D., and Raymond, R. 1978. Comparison of presettlement and modern upland boreal forest on Isle Royale National Park. For. Sci. **24**: 115–121.

Jost, L. 1993. A simple distance estimator for plant density in nonuniform stands. Appendix to Texabama Croton Investigations, DLS Associates, Austin, Tex (U.S. Army project No. 93-396). Available from www.loujost.com/Statistics%20and%20Physics/PCQ/PCQJournalArticle.htm [accessed 17 August 2006].

Manies, K.L., and Mladenoff, D.J. 2000. Testing methods to produce landscape-scale presettlement vegetation maps from the U.S. public land survey records. Landscape Ecol. **15**: 741–754. doi:10.1023/A:1008115200471.

Manies, K.L., Mladenoff, D.J., and Nordheim, E.V. 2001. Assessing large-scale surveyor variability in the historic forest data of the original U.S. Public Land Survey. Can. J. For. Res. **31**: 1719–1730. doi:10.1139/cjfr-31-10-1719.

Mladenoff, D.J., Dahir, S.E., Nordheim, E.V., Schulte, L.A., and

Guntenspergen, G.G. 2002. Narrowing historical uncertainty: probabilistic classification of ambiguously identified tree species in historical forest survey data. Ecosystems (N.Y., Print), **5**: 539–553.

Nelson, J.C. 1997. Presettlement vegetation patterns along the 5th principal meridian, Missouri territory, 1815. Am. Midl. Nat. **137**: 79–94. doi:10.2307/2426757.

Palik, B.J., and Pregitzer, K.S. 1992. A comparison of presettlement and present-day forests on two Bigtooth aspen-dominated landscapes in Northern Lower Michigan. Am. Midl. Nat. **127**: 327–338. doi:10.2307/2426539.

Pielou, E.C. 1959. The use of point to plant distances in the study of the pattern of plant poulations. J. Ecol. **47**: 607–613.

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

Radeloff, V.C., Mladenoff, D.J., He, H.S., and Boyce, M.S. 1999. Forest landscape change in the northwestern Wisconsin Pine Barrens from pre-European settlement to the present. Can. J. For. Res. **29**: 1649–1659. doi:10.1139/cjfr-29-11-1649.

Seischab, F.K. 1992. Forest of the Holland Land Company in western New York, circa 1798. N.Y. State Mus. Bull. **484**: 36–53.

Siccama, T.G. 1971. Presettlement and present forest vegetation in northern Vermont with special reference to Chittenden County. Am. Midl. Nat. **85**: 153–172. doi:10.2307/2423919.

Steinke, I., and Hennenberg, K.J. 2006. On the power of plotless density estimators for statistical comparisons of plant populations. Can. J. Bot. **84**: 421–432. doi:10.1139/B05-135.

Wang, Y.-C. 2004. A geographic information approach to analyzing and visualizing presettlement land survey records for vegetation reconstruction. Ph.D. dissertation, University at Buffalo, The State University of New York, Buffalo, N.Y.

Wang, Y.-C. 2005. Presettlement land survey records of vegetation: geographic characteristics, quality and modes of analysis. Prog. Phys. Geogr. **29**: 568–598.

White, C.A. 1984. A history of the rectangular survey system. Bureau of Land Management, Government Printing Office, Washington, D.C.

White, M.A., and Mladenoff, D.J. 1994. Old-growth forest landscape transitions from pre-European settlement to present. Landscape Ecol. **9**: 191–205. doi:10.1007/BF00134747.

Whitney, G.G., and DeCant, J. 2001. Government land office survey and other early land surveys. *In* The historical ecology handbook: a restorationist's guide to reference ecosystems. *Edited by* D. Egan and E.A. Howell. Island Press, Covelo, Calif. pp. 147–172.

Wyckoff, W. 1988. The developer's frontier: the making of the western New York landscape. Yale University Press, New Haven, Conn.

Appendix A. Derivation of correction factors

Quadrant configuration inconsistency correction factor

The following variables are defined as in the main text: c is the number of corners, b is the number of bearing trees recorded at each corner, n is the total number of bearing trees ($= c \times b$), and μ is the mean area of trees at all corners. Equation 2 is used as the base estimator $E(\lambda)$ throughout. Further denote p as the proportion of corners with one tree in each quadrant, μ_q as the mean area of trees at corners with one tree in each quadrant, and ϕ as the probability of four nearest trees occurring in four quadrants due to chance.

If surveyors select the b nearest trees to each corner, the appropriate correction factor will be the ratio of eq. 4 to eq. 2:

$$c(b+1)/2(bc-1)$$

This ratio should only be applied, however, to the subset of corners at which point-quarter sampling was not followed. The observed proportion p does not include "false conformities" at which the surveyors simply picked the nearest b trees, but these happened to fall into the prescribed configuration. The best estimate of the proportion of false conformities is

$$\frac{\phi}{1-\phi}(1-p)$$

Therefore, the estimated proportion of corners at which surveyors conformed to the rule is

$$p-\frac{\phi}{1-\phi}(1-p)=\frac{p-\phi}{1-\phi}$$

and the proportion of corners at which they did not is

$$1-p+\frac{\phi}{1-\phi}(1-p)=\frac{1-p}{1-\phi}$$

It is necessary to adjust the mean area of the corners at which the rule was followed as well, since this may differ substantially from that of the false conformities. The best estimate of the latter is simply the mean area of the nonconforming corners:

$$\frac{\mu-p\mu_q}{1-p}$$

When this is factored out, the mean area at corners where point-quarter sampling was conducted is

$$\frac{p\mu_q-\phi\mu}{p-\phi}$$

Since estimated density is inversely proportional to mean area, the correction factor is the ratio between the weighted average of the inverses of the estimated mean areas at the two sets of corners and the inverse of the overall mean area:

$$\frac{E(\lambda')}{E(\lambda)}=\frac{\left(\frac{p-\phi}{1-\phi}\right)\left(\frac{p-\phi}{p\mu_q-\phi\mu}\right)+\left(\frac{c(b+1)}{2(cb-1)}\right)\left(\frac{1-p}{1-\phi}\right)\left(\frac{1-p}{\mu-p\mu_q}\right)}{1/\mu}=\frac{(p-\phi)^2}{(p\mu_q/\mu-\phi)(1-\phi)}+\left(\frac{c(b+1)}{2(cb-1)}\right)\frac{(1-p)^2}{(1-p\mu_q/\mu)(1-\phi)}$$

which is the same as eq. 6.

Bearing angle bias correction factor

Suppose that trees from only a limited proportion α ($0 < \alpha \leq 1$) of angles out of the 360° circle are considered as bearing tree candidates. A density estimate from such a set of bearing trees will then represent only the proportion α of the forest and must be multiplied by $1/\alpha$ to obtain the true density. Alternatively, consider the fact that the selected bearing trees in this case will often not be the nearest to the post. Some bearing trees will be the second nearest, a few will be the third nearest, even fewer still the fourth nearest, and so on. The expected distance rank for a given tree can be expressed as the infinite series

$$E(k) = \beta + (1-\beta)\beta + (1-\beta)^2\beta + (1-\beta)^3\beta + ... + (1-\beta)^\infty\beta = 1/\beta$$

giving the same result. This line of reasoning can be extended to situations where surveyors avoided certain angular sections some but not all of the time. Let p_i denote the proportion of bearing trees recorded in each angular section α_i. The angular section with the maximum density of bearing trees can be considered to represent an unbiased sample of a proportion α_i of the forest. The density of bearing trees in this angular section will be greater than the density of all bearing trees by a factor of p_i/α_i, which is the correction factor expressed in eq. 7.

Species bias correction factor

The following variables are carried over from the Species bias section: s is the number of distinct species, p_j is the observed survey frequency of species j:

$$\sum_{j=1}^{s} p_j = 1$$

ϕ_j is the actual frequency of species j:

$$\sum_{j=1}^{s} \phi_j = 1$$

and μ_j is the observed mean area of bearing trees of species j. For each species j, further define ξ_j as a bias parameter such that the "perceived" distance r^* of a tree of species j is given by

$$r^* = \sqrt{\xi_j} r$$

A low value of ξ_j indicates that trees of species j will be preferentially selected, and a high value indicates that they will be preferentially avoided. Under this model, bearing tree selection with species bias will be equivalent to unbiased selection in a hypothetical forest in which the density of each species is a fraction ξ_j of its density in the real forest. Put differently, in the biased sample of surveyor-selected trees, the mean area of each species will be overstated to the same degree that its frequency is understated (or vice versa). This is expressed by the following equality:

$$\frac{\phi_j}{p_j} = \frac{\mu'}{\mu_j}$$

where μ' is the expected mean area absent of surveyor bias. Taking sample size of each species into consideration, we solve for the true species proportions ϕ and overall mean area μ' by minimizing

$$\sum_{j=1}^{s} \sqrt{np_j}\left(\frac{\phi_j}{p_j} - \frac{\mu'}{\mu_j}\right)^2$$

subject to the following constraints:

$$\sum_{j=1}^{s} \phi_j = 1$$

$$\phi_j > 0$$

$$\mu > 0$$

Using a LaGrangian multiplier to enforce the first constraint, the following solution emerges:

$$\phi_j = p_j\mu'/\mu_j$$

$$\mu' = \frac{1}{\sum_{j=1}^{s}\left(p_j/\mu_j\right)}$$

It can be seen that no change of sign will occur, so the second and third constraints will hold automatically. By substitution, the actual frequencies of each species can be estimated from the observed frequencies and mean areas:

$$\phi_j = p_j/\mu_j \sum_{j=1}^{s}\left(p_j/\mu_j\right)$$

Since density is an inverse function of mean area, the appropriate correction factor can be derived as

$$\frac{E'(\lambda)}{E(\lambda)} = \frac{\mu}{\mu'} = \mu\sum_{j=1}^{s}\left(p_j/\mu_j\right)$$

which is the same as eq. 8.