Comparison of Ground Sampling Methods for Estimating Canopy Cover

Michael S. Williams, Paul L. Patterson, and H. Todd Mowrer

ABSTRACT. Knowledge of the canopy structure is essential to improving our understanding of forest structure. While numerous sampling techniques have been developed to estimate attributes of the forest canopy, these require either additional measurements or a sampling design and measurement techniques that differ substantially from the ones that are used to estimate more traditional forest attributes, such as basal area, number of stems, or volume. The root of the problem is that the sample element for a design that estimates canopy attributes is the tree crown, whereas the sample element is the bole for a design that estimates an attribute such as basal area. For example, if a fixed-area plot is used to estimate basal area, canopy cover cannot be estimated using the same design because a portion of the plot invariably is covered by the crowns of trees whose boles lie outside the plot boundary and would not be included in the sample under the standard sampling design. In this study, a technique called "morphing" is used to model the trees outside the plot boundary. For the purpose of comparison, the morphing technique is used to estimate canopy cover using data from a circular fixed-area plot, and this technique is compared with both dot count and line intersect sampling using a simulation study and two small forest populations. For the study, the populations were sampled using circular fixed-area plots with radii ranging from $\rho = 3.05$–6.10 m (10–20 ft) and line lengths ranging from $L = 3.05$–22.9 m (10–75 ft). For both populations, the bias of the canopy cover estimator derived from the morphing technique was negligible. The estimator based on line intersect sampling is design-unbiased, but it generally had a much larger variance than the one based on the morphing technique. The dot count method consistently had the highest variance. For. Sci. 49(2):235–246.

Key Words: Canopy structure, morphing, torus edge-correction.

The primary objective of most forest inventories has been to assess standing timber resources. Thus, the sample elements are usually the tree bole with the criterion for inclusion in the sample being distance from the sample point and a minimum diameter measured at 1.37 m above the ground. Because the probability of inclusion of a tree bole can be determined, design-unbiased and efficient estimators exist for making inferences about attributes associated directly with the bole. Examples are diameter, basal area, growth increment, height, age, percent defect, and volume. Another important class of forest attributes is the one related to the forest canopy. Forest canopy structure is of interest because it controls a host of different processes including heat and mass transfer; temperature and moisture of soils (Campbell and Norman 1998); quantity and quality of light reaching the forest floor (Jennings et al. 1999); animal habitat, understory regeneration (Lowman and Nadkarni 1995); and the

Michael S. Williams, Mathematical Statistician, Rocky Mountain Research Station, USDA Forest Service, 2150 A Center Drive, Suite 361, Fort Collins, Colorado 80526—Phone: (970) 295-5974; Fax: (970) 295-5959; E-mail: usdafs@lamar.colostate.edu. Paul L. Patterson, Mathematical Statistician, Rocky Mountain Research Station, USDA Forest Service, 2150 A Center Drive, Suite 350, Fort Collins, Colorado 80526. H. Todd Mowrer, Project Leader, Rocky Mountain Research Station, USDA Forest Service, 240 W. Prospect Road, Fort Collins, Colorado 80526-2098.

Acknowledgments: The authors thank the Associate Editor, and three anonymous reviewers, for their constructive comments.

Manuscript received June 22, 2001, accepted May 16, 2002.

This article was written by U.S. Government employees and is therefore in the public domain.

reflected light which is the source of remotely sensed data. An example of a canopy-related attribute is canopy cover, which will be the focus of this study and is defined as the proportion of the forest floor covered by the vertical projection of the tree crowns (Jennings et al. 1999).

For a large-scale survey that uses circular fixed-area plot sampling, such as the Forest Inventory and Analysis (FIA) program of the United States (see Frayer and Furnival 1999), two problems exist when estimating canopy cover using standard forest inventory plots. The first is that when estimating canopy cover, the elements of the population are the tree crowns rather than the tree boles. Thus, an unbiased estimator of the proportion of a fixed-area plot covered by tree crowns cannot be obtained without including those trees whose boles fall outside the plot boundary, but whose crowns cover a portion of the plot. The bias associated with ignoring these crowns was studied by Nelson et al. (1998). The other problem is that canopy cover is the sum of the individual crown areas minus the intersection of their overlapping crown areas, rather than the sum of the crown areas. Thus, a sample design whose basic sample elements are tree boles is not appropriate for the estimation of canopy cover. Standard sampling designs, such as dot counts and line-intersect sampling, are viable alternatives. However, these alternatives can add considerable time and expense to a survey. Another alternative is to use a model to predict canopy cover from other variables measured on the plot (Crookston and Stage 1999, Gill et al. 2000). The primary concern with this approach is that estimators based on model predictions can have a very high mean square error in situations where the models do not adequately describe the forest condition. In this study we review some canopy cover sampling techniques and test a technique, called the morphing technique. This procedure is used to model the forest condition surrounding the plot. This information is then combined with the actual plot measurements to model the missing canopy information along the plot boundary. While inference for this sampling strategy relies on a model to describe the forest condition beyond the plot boundary, the reliance on models is not to the same degree as the regression models proposed by Gill et al. (2000) and the similar analytical result used by Crookston and Stage (1999). Another advantage of this procedure is that it can be used in the estimation of most canopy-related attributes using traditional circular fixed-area plot sampling without collecting any information beyond the plot boundary. Thus, it could be used in conjunction with techniques, such as those described in Song et al. (1997), to model canopy structure from traditional survey data. The properties of this new technique are illustrated and compared to dot count and line intersect sampling in a simulation study where percent canopy cover is estimated using two small forest populations. Throughout this study, some additional observations are made regarding the dimension of the plot in relation to the dimension of the elements that make up the population of interest.

Review of Estimation Techniques

When estimating canopy cover on a forested tract, denoted by A, the elements of the population are the set of connected crown masses within the boundary of A. The attribute to be estimated is two dimensional (area). A number of different techniques exist for estimating canopy cover. The simplest is based on a count of points or dots. Jennings et al. (1999) provide an overview of the dot-count technique and summarize the results of numerous studies. For this technique, the proportion of the plot covered by the canopy is estimated by visiting n randomly located points within the plot. At each point, a vertical measurement is taken to determine whether or not a vertical line from the point would intersect a tree crown. The indicator variable δ is given the value of 1 if the point is covered by a tree's crown and a value 0 otherwise. The dot-count estimator of canopy cover is

$$\hat{Z}_{DOT} = \sum_{j=1}^{n} \delta_j / n.$$

The mean and variance of $\hat{Z}_{DOT}$ can be calculated by assuming that δ is distributed as a binomial random variable. Each measurement is taken at a point so the dimension of each plot is zero. Recommendations for the number of sample points vary, with suggestions for as few as $n = 20$ (e.g., Ganey and Block 1994) existing in the literature. The reality is that n should probably be much larger. As pointed out in Jennings et al. (1999), "With 20 observations per plot, an estimated canopy cover of 50 percent has 95 percent confidence intervals that range from less than 30 percent to in excess of 70 percent. Any estimate made with fewer than 100 observations will be of very little utility in distinguishing between forest plots with all but the grossest differences in canopy cover." While this technique is straightforward to implement, and individual measurements are quick and easy, the overall cost can still be quite high if estimates of reasonable precision are required. For this reason, the dot-count technique will play a diminished role in this study. Instead, the properties of the binomial distribution will be used to make comparisons between the dot-count technique and the two methods described below.

The standard alternative to the dot-count technique is line-intersect sampling (LIS). This technique relies on the random placement of lines within the forested area A. In practice, a linear tape is randomly located, and the edges of the connected crown masses are used to determine the length of the tape covered by the canopy. Each measurement is the length of a line so each plot is one-dimensional. Results for LIS using design-based inference have been presented by numerous authors, with Valentine et al. (2001) discussing LIS as an application of rectangular fixed-area plot sampling where the width of the plot goes to zero. Gregoire (1998) provides a broad overview and dispels the popular notion that the orientation of the lines must be random. Kaiser (1983) derives the estimators for

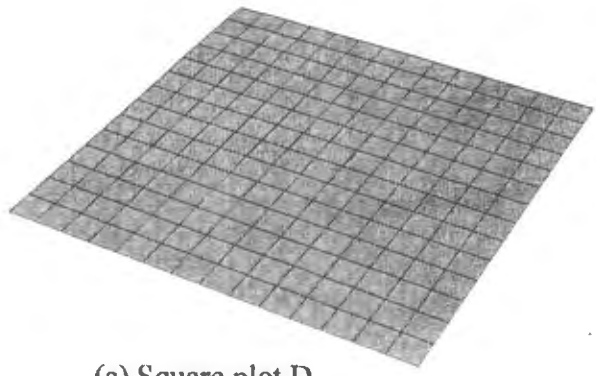
LIS taking the traditional design-based approach to inference, where n lines of fixed total length L and either fixed or random orientation are located by some point (usually the midpoint) on the line that is determined in accordance with a uniform distribution over A. An advantage of LIS is that the length of the line can be very long, which tends to decrease the variance between the n lines and produce precise estimates in comparison with the dot count technique. The LIS estimator for the proportion of A covered by the canopy with n lines with fixed orientation is given by

$$\hat{Z}_{LIS} = \frac{1}{nL} \sum_{i=1}^{n} \sum_{j=1}^{m_i} l_{ij}$$

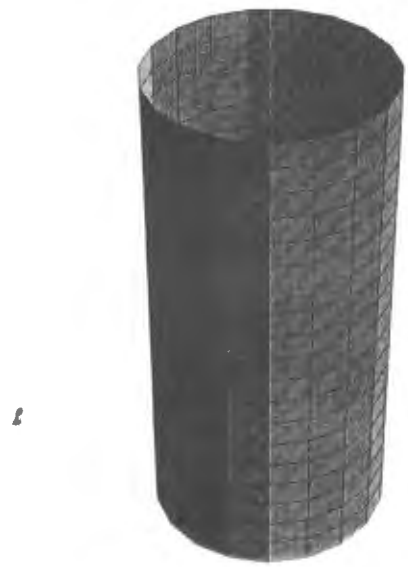
where l_{ij} is the length of the intersection of the ith line with the jth connected crown mass and m_i is the number of connected crown masses intersected by the ith line. For our purposes, the orientation of the line will be fixed.

O'Brien (1989) performed a field test of LIS and used the method of running means (Kershaw and Looney 1985) to determine that at least $n = 8$ lines, each of length 30.5 m (100 ft), were needed to produce an overall estimate that varied by less than 10%. This study also compared the LIS estimates with estimates derived from stereo-plotted aerial photography and found that while LIS was quite time-consuming, the accuracy of measurements derived from aerial photography was poor, with the aerial photo measurements falling outside of a 95% confidence internal based on $n = 10$ LIS samples about one half of the time. Due to the additional issues associated with estimating canopy cover from aerial photography (e.g., resolution, bias, shadows, cost, and age of photography), the remainder of the article will only address ground sampling methods.

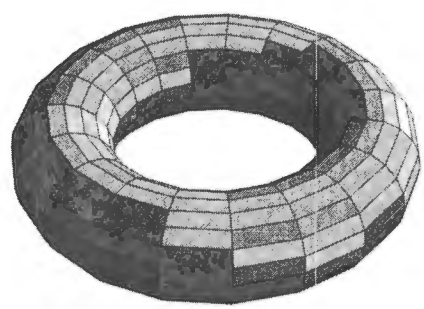
Just as in fixed- and variable-radius plot sampling, an edge-effect bias can occur when sampling along the boundary of A. Various adjustment techniques have been derived to avoid this design-based bias in $\hat{Z}_{LIS}$. For example, Gregoire and Monkevich (1994) adapted the mirage method (Gregoire 1982), which is one practical technique that can be implemented in the field. Kaiser (1983) shows that $\hat{Z}_{LIS}$ is design-unbiased when the portion of transects that falls outside of the population is mapped back into A at a random displacement. A similar alternative that is probably more suited to simulation studies on rectangular populations allows a line that falls across the boundary of A to be mapped back into A on the opposite side. This method is similar to the torus mapping technique as described in Fraser and Van Den Driessche (1972). This technique maps a rectangular area onto a connected three-dimensional surface which yields a continuous space with no boundaries. This technique is implemented as follows: For the rectangular area D, connect the top and bottom boundaries to form a cylinder. Then connect the ends of the cylinder to form a torus (Figure 1). An equivalent interpretation is to replicate D eight times to form a grid of nine identical areas. The analysis is then carried out on the central area with the replicated areas providing the needed boundary information (Figure 2). From a spatial modeling perspective, torus mapping makes the assumption that the point pattern observed on the plot also exists beyond the plot boundary, which is not always the case in a highly heterogeneous forest. However, Ripley (1979) found torus



(a) Square plot D



(b) Connect top and bottom of (a)



(c) Connect open ends of (b) to form the torus

Figure 1. Torus mapping for a square plot (a) is performed by first connecting the top and bottom edges of the square to form a cylinder (b), then ends of the cylinder are connected to form a torus (c).

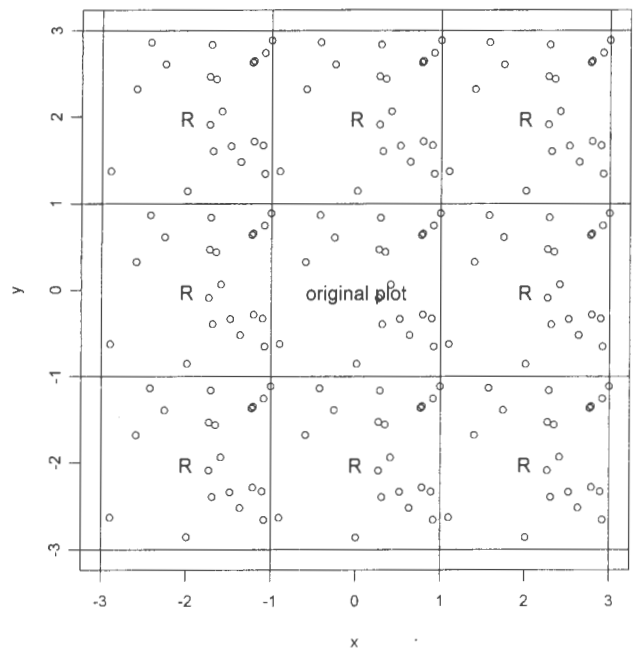


Figure 2. Alternative interpretation of the torus edge correction procedure. The data represent the location and crown diameter, which are replicated eight times, and a grid of plots is formed. Analyses are carried out on the center plot.

mapping to be an effective method of edge-correction in the field of spatial modeling. Its utility is limited in forestry applications to applications where rectangular plots are used. No equivalent method exists for circular plots, which are common in forest survey. Williams et al. (2001b) found the simplicity of the torus mapping appealing and derived a technique, called morphing, for transforming the data on a circular fixed-area plot into a square plot of equal area. The logical solution to estimating canopy related attributes is to morph the data from a circular fixed-area plot to a square one and then use torus mapping to model the crowns of those trees whose boles fell outside the circular fixed-area plot boundary. The measurement taken on each plot is the area of the plot covered by the canopy, so the dimension of the plot and the dimension of the attribute are both two-dimensional.

To present the morphing technique, both circular and square fixed-area plots have to be defined as follows: Define two spaces, the first being all points contained within a circle C with the origin of the polar coordinate system as its center and a radius of ρ. The other space, D, is all points contained by a square of equal area, with its center also at the origin. A formal definition of the spaces is

$$C = \{(x,y): x^2 + y^2 \le \rho^2\}$$

and

$$D = \left\{(x,y): |x| \le \frac{\rho\sqrt{\pi}}{2}, |y| \le \frac{\rho\sqrt{\pi}}{2}\right\}$$

The morphing transformation maps any point $(x,y) \in C$ into a point $(x', y') \in D$. Two premises were used to define the morphing transformation. First, if the point $(x,y) \in C$ lies on a circle with the origin as its center and a radius $r \le \rho$, then its transformed point $(x', y') \in D$ lies on a square with the origin as its center and with an area (πr^2) equal to that of the circle. This ensures that the number of units/area is preserved. Second, the ratio of the distance along this circle from $(r,0)$ to (x,y) divided by the circumference $2\pi r$ is equal to the ratio of the distance along the perimeter of the square from $(r\sqrt{\pi}/2, 0)$ to (x', y') divided by the perimeter of the square $4r\sqrt{\pi}$. This ensures that membership in any quadrant is preserved.

The equations for the transformation are derived from the equations that define the two premises and are

$$x' = \left[I_{(0,\frac{3\pi}{4}]}(\theta) + I_{(\frac{7\pi}{4},2\pi]}(\theta) - I_{\frac{3\pi}{4},\frac{7\pi}{4}}(\theta) \right] \left(\frac{s}{2}\right) - I_{(\frac{\pi}{4},\frac{3\pi}{4}]}(\theta) \left[\frac{2s(\theta - \frac{\pi}{4})}{\pi}\right] + I_{(\frac{5\pi}{4},\frac{7\pi}{4}]}(\theta) \left[\frac{2s(\theta - \frac{5\pi}{4})}{\pi}\right] \quad (1)$$

$$y' = \left[I_{(\frac{\pi}{4},\frac{5\pi}{4}]}(\theta) - I_{(\frac{5\pi}{4},2\pi]}(\theta) \right] \left(\frac{s}{2}\right) + I_{(0,\frac{\pi}{4}]}(\theta) \left[\frac{2s\theta}{\pi}\right] - I_{(\frac{3\pi}{4},\frac{5\pi}{4}]}(\theta) \left[\frac{2s(\theta - \frac{3\pi}{4})}{\pi}\right] + I_{(\frac{7\pi}{4},2\pi]}(\theta) \left[\frac{2s(\theta - \frac{7\pi}{4})}{\pi}\right] \quad (2)$$

where $I_{(a,b]}(\theta)$ is 1 for $a < \theta \le b$, 0 otherwise and the following substitutions are defined:

$$s = \sqrt{\pi(x^2 + y^2)} \quad (3)$$

$$\theta = \left[\tan^{-1}\frac{y}{x} + \pi I_{(-\infty,0)}(x)\right]\left[1 - I_{(0,\infty)}(x) I_{(-\infty,0]}(y)\right] + \left[\tan^{-1}\frac{y}{x} + 2\pi\right] I_{(0,\infty)}(x) I_{(-\infty,0]}(y) \quad (4)$$

where θ is in radians. These equations appear to be quite complicated because they are fragmented by many indicator variables. This occurs because while a circle can be easily defined by a single equation ($x = r\cos\theta$, $y = r\sin\theta$) a square requires one equation for each side. Figure 3 provides a graphical representation of the morphing technique where a circle of radius $\rho = 1$ contains 200 points. The arrows on the figure represent the direction and distance each point is moved in the transformation from a circular to square plot.

In most cases it is desirable to transform the morphed and replicated data back to a circular plot. This will be referred to

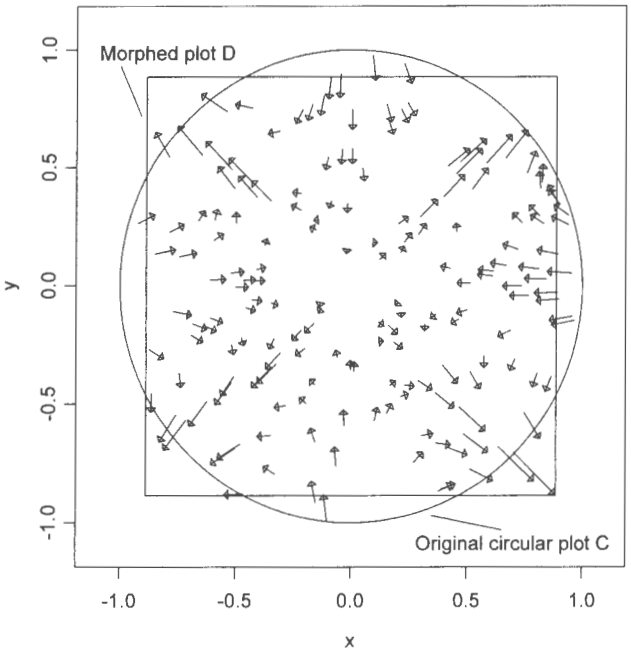


Figure 3. Movement of points for the morphing technique. Arrows represent the distance and direction of the change in location for the points within the square plot.

as demorphing. The inverse of the morphing technique to transform a point (x', y') on a square plot back into a circular plot of radius ρ is given by:

$$r = \frac{2x'}{\sqrt{\pi}}\left[I_{(0,\frac{\pi}{4}]}(\theta') + I_{(\frac{3\pi}{4},\frac{5\pi}{4}]}(\theta') + I_{(\frac{7\pi}{4},2\pi]}(\theta') \right] + \frac{2y'}{\sqrt{\pi}}\left[I_{(\frac{\pi}{4},\frac{3\pi}{4}]}(\theta') + I_{(\frac{5\pi}{4},\frac{7\pi}{4}]}(\theta') \right] \quad (5)$$

$$\theta = \frac{\pi y'}{4x'}\left[I_{(0,\frac{\pi}{4}]}(\theta') + I_{(\frac{3\pi}{4},\frac{5\pi}{4}]}(\theta') + I_{(\frac{7\pi}{4},2\pi]}(\theta') \right] + \frac{\pi x'}{4y'}\left[I_{(\frac{\pi}{4},\frac{3\pi}{4}]}(\theta') + I_{(\frac{5\pi}{4},\frac{7\pi}{4}]}(\theta') \right] \quad (6)$$

$$x = r\cos\theta \quad (7)$$

and

$$y = r\sin\theta$$

where the angle θ' is the angular displacement of the point (x', y') in polar coordinates.

The following algorithm is proposed for estimating canopy cover on a circular plot of radius ρ:

1. Morph the (x, y) coordinates of the tree locations on circular plot and their crown diameters into a square plot with coordinates (x', y') using Equations (1) and (2).

2. Replicate the morphed plot eight times and tile the information in accordance with the torus edge-correction method.

3. Demorph the replicated plot data back to a circular plot of radius 2ρ.

4. Estimate the canopy cover on the original plot of radius ρ by calculating the proportion of the plot covered by tree canopies. This estimate uses the original crowns found on the plot in conjunction with the crowns provided by the replicated data, which model the portion of the forest canopy that was not sampled by the original circular fixed-area plot. This is referred to as the morphed-torus estimator $\hat{Z}_{MT}$.

Williams et al. (2001b) show that if points $(x_1, y_1),\ldots(x_m, y_m)$ are an independently and identically distributed (iid) sample from an Uniform distribution over C, then the morphed points $(x_1', y_1'),\ldots(x_m', y_m')$ are an iid sample from a Uniform distribution over D. The Uniform distribution of points is equivalent to saying that the m points are a realization of a stationary Poisson process (Stoyan et al. 1995, p. 102) with intensity $\lambda = m/|A|$. This model has been used to describe tree counts and locations in numerous publications (e.g., Lappi 1991, Mandallaz and Ye 1999, Williams et al. 2001a). Further results on transforming the points of a Poisson process are discussed by Resnick (1992, p. 308–321). Some of the shortcomings of this modeling process are that the results only hold for stationary Poisson process models, and the size of individual tree crowns is not accounted for in the morphing transformation. Thus, it is possible for some large trees to be placed closer together than would normally occur. Marked-point process models are a class of models that could account for the size of the crowns in relation to the locations of the trees (see Penttinen et al. 1992). Unfortunately, fitting these models requires large data sets as well as adjustments to account for plot and population boundaries (Stoyan et al. 1995, p. 133–136). Williams et al. 2001 studied the performance of the morphing transformation for estimating canopy cover when the distribution of points was either more regularly spaced or more clustered than points from a Uniform distribution. They found that the differences between actual and estimated canopy cover never exceeded 1.3%, even when the spatial distribution of tree locations was more clustered than would be found in any real forest population. Thus, it seems reasonable to assume that estimates derived from this technique will be sufficiently accurate for most purposes.

The morphing technique has advantageous properties related to point process models. Thus, it could be used as a tool for studying the spatial properties of the trees on a circular fixed-area plot using standard spatial analysis techniques (e.g., Cressie 1993). However, it is felt that many potential users are more interested in estimating the true canopy cover, denoted by Z_{CC}, for a specific area using a sample of n fixed-area plots. The study area may be a forest stand or it could be the land area represented by a group of pixels on a satellite image. Thus, properties of the morphing estimator $\hat{Z}_{MT}$ must be addressed in the context of survey sampling. In deference to Gregoire (1998), a terse description of the properties of $\hat{Z}_{MT}$ is follows.

The area of interest is a two-dimensional plane, A, of area $|A|$ and the target parameter is the proportion of A covered by the forest canopy. Inference follows the design-based framework, which views the population, is fixed. The only source of randomness is the selection of sampling locations, which are determined by generating the (x,y) coordinates of n independent sampling locations from a Uniform distribution over A. Thus, the probability density function of the sample locations is $f(x,y) = 1/|A|$. At each location, a fixed-area plot of radius ρ is installed, and tree locations and crown diameters are measured for all trees on the plot. These data are morphed, replicated, demorphed, and then used to determine the morphing based measurement of crown cover at the point (x,y), denoted by $\hat{Z}_{MT}(x,y)$. The infinite set of $\hat{Z}_{MT}(x,y)$ values over A form the reference set, with the reference distribution being the infinite set of all equally likely samples of size n.

In the design-based framework, the expected value of $\hat{Z}_{MT}$ is given by

$$\mu_{MT} = E[\hat{Z}_{MT}] = \iint_A f(x,y)\hat{Z}_{MT}(x,y)dx,dy$$
$$= \frac{1}{|A|}\iint_A \hat{Z}_{MT}(x,y)dx,dy$$

The estimator of canopy cover using the n plots is

$$\hat{Z}_{MT} = 1/n\sum_{i=1}^{n} \hat{Z}_{MTi}$$

where $\hat{Z}_{MTi}$ is the morphed-torus estimator for plot i.

As illustrated by the simulation study in Williams et al. (2001b), the morphing estimator of canopy cover is not an unbiased estimator of the true canopy cover, with the bias being

$$E[\hat{Z}_{MT} - Z_{CC}] = \mu_{MT} - Z_{CC}.$$

While the magnitude of the bias cannot be determined, it is assumed to be small in most cases.

The variance is given by

$$Var[\hat{Z}_{MT}] = \frac{1}{|A|}\iint_A (\hat{Z}_{MT}(x,y) - \mu_{MT})^2 dxdy$$

The sample-based variance estimator is derived from the observed variation among the n sample plots. Thus, for the n randomly located plots, the design-unbiased estimator of the variance is

$$v\left[\hat{Z}_{MT}\right] = \frac{1}{n(n-1)}\sum_{i=1}^{n} (\hat{Z}_{MTi} - \bar{Z}_{MT})^2$$

Data Description

Two data sets were used in this study. The first data set was used to test $\hat{Z}_{MT}$ over a range of spatial patterns (Williams et al. 2001) and was gathered from a 0.58 ha square plot located on the Fraser Experimental Forest in Colorado. The data are an uneven-aged mixture of Engelmann spruce (*Picea engelmanii* [Parry]Engelmann) and subalpine fir (*Abies lasiocarpa* Nutt.). The locations of all 1,193 trees diameter at breast height greater than 2.54 cm were mapped. For each tree, the width of the projected crown was measured along the longest axis and at right angles to it. Area of crown was then derived, assuming a circular shape, from the quadratic mean of the two measurements. The largest crown in the data set had a width of 5.6 m (18.4 ft). We refer to these data as the Fraser data set.

A simple analysis of the spatial properties of the tree locations was performed. Pielou's index of nonrandomness (Cressie 1993, p. 603–605) was 1.44, which fell outside the confidence interval (0.91, 1.09), indicating a significant degree of clustering. Ripley's L function (Cressie 1993, p. 615) suggests an nonstationary point process. Figure 4 shows a plan of the projected crowns in the data set.

While numerous stem-mapped data sets exist, none that included crown diameter information were available. This led to the creation of a second data set that was based on a stem-mapped mature hardwood stand measured in central New Jersey. Crown diameters for each tree were generated using the white fir (*Abies concolor* Lindl. and Gord.) crown radius model found in Gill et al. (2000). The largest crown diameter in the data set was 2.6 m (8.6 ft). This model probably does not represent the true crown widths of the original stand. However, eastern hardwood models (e.g., Gering and May 1995) produced a population with 100% cover, which has no utility for making comparisons because $\hat{Z}_{DOT} = \hat{Z}_{LIS} = \hat{Z}_{MT}$ for every measurement.

Figure 5 illustrates the location and relative crown size of each tree. The New Jersey data covers 0.26 ha and contains 1,250 stems. This data set has three interesting

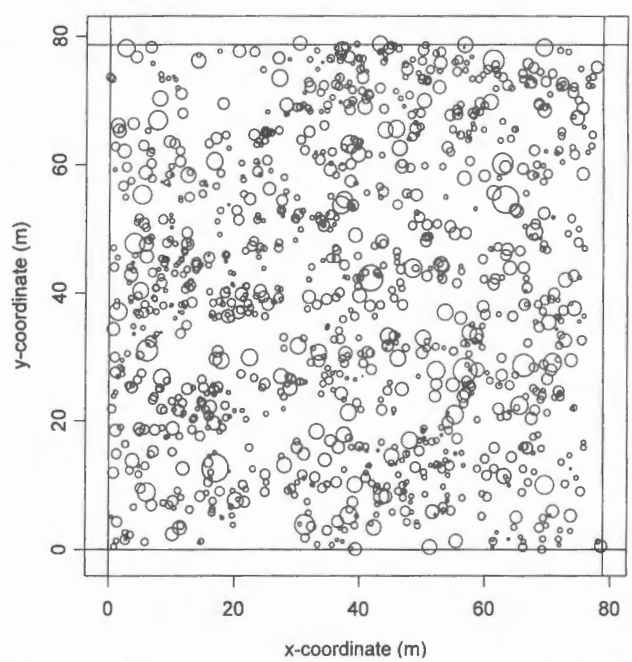


Figure 4. Plan of the projected crown diameters for the Fraser data. The diameter of each circle is proportional to the crown diameter.

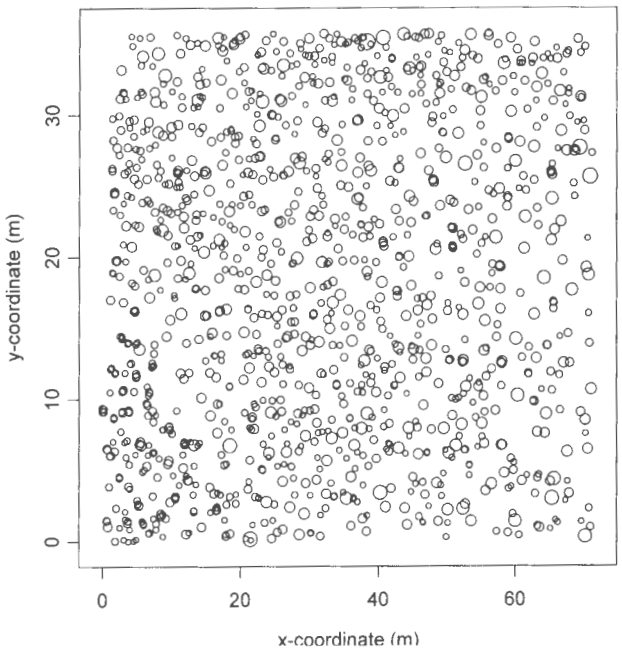


Figure 5. Plan of the projected crown diameters for the New Jersey data. The diameter of each circle proportional to the model generated crown diameter.

features. The first is the difference in forest structure between the lower left corner and rest of the stand. Thus, it is probably not reasonable to assume that any stationary point process model (Stoyan et al. 1995, p. 102) adequately describes this population. The second feature is the large number of overlapping crowns. These are probably due to forking of the main stem below breast height. The other feature of interest is that a visual inspection suggests numerous areas where the tree locations run parallel to the x-axis, though no pattern can be confirmed.

Pielou's index of nonrandomness was 1.04, which fell within the confidence interval (0.91, 1.09), indicating that there was no reason to reject the assumption of complete spatial randomness. However, it seems unlikely that any of the summary statistics of spatial pattern are meaningful due to the obvious nonstationarity of the point process.

The true canopy cover was approximately 35.6% and 48.8% for the Fraser and New Jersey data sets, respectively. These values were calculated by superimposing a fine dot grid over the population. At each location on the grid, the crown information was used to determine if the location was covered by a crown. While it is impossible to assess exactly how accurate this approximation is, further decreasing the grid spacing did not change true crown cover to four significant digits.

An important point to note is that the variance of all three canopy cover estimators is identical when the cover is either 0 or 100% (e.g., $\hat{Z}_{DOT} = \hat{Z}_{LIS} = \hat{Z}_{MT} = 0$ or 1 for every measurement) regardless of spatial arrangement, basal area, number of stems, etc. Thus, forest populations where the canopy cover ranges from approximately 30 to 70% are likely to show the greatest differences in the performance of the estimators.

Simulation Study

A Monte Carlo simulator was used to compare *LIS* and the morphing technique. The dot-count technique was not included in the simulation study because a simple closed form solution exists for the mean and variance of $\hat{Z}_{DOT}$. The simulation study was designed to draw a large number of samples, where each sample consisted of $n = 1$ plot or line randomly established within the boundary of the population.

The goal was to assess if the difference in the mean square of $\hat{Z}_{LIS}$ and $\hat{Z}_{MT}$ was sufficiently large to conclude that $\hat{Z}_{MT}$ would be superior in terms of mean square error in most field applications. The simulator established random coordinates within the boundaries of each data set at which both a line of length L and a circular plot of radius ρ were located. The L values ranged from 3.05–22.9 m (10–75 ft) in 1.52 m (5 ft) increments. The ρ values ranged from 3.05–6.1 m (10-20 ft) in 0.305 m (1 ft) increments. While plot radii in the 3.05 m range are smaller than would be used in most forest inventories, these small fixed-area plot sizes were chosen to place the morphing technique at the greatest disadvantage because they create a large boundary to interior ratio as well as a small number of trees from which to model the surrounding condition.

The locations and crown diameters of all trees within this plot were recorded. For the morphing technique, the data from the circular sample plot were morphed into a square plot, torus-mapped and demorphed back into an edge-corrected circular plot of radius 2 ρ. The canopy cover was approximated by establishing a two-dimensional grid with the width of each cell being approximately 0.15 m. Cells which were covered by a tree crown were given a value of 1, those not covered a value of 0. Canopy cover was calculated by summing the values for cells within the morphed and torus edge-corrected sample plot of radius ρ. Concurrently, at each of the random coordinates, a line with fixed orientation and length L was also established. The proportion of the line covered by tree crowns was estimated using the same two-dimensional grid. To avoid biases associated with sampling along the boundary of the population, two different techniques were used depending on the estimator. For $\hat{Z}_{MT}$, the torus edge-correction was applied to the entire population, while for $\hat{Z}_{LIS}$, any line that intersected the boundary was mapped so that it re-entered the boundary on the opposite side.

To give some indication of the variance of the canopy cover estimators in relation to the estimators of other more common forest attributes, such as basal area or volume, the total number of trees (N) was simultaneously estimated using the sample of trees derived from the circular fixed-area plot. The estimator used was

$$\hat{Z}_N = \sum_{k=1}^{K} 1 / \pi_k$$

where K is the number of trees tallied on the plot and π_k is the inclusion probability for tree k. The simulation was repeated $M = 20{,}000$ times. The mean and standard error of canopy

cover and number of trees estimates over the M samples were used to compare the estimators. In order to present and discuss all the results simultaneously, the results were expressed using the coefficient of variation calculated from the M samples. The formula used was

$$CV_* = 100 \frac{\sqrt{\left(\sum_{m=1}^{M} \hat{Z}_*^2 - \left(\sum_{m=1}^{M} \hat{Z}_*\right)^2 / M\right) / (M-1)}}{\sum_{m=1}^{M} \hat{Z}_* / M}$$

where $\hat{Z}_*$ is the estimator of either canopy cover or number of trees derived from a sample of size $n = 1$. Another useful metric is the Monte Carlo relative efficiency of the morphing technique, which is given by

$$RE = \frac{Var[\hat{Z}_*]}{Var[\hat{Z}_{MT}]}$$

The advantage of using RE to compare the methods is that the number of points required to achieve equal variance (n_{EV}) between either $\hat{Z}_{DOT}$ or $\hat{Z}_{LIS}$ or $\hat{Z}_{MT}$ using a sample of size n is $n_{EV} = nRE$.

Results

The difference between the true canopy cover and the mean of $\hat{Z}_{MT}$ over the 20,000 simulation samples was less than 0.03%. Thus, for the populations studied, the bias associated with the morphing technique made no meaningful contribution to the mean squared error of $\hat{Z}_{MT}$ and the comparison between $\hat{Z}_{LIS}$ and $\hat{Z}_{MT}$ will be done strictly on the coefficient of variation of the estimators. The difference in the performance of the canopy cover estimators is summarized in Figures 6 and 7 for the Fraser data set. Figure 6 shows three interesting points: The first is that the coefficient of variation of the two estimators is equal only when L is much greater than ρ. The RE values can also be derived from Figure 6, with RE being the squared ratio of CV_{LIS} to CV_{MT}. The range of RE values across the entire range of plot radii and line lengths was RE = 0.56–10.1, which shows that only about 60% as many 22.9 m lines are needed to achieve an equal variance as would be achieved from an inventory that used 3.05 m fixed-area plots. At the other end of the scale, more than 10 times as many L = 3.05 m lines are needed to achieve a sampling variance that equaled that derived from an inventory that used 6.1 m fixed-area plots. The second point of interest is that the coefficient of variation of $\hat{Z}_{MT}$ decreases at a rate much faster than that of $\hat{Z}_{LIS}$. For example, when ρ = 3.05 m (10 ft) the coefficient of variation of the two estimators is equal when $L \approx 11.25$ m (37 ft). When the plot radius is increased by less than a third of a meter (1 ft), the value of L needed to equalize the variance is approximately L = 14.6 m (48 ft). The final point of interest is that the coefficient of variation for $\hat{Z}_{MT}$

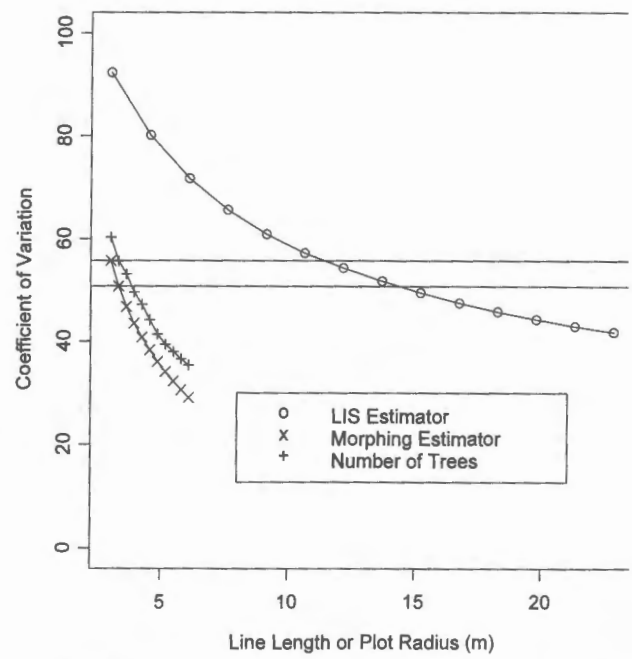
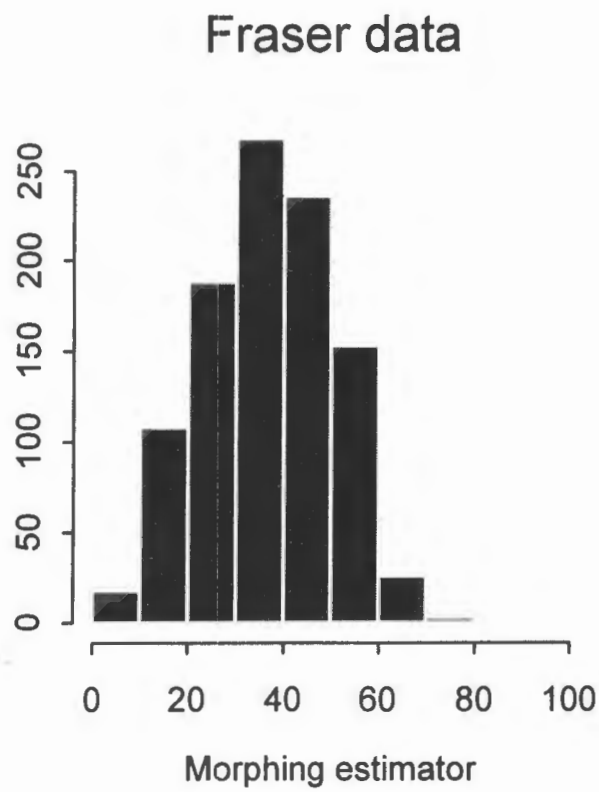


Figure 6. Comparison of the coefficient of variation for the Fraser data set. The line intersect sampling estimator (Z_{LIS}) is graphed against the morphed and torus corrected estimator (Z_{MT}) as a function of line length L and fixed-area plot radius ρ. Where the horizontal lines intersect the two curves is the line length and plot radius at which approximately equal standard errors are achieved for this data set. The relative efficiency (*RE*) is the squared difference between the two curves.

is smaller than $\hat{Z}_N$. Figure 7 shows the empirical sampling distribution for $\hat{Z}_{MT}$ and $\hat{Z}_{LIS}$ using a plot radius of $\rho = 4.6$ m (15 ft) and a line length of 16.8 m (55 ft). The shape of the two distributions is similar, with the one for LIS having slightly fatter tails.

Figures 8 and 9 summarize the results of the simulation study when the New Jersey data set was used. The range of RE values across the entire range of plot radii and line lengths was RE = 13.0–53.8, which shows that even in the least favorable situation for the morphing technique (ρ = 3.05 m and L = 22.9 m) the sample size for LIS needs to be 13 times larger to achieve an equal variance. As shown in Figure 8, there is no case in which the coefficient of variation of $\hat{Z}_{LIS}$ is even remotely similar to that of $\hat{Z}_{MT}$. This occurs because a large number of the samples are such that the line is either almost completely uncovered or completely covered by tree crowns, which results in numerous estimates where $\hat{Z}_{LIS}$ is either less than 10% or greater than 90%. This result is summarized in Figure 9, which compares the empirical sampling distributions of $\hat{Z}_{LIS}$ and $\hat{Z}_{MT}$. In this figure, the bimodal distribution of estimates for LIS is clearly visible. From this we conclude that the LIS estimator can have a very large variance in any situation where the degree of canopy cover is moderate (30–70%) and the spatial distribution of the forest is such that it is likely that some lines are either completely covered or not covered by tree crowns. In these situations, the number of LIS samples required to generate estimates of reasonable precision may be much larger than any of the results in the literature would suggest, with the number of



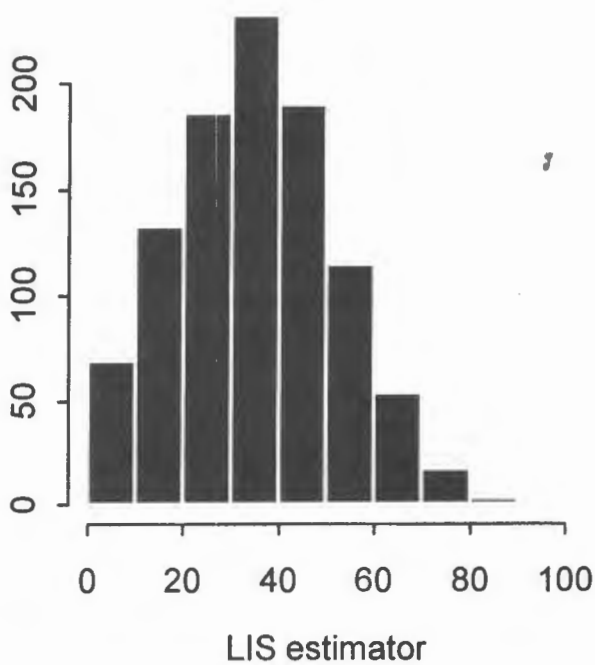


Figure 7. Empirical sampling distribution generated from 1,000 samples drawn from the Fraser data set. The line length and plot radius were L = 16.8 m (55 ft) and 4.6 m (15 ft), respectively.

samples easily being in excess of $n = 20$. Another point of interest is that the coefficient of variation for $\hat{Z}_{MT}$ is again smaller than $\hat{Z}_N$ for this data set. Also note that the rate of the reduction in the coefficient of variation is similar for

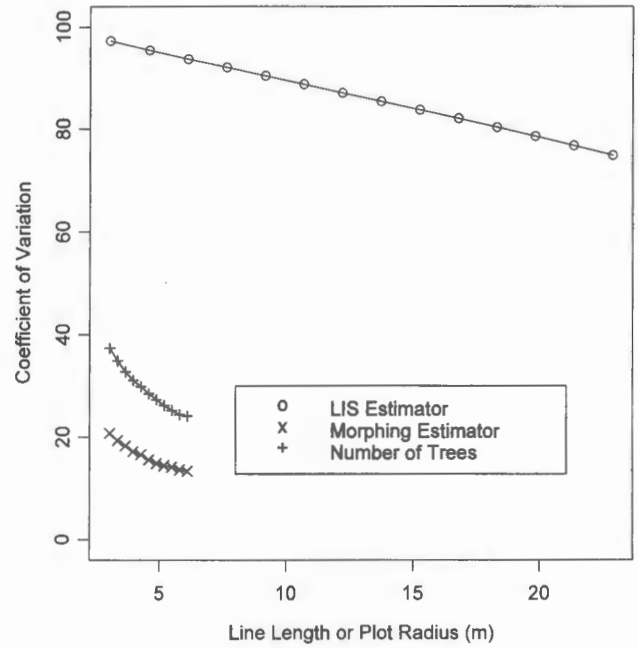


Figure 8. Comparison of the coefficient of variation for the New Jersey data set. The line intersect sampling estimator (Z_{LIS}) is compared to the morphed and torus corrected estimator (Z_{MT}) as a function of line length L and fixed-area plot radius ρ. The relative efficiency (*RE*) is the squared difference between the two curves.

both $\hat{Z}_{MT}$ and $\hat{Z}_N$. Thus, it seems likely that canopy cover can be estimated using $\hat{Z}_{MT}$ with a degree of precision that is better than many of the common forest attributes (e.g., number of trees, basal area, and volume).

For the sample size $n = 1$, the coefficient of variation for $\hat{Z}_{DOT}$ is

$$CV_{DOT} = 100\frac{\sqrt{Z_{CC}(1-Z_{CC})}}{Z_{CC}}$$

where Z_{CC} is the true canopy cover. Using this result, the coefficient of variation is 134.5 and 102.7 for the Fraser and New Jersey data sets respectively. Figures 7 and 9 can provide a better understanding of the performance of the dot count method ($\hat{Z}_{DOT}$) in comparison with both LIS and the morphing technique. This is because the empirical distribution of $\hat{Z}_{DOT}$ for $n = 1$ is a Bernoulli random variable whose discrete distribution places all of the mass on the points 0 and 100%. Thus, $\hat{Z}_{DOT}$ is in some sense the maximum variance unbiased estimator of canopy cover and can never have a smaller variance than either $\hat{Z}_{MT}$ or $\hat{Z}_{LIS}$ for an equal number of sample locations (n). Comparing the *RE* of $\hat{Z}_{DOT}$ against the other two methods yields two interesting results. The first is that $\hat{Z}_{DOT}$ tends to be very inefficient when compared to $\hat{Z}_{MT}$, with the relative efficiency ranging from *RE* = 5.8–60.0, and the smallest *RE* occurring for the Fraser data. To try to put this in perspective, note that even in the worst case scenario (*RE* = 5.8) an inventory crew would have to establish six random locations within the 0.58 ha area of the Fraser data set in the same amount of time as it took to travel to one

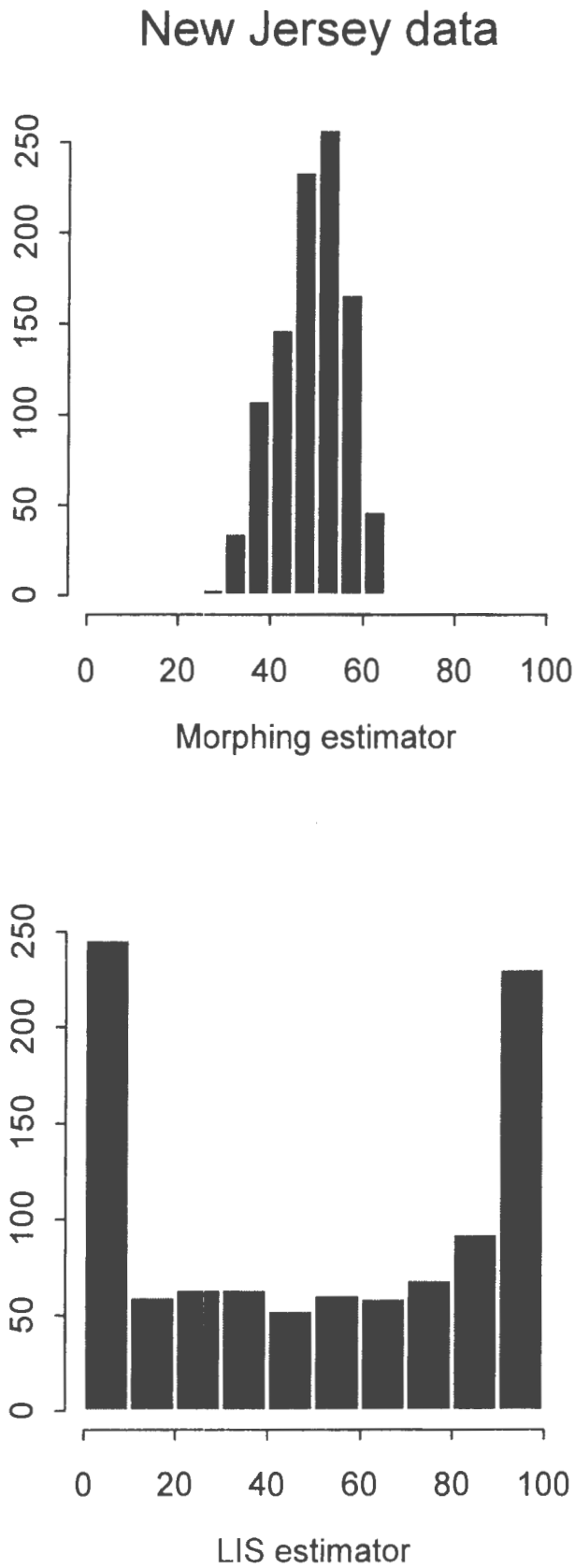


Figure 9. Empirical sampling distribution generated from 1,000 samples drawn from the New Jersey data set. The line length and plot radius were L = 16.8 m (55 ft) and 4.6 m (15 ft), respectively.

location and stem- and crown map approximately 6 trees. (For the Fraser data set, an average of approximately 5.6 trees were sampled on each plot when the plot radius was $\rho = 3.05$ m.)

When comparing $\hat{Z}_{DOT}$ and $\hat{Z}_{LIS}$, the relative efficiency ranged from $RE = 1.1$–10.3, with the smallest RE occurring for the New Jersey data set. The most interesting point is that $\hat{Z}_{LIS}$ was only slighty more efficient for the New Jersey data set, with the maximum RE being 1.9 for $L = 22.9$ m. Thus, it can be concluded that situations exist where $\hat{Z}_{LIS}$ may not perform much better than $\hat{Z}_{DOT}$.

Discussion

The results of this study and experience with numerous other mensuration problems suggests that it is usually wise to match the dimension of the attribute and the measurement. Thus, when attempting to estimate area, a sampling strategy based on a fixed-area plot (two dimensions) will probably be superior to a sampling strategy that employs either lines or points, though counter examples exist (Williams and Patterson 2003). This observation also helps to explain why the dot count canopy cover estimator tends to have a much higher variance than the LIS estimator given an equal number of sample points (n). Whether this observation can be extended to other applications where either points, lines, or fixed-area plots can be used (e.g., estimating coarse woody debris or other canopy related attributes) cannot be determined without further study.

One of the advantages of the morphing technique is that it can be used to estimate canopy cover without actually measuring the crown diameters on the plot. Numerous authors (e.g., Gill et al. 2000, Gering and May 1995, and references therein) have found that crown diameter is strongly correlated with diameter at breast height, and models already exist for most common species. These models can also be used in conjunction with the morphing technique to estimate canopy cover for existing stem-mapped data sets. An example of such a data set is the approximately 120,000 forested ground plots measured by the FIA program. Another advantage of this technique is that because fixed-area plots and the pixels of a satellite image are both two dimensional, the linkage between a canopy cover measurement on a fixed-area plot and the classification of a pixel or group of pixels is likely to be better than the linkage between a LIS estimate and the same image.

The primary disadvantage of the morphing technique is determining how to best model plots that straddle either a forest/nonforest boundary or stand boundaries where the structure of the stands is dissimilar. Provided the boundary is relatively straight, the data can be rotated on the circular plot so that the boundary is maintained within the replicated data set, and the crowns that would cover the nonforested portion of the plot could be deleted. Another solution would be to morph the circular plot into a square and reflect the trees across a randomly placed line as described in Radtke and Burkhart (1998). A plot that covers a corner point or irregular boundary will prove to be problematic with no clear method for implementing the

morphing technique. In these cases, it may be advantageous to actually measure the true canopy cover for this subset of the plots because the reduced tree tally probably offsets the cost of additional measurements.

Another concern with the morphing technique is the potential for an additional bias due to irregularly shaped crowns. This bias occurs any time it is assumed that crowns are circular. This problem is similar to the bias associated with all of the common basal area measurements (see Matern 1990, and references therein).

Comparing LIS and the morphing technique in terms of mean squared error and the cost of data collection is not straightforward. For an inventory such as FIA, where the four large fixed-area subplots ($\rho = 7.3$ m = 24 ft) at each sampling location are already stem mapped, the only additional cost is the crown diameter measurement on each tree. If crown diameters are estimated from a regression model, then there is essentially no additional cost. Thus, if the data are already stem mapped, it is hard to imagine how the dot count or LIS would ever be competitive in terms of overall efficiency (e.g., mean square error and cost). On the other hand, if the purpose of the survey was only to estimate canopy cover for a stand, it is possible for LIS to be more efficient in terms of both mean square error and overall cost. This could occur because only the edges of the connected crown masses along the line require a measurement, rather than having to measure each crown diameter and tree location. Thus, in a stand with nearly 100% canopy cover, there would be little variation between LIS samples and only a small number of measurements would be required with LIS.

Conclusions

The morphing technique is a flexible modeling technique that can be used to estimate canopy attributes that might otherwise be ignored due to the additional cost and complexity required for estimation. It is expected that the greatest difference in the mean squared errors of $\hat{Y}_{LIS}$ and $\hat{Y}_{MT}$ will occur when the true canopy covers falls in the 30 to 70% range and the spatial arrangement of tree tends to be regular. Simulation results suggest that the bias of an estimator based on the morphing technique is likely to be small when estimating canopy cover. Thus, while design-unbiased estimators, such as $\hat{Z}_{LIS}$, have many theoretical advantages, they are not likely to be competitive in terms of mean squared error without a substantial amount of field work.

Software to implement the morphing technique is available. Two versions are available. The first, which is written in *S*, is available at http://lib.stat.cmu.edu/S/morph. FORTRAN subroutines are available from the authors.

Literature Cited

Campbell, G.S., and J.M. Norman. 1998. An introduction to environmental biophysics. Ed. 2. Springer-Verlag, New York.

Cressie, N.A.C. 1993. Statistics for spatial data. Wiley, New York. 900 p.

Crookston, N.L., and A.R. Stage. 1999. Percent canopy cover and stand structure statistics from the forest vegetation simulator. USDA For. Serv. Gen. Tech. Rep. RMRS-GTR-24. 11 p.

Fraser, A.R., and P. Van Den Driessche. 1972. Triangles, density and pattern in point population, P. 277–286 in Proc. of the Third Conf. Advisory Group of For. Stat., Int. Union for. Res. Organ., Inst. Nat. Rech. Agric., Jouy-en-Josas, France.

Frayer, W.E., and G.M. Furnival. 1999. Forest survey sampling designs: A history. J. For. 97:4–8.

Ganey, J.L., and W.M. Block. 1994. A comparison of two techniques for estimating canopy closure. West. J. Appl. For. 9:21–23.

Gering, L.R., and D.M. May. 1995. The relationship of diameter at breast height and crown diameter for four species groups in Harding County, Tennessee. South. J. Appl. For. 19:177–181.

Gill, S.J, G.S. Biging, and E.C. Murphy. 2000. Modeling conifer treed crown radius and estimating canopy cover. For. Ecol. Manage. 126:405–416.

Gregoire, T.G. 1982. The unbiasedness of the mirage correction procedure for boundary overlap. For. Sci. 28:504–508.

Gregoire, T.G. 1998. Design-based and model-based inference in survey sampling: Appreciating the difference. Can. J. For. Res. 28:1429–1447.

Gregoire, T.G., and N.S. Monkevich. 1994. The reflection method of line-intersect sampling to eliminate boundary bias. Environ. Ecol. Stat. 1:219–226.

Jennings, S.B., N.D. Brown, and D. Sheil. 1999. Assessing forest canopies and understory illumination: Canopy closure, canopy cover and other measures. Forestry 72:59–73.

Kaiser, L. 1983. Unbiased estimation in line-intersect sampling. Biometrics 39:965–976.

Kershaw, K.A., and J.H.H. Looney. 1985. Quantitative and dynamic plant ecology. Ed. 3. Edward Arnold Publishing, Baltimore, MD. 282 p.

Lappi, J. 1991. Estimating the distribution of a variable measured with error: Stand densities in a forest inventory. Can. J. For. Res. 21:469–473.

Lowman, M.D., and N.M. Nadkarni. 1995. Forest canopies. Academic Press, San Diego, CA. 626 p.

Mandallaz, D., and R. Ye. 1999. Forest inventory with optimal two-phase, two-stage sampling schemes based on the anticipated variance. Can. J. For. Res. 29:1691–1708.

Matern, B. 1990. On the shape of the cross-section of a tree stem: An empirical study of the geometry of mensurational methods. Swedish Univ. Agric. Sci. Sec. For. Biometry. Rep. 28. 46 p.

Nelson, R.F., T.G. Gregoire, and R.G. Oderwald. 1998. The effects of fixed-area plot width on forest canopy height simulation. For. Sci. 44:438–444.

O'Brien, R.A. 1989. Comparison of overstory canopy cover estimates on forest survey plots. USDA For. Serv. Gen. Tech. Rep. INT-GTR-417. 5 p.

Penttinen, A., D. Stoyan, and H.M. Henttonen. 1992. Marked point processes in forest statistics. For. Sci. 38:806–824.

Radtke, P.J., and H.E. Burkhart. 1998. A comparison of methods for edge-bias compensation. Can. J. For. Res. 28:942–945.

Resnick, S.I. 1992. Adventures in stochastic processes. Birkhauser, Boston. 626 p.

RIPLEY, B.D. 1979. Tests of randomness for spatial point processes, J. Royal Stat. Soc. B. 41:368–374.

SONG, B., J. CHEN, P.V DESANKER, D.D. REED, G.A. BRADSHAW, AND J.F. FRANKLIN. 1997. Modeling canopy structure and heterogeneity across scales: From crowns to canopy. For. Ecol. Manage. 96:217–229.

STOYAN, D., W.S. KENDALL, AND J. MECKE. 1995. Stochastic geometry and its applications. Ed. 2. Wiley, New York. 436 p.

VALENTINE, H.T., J.H. GOVE, AND T.G. GREGOIRE. 2001. Monte Carlo approaches to sampling forested tracts with lines or points. Can. J. For. Res. 31:1410–1424.

WILLIAMS, M.S., H.T. SCHREUDER, AND R.L. CZAPLEWSKI. 2001a. Accuracy and efficiency of area classifications based on tree tally. Can. J. For. Res. 31:556–560.

WILLIAMS, M.S., M.T. WILLIAMS, AND H.T. MOWRER. 2001b. A boundary reconstruction method for circular fixed-area plots in environmental survey. J. Agric. Biol. and Environ. Stat. 6:479–494.

WILLIAMS, M.S., AND P.L. PATTERSON. 2003. Comparison of area estimation techniques for established surveys using remotely sensed and ground data. For. Sci. (in press).