Environmental and Ecological Statistics **10,** 263–279, 2003

Line intersect sampling: Ell-shaped transects and multiple intersections

TIMOTHY G. GREGOIRE[1] and HARRY T. VALENTINE[2]

[1]*School of Forestry and Environmental Studies, Yale University, New Haven, CT 06511-2189, U.S.A.*
E-mail: timothy.gregoire@yale.edu
[2]*USDA Forest Service, Northeastern Research Station, Durham, NH 03824-0640, U.S.A.*
E-mail: hvalentine@fs.fed.us

Received January 2002; Accepted June 2002

The probability of selecting a population element under line intersect sampling depends on the width of the particle in the direction perpendicular to the transect, as is well known. The consequence of this when using ell-shaped transects rather than straight-line transects are explicated, and modifications that preserve design-unbiasedness of Kaiser's (1983) conditional and unconditional estimators are presented. A case against treating multiple intersections as multiple probabilistic events is argued on the basis, also, of preserving design-unbiased estimation.

Keywords: conditional estimator, design-unbiasedness, line intercept sampling, particles

1352-8505 © 2003 Kluwer Academic Publishers

1. Introduction

Since its introduction to ecologists by Canfield (1941), line intersect sampling has found widespread application for the purpose of estimating plant abundance and cover. Also known as line intercept sampling (despite a quibbling distinction between the two terms by Pielou, 1985), the technique of using a line as a sampling unit and measuring some feature of the objects crossed by it has origins in many other fields, indeed some predating Canfield's exposition by many years. Much recent attention on monitoring ecosystems for biodiversity has focused on the use of line intersect sampling (LIS) for the assessment of coarse woody debris (CWD), e.g., Marshall *et al.* (2000). Ringvall and Ståhl's (1999) study of errors in the field implementation of LIS to assess CWD is a refreshing reminder that the reported accuracy of estimates of population attributes often will be overly optimistic.

From a statistical viewpoint, some of the literature on LIS is misleading, or worse, incorrect. In particular, misconceptions about the theoretical and probabilistic underpinnings of LIS appear distressingly often. Most authors fail both to distinguish design-based inference from model-based inference and to stipulate the inferential paradigm they implicitly invoke. An exception is Kaiser (1983) who treated the objects comprising the population as fixed, thereby obviating the need for the oft stated and unrealistic assumption that the objects to be sampled are distributed and oriented uniformly at random over the region of interest. Instead he derived design-unbiased estimators of population attributes by assuming that randomness arises only through the random location, and possibly random orientation, of the sample transects.

While the work reported here was prompted by consideration of LIS for CWD assessment, wherein needle-shaped logs, branches, and twigs constitute the population of interest, we adopt the generality of Kaiser by considering a population of arbitrarily shaped objects, termed particles. Likewise we take a design-based approach to inference, which implies that the sampling properties of estimators are derived from the randomization distribution of estimates from all, and perhaps infinite number of, samples possible under the sampling design (*vide* Gregoire, 1998; Dorazio, 1999). In addition to straight-line transects we consider ell-shaped transects formed by two line segments joined perpendicularly, as in Fig. 1. Ell-shaped transects have been discussed elsewhere (Hazard and Pickford, 1986; Marshall *et al.*, 2000) with the apparent presumption that the same estimators as developed for straight-line transects would be used. So doing may sacrifice design-unbiasedness, as the probability of selecting a particular particle with an ell-shaped

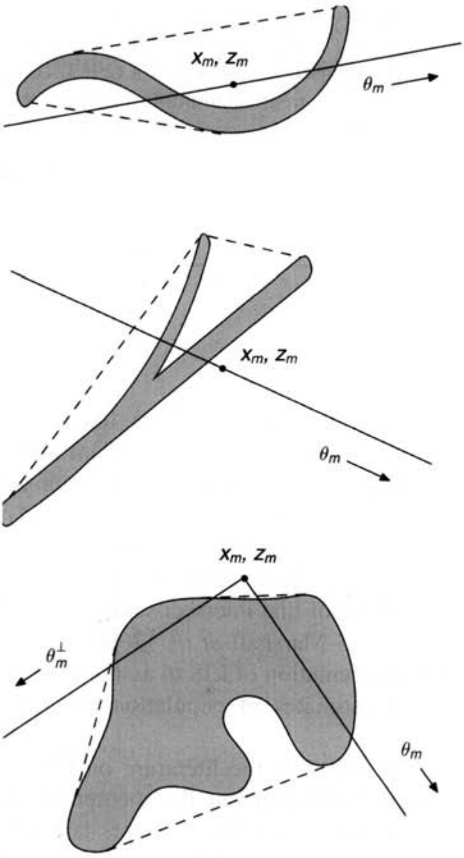

Figure 1. Example intersections of particles by straight-line transects and ell-shaped transects. Transect is located by (x_m, z_m) and oriented by θ_m. The dashed lines indicate the convex hulls of the respective particles.

transect may differ from what it is when using a straight-line transect. We present a sampling protocol and corresponding estimators which specifically account for a particle's selection probability when using ell-shaped transects. Following Kaiser (1983) we consider estimation conditionally upon the orientation of the sampling transects, as well as unconditionally over all possible orientations.

We also consider the case where a straight-line transect or segment of an ell-shaped transect intersects a particle more than once as illustrated in Fig. 1. Specifically, we examine, and are critical of, the practice of treating each intersection of a particle as an independent observation, as recommended in Marshall *et al.* (2000).

2. Notation

The notation developed in this and later sections is summarized in Appendix B.

Let the population of interest, $\mathscr{P}$, be the set of N discrete particles: $\mathscr{P} = \{P_1, \dots, P_N\}$. Each particle is assumed to be connected in the sense of Buck (1965, page 29), but nothing about the shape of a particle is presumed. Nor is anything about the spatial distribution of $\mathscr{P}$ on the planar region of interest, $\mathscr{A}$, assumed; indeed particles may overlap. We do not deal explicitly with LIS in a stratified population other than to mention that stratification may be usefully employed as a design strategy to create strata within which the spatial distribution of $\mathscr{P}$ is fairly homogeneous.

Remark 1 *There appears to be abiding confusion in the LIS literature over the necessity and desirability of assuming that the particles are randomly located and oriented in $\mathscr{A}$. While inference about $\mathscr{P}$ following LIS surely may be based on a model of complete spatial randomness, the implicit adoption of such a model hardly is necessary. Eberhardt (1978) recognized but did not elaborate on this point. Kaiser (1983) addressed the issue directly. We emphasize that the random placement of the line transects presumed in the present work is not in any way intended as a proxy for a failed assumption of random placement of particles in $\mathscr{A}$, but instead is a deliberate and direct appeal to classical sampling theory, as embodied* inter alia *in Cochran (1977), which bases inference solely on the distribution of possible estimates under the design.*

Associated with the i-th particle, P_i, is a measurable characteristic, say y_i, whose value is fixed in the sense that it in no way depends on whether P_i is included in the sample or how it is intersected by a transect. The total y in the population is

$$\tau = \sum_{i \in \mathscr{P}} y_i.$$

Letting the area of $\mathscr{A}$ be denoted by A, a population descriptive parameter to be estimated from the sample data might be the density of $\mathscr{P}$ on $\mathscr{A}$:

$$\lambda = \frac{\tau}{A}.$$

Thus, when $y_i = 1$ for all P_i, then λ is the usual expression of population density in terms of the number of individuals per unit area, *viz.* $\lambda \equiv \lambda_N = N/A$. Alternatively, when y_i is the cross-sectional area of P_i when orthogonally projected onto $\mathscr{A}$, then λ is the conventional expression of cover (*vide* McIntyre, 1953). While these two choices for y are special cases of common interest, we do not restrict ourselves to just these. Indeed, y_i may take on values of just 1 or 0, indicating respectively the presence or absence of some attribute. In this case it might be of interest to estimate the proportion, μ, of $\mathscr{P}$ with this attribute:

$$\mu = \frac{\tau}{N}.$$

For more general y_i, μ is a measure of mean size. In this article, we focus on the estimation of λ, μ, and N.

We imagine that a sample of M transects, each of fixed length L, is established on $\mathscr{A}$, and that the m-th of these is located by the coordinate pair (x_m, z_m) in $\mathscr{A}$. For sake of specificity, we adopt the protocol that this coordinate location of the transect coincides with the midpoint of a straight-line transect (see Fig. 1). For an ell-shaped transect, we specify the coordinate location as the join point of the two segments, where each segment has length $L/2$. We assume that edge conditions are handled in the manner described by Kaiser (1983), the reflection method introduced by Gregoire and Monkevich (1994), or some other means that does not insinuate a design-bias into the estimators we consider.

We further assume that the transect locations, $(x_m, z_m, m = 1, \dots, M)$ are assigned at random in $\mathscr{A}$, or that at least one transect is randomly located and the remaining transects are spaced systematically around it. While it is unlikely that any particular particle will be intersected by more than one line, nothing in the sampling protocol prevents this event from occurring. Hence, in the common vernacular of the sample survey literature (*cf.* Cochran, 1977), LIS is a form of sampling with replacement.

Let θ_m denote the orientation of the m-th transect with respect to some arbitrarily established reference direction $\theta = 0$, as illustrated in Fig. 2. For ell-shaped transects we assume that one segment has direction θ_m, and that this serves to indicate the orientation of the transect (see Fig. 3). The direction of the other segment is $\theta_m^{\perp} = \theta_m \pm \pi/2$.

Remark 2 *It is advisable to adopt a consistent protocol which assigns θ_m and $\theta_m^{\perp}$ in advance of sampling to the two segments; e.g., from the vertex, let θ_m orient one segment of the ell, and then orient the other segment always in direction $\theta_m^{\perp} = \theta_m - \pi/2$ or always in direction $\theta_m^{\perp} = \theta_m + \pi/2$. We advise particularly against letting field conditions influence the orientation assigned to one or the other segment of the transect, because this practice makes it impossible to truly discern the selection probabilities of particles in $\mathscr{A}$.*

In the following we consider two cases. In one, θ_m is selected uniformly at random over the interval $[0, \pi)$ for each transect, and henceforth denoted as $\theta_m \sim U[0, \pi]$; in the other, θ_m is fixed in advance of sampling at some common value, say $\theta_m = \tilde{\theta}$ for $m = 1, \dots, M$.

Regardless of whether θ_m is random or fixed a priori, we let $\widehat{\lambda}^c$ denote the estimator of λ conditionally on θ_m, *viz.*

$$\widehat{\lambda}^c = \frac{1}{M} \sum_{m=1}^{M} \widehat{\lambda}_m^c, \tag{1}$$

where $\widehat{\lambda}_m^c$ is the conditional estimator of λ based on the m-th transect only. Exact expressions for $\widehat{\lambda}_m^c$ are given in the next section.

Similarly, but only for the case where $\theta_m \sim U[0, \pi]$, let $\widehat{\lambda}_m^u$ denote the unconditional

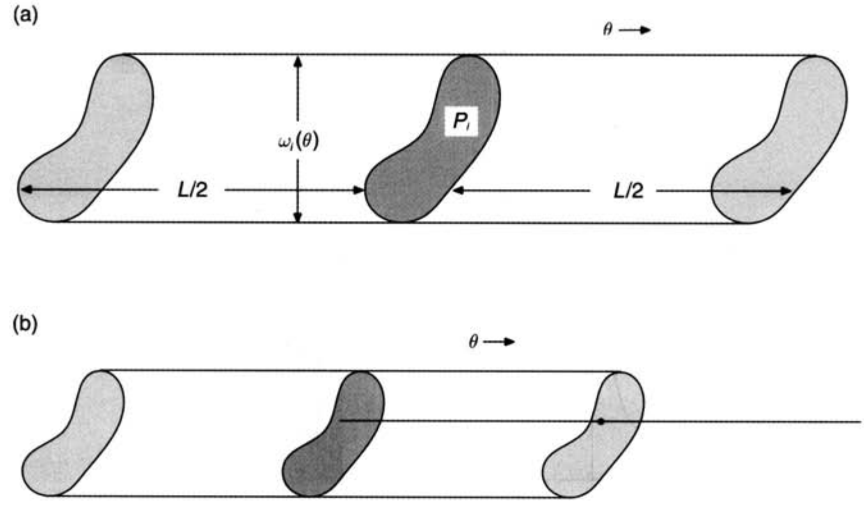

Figure 2. (a) Inclusion area of particle P_i for a straight-line transect whose location is its midpoint; (b) an example of a straight-line transect which partially intersects P_i.

estimator of λ based on the m-th transect only, i.e., the estimator which accounts for the randomness of θ_m. These then are averaged over all M transects yielding

$$\widehat{\lambda}^u = \frac{1}{M}\sum_{m=1}^{M}\widehat{\lambda}^u_m. \tag{2}$$

In an evident fashion, $\widehat{\mu}^c$ and $\widehat{\mu}^u$ will be used to denote the conditional and unconditional estimators of μ.

When transects are randomly located, the sampling variance of $\widehat{\lambda}^c$ can be estimated design-unbiasedly by

$$\widehat{\mathrm{var}}\left(\widehat{\lambda}^c\right) = \frac{1}{M(M-1)}\sum_{m=1}^{M}\left(\widehat{\lambda}^c_m - \widehat{\lambda}^c\right)^2.$$

Similarly,

$$\widehat{\mathrm{var}}\left(\widehat{\lambda}^u\right) = \frac{1}{M(M-1)}\sum_{m=1}^{M}\left(\widehat{\lambda}^u_m - \widehat{\lambda}^u\right)^2,$$

provides a design-unbiased estimator of the variance of $\widehat{\lambda}^u$. Multiplication by $(A/N)^2$ converts the above to variance estimators of $\widehat{\mu}^c$ and $\widehat{\mu}^u$, respectively. Under systematic placement of transects, these variance estimators are likely to overstate the sampling variance of $\widehat{\lambda}^c$ and $\widehat{\lambda}^u$, respectively.

3. Estimation of density

We consider P_i to be in the sample selected by the m-th transect if it is intersected completely by it, or with some restrictions, partially intersected by it. In the latter case we

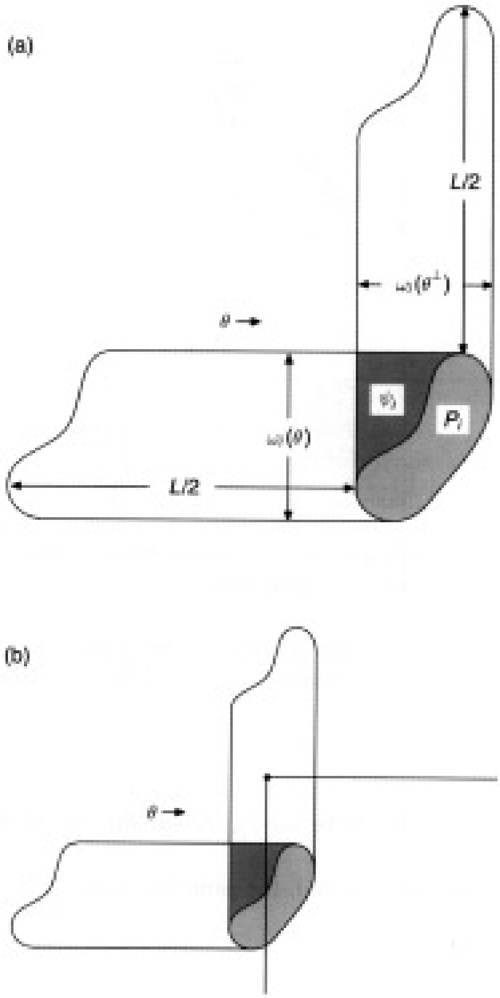

Figure 3. (a) Inclusion area of particle P_j for a ell-shaped-line transect whose location is its vertex; (b) an example of an ell-shaped transect which completely intersects P_j.

assume that the transect length is sufficiently long to preclude both ends of a straight-line transect or both ends of a segment of an ell-shaped transect from falling within P_j. The restriction is that a partial intersection of P_j qualifies it for inclusion in the sample with probability 1/2, as per McIntyre (1953), Lucas and Seber (1977), and Kaiser (1983). This can be accomplished by choosing one end of each transect in advance of sampling to be the end for which partial intersections will be ignored. Or one can decide in an adaptive fashion by flipping a fair coin upon each occurrence in the field of a partial intersection.

For ell-shaped transects, a broader spectrum of intersections must be considered. For sake of specificity we adopt the convention of allowing only those partial intersections that do not involve the vertex of the transect, i.e., we consider particles to be excluded from the sample if the vertex happens to be located beneath P_j.

For the m-th straight-line transect located at its midpoint (x_m, z_m), P_j will be included into the sample if (x_m, z_m) falls within the unshaded portion outside P_i, and it will be included with probability 1/2 if (x_m, z_m) falls within the lightly-shaded regions outside P_i (Fig. 2(b)). Note that the particle, P_i, is partially intersected whenever the midpoint of the transect line falls in one of the lightly shaded areas depicted in Fig. 2(b). If we let $t_{im} = 1$ if P_i is included in the sample from the m-th straight-line transect, and $t_{im} = 0$ otherwise, then as demonstrated by Kaiser (1983)

$$\mathrm{Prob}(t_{im} = 1|\theta_m) = \frac{Lw_i(\theta_m)}{A},$$

where $w_i(\theta_m)$ is the width of P_i perpendicular to θ_m. (We assume trivially that all measures of distance such as L and $w_i(\theta_m)$ have identical units, which multiply together to give a measure of area with identical units to that of A. This avoids the notational messiness of having to account explicitly for conversion factors, which, if neglected, can cause non-trivial aberration in the computed estimates!)

When using ell-shaped transects, P_i will be included in the sample from the mth transect if (x_m, z_m) falls within the shaded or unshaded area around P_i (Fig. 3(a)). If (x_m, z_m) falls within the shaded region, then P_i will be completely intersected by both segments of the transect; let the area of this region be denoted by $\psi_i(\theta_m, \theta_m^\perp)$. In parallel fashion to the development for straight-line transects, let $t_{im} = 1$ if P_i is intersected by only one segment of the transect; and $t_{im} = 2$ if it is intersected by both segments; and $t_{im} = 0$ otherwise. Therefore,

$$\mathrm{Prob}(t_{im} = 1|\theta_m) = \left(\left(\frac{L}{2}w_i(\theta_m) - \psi_i(\theta_m, \theta_m^\perp)\right)\right.$$
$$\left.+\left(\frac{L}{2}w_i(\theta_m^\perp) - \psi_i(\theta_m, \theta_m^\perp)\right)\right)/A,$$

where $w_i(\theta_m)$ is the width of P_i perpendicular to θ_m and $w_i(\theta_m^\perp)$ is its width perpendicular to $\theta_m^\perp$. Also,

$$\mathrm{Prob}(t_{im} = 2|\theta_m) = \frac{\psi_i(\theta_m, \theta_m^\perp)}{A}.$$

These probability results for the random variable t_{im} can be used to derive its expected value conditionally upon the orientation of the m-th transect. For a straight-line transect the two coincide:

$$E[t_{im}|\theta_m] = \mathrm{Prob}(t_{im} = 1|\theta_m) = \frac{Lw_i(\theta_m)}{A},$$

whereas for the ell-shaped transect it is

$$E[t_{im}|\theta_m] = \mathrm{Prob}(t_{im} = 1|\theta_m) + 2 \times \mathrm{Prob}(t_{im} = 2|\theta_m) = \frac{L\overline{w}_i(\theta_m)}{A}, \tag{3}$$

where $\overline{w}_i(\theta_m) = \left(w_i(\theta_m) + w_i(\theta_m^\perp)\right)/2$.

A design-unbiased estimator of λ which is conditional on the orientation of the m-th transect is

$$\widehat{\lambda}_m^c = \frac{1}{A}\sum_{i\in\mathscr{P}} \frac{t_{im}y_i}{E[t_{im}|\theta_m]}, \tag{4}$$

where $i \in \mathscr{P}$ is used to indicate that the summation extends over all particles, P_i, in the population. Evidently, (4) is a linear, homogeneous estimator (*vide* Heyadat and Sinha, 1991, page 23). For straight-line transects (4) collapses to $\widehat{\lambda}_m^c = 1/L\sum_{i\in\mathscr{S}_m} y_i/w_i(\theta_m)$, where $i \in \mathscr{S}_m$ indicates summation over all particles, P_i, included in the sample from the m-th transect; for ell-shaped transects it can be simplified only to $\widehat{\lambda}_m^c = 1/L\sum_{i\in\mathscr{S}_m} t_{im}y_i/\overline{w}_i(\theta_m)$, because t_{im} will either be 1 or 2 depending on how a sampled particle is intersected. The implication of the latter result for field implementation of ell-shaped transects is that the maximal width of P_i perpendicular to both θ_m and $\theta_m^\perp$ must be measured regardless of whether P_i is intersected by both segments of the transect (multiple intersections of P_i by a straight-line transect or by one segment of an ell-shaped transect are discussed in Section 6.

For the case where $\theta_m \sim U[0,\pi], m = 1,\ldots,M$, one could consider, as an alternative to λ_m^c, an estimator that unbiasedly estimates λ when accounting not only for all possible locations but also all orientations of sample transects. For this alternative estimator, the unconditional expected value of t_{im} is needed. Since $E[w_i(\theta_m)] = E[w_i(\theta_m^\perp)] = E[w_i]$, say, then $E[t_{im}] = L \times E[w_i]/A$ for both straight-line and ell-shaped transects, providing that the procedures for dealing with partial intersections are those described above. (This result is closely related to the Buffon needle problem (*vide* Beckmann, 1971).) Accordingly, the unconditional estimator of λ is

$$\begin{aligned}\widehat{\lambda}_m^u &= \frac{1}{A}\sum_{i\in\mathscr{S}_m} \frac{t_{im}y_i}{E[t_{im}]} \\ &= \frac{1}{L}\sum_{i\in\mathscr{S}_m} \frac{t_{im}y_i}{E[w_i]}.\end{aligned} \tag{5}$$

Example 1 *Using (4) for the case where $y_i = 1$ for all particles yields a conditional estimator of $\lambda_N = N/A$ as $\widehat{\lambda}_N^c = \widehat{N}^c/A$, where for the mth straight-line transect*

$$\widehat{N}_m^c = \frac{A}{L}\sum_{i\in\mathscr{S}_m} \frac{1}{w_i(\theta_m)};$$

and for the m-th ell-shaped transect it is

$$\widehat{N}_m^c = \frac{A}{L}\sum_{i\in\mathscr{S}_m} \frac{t_{im}}{\overline{w}_i(\theta_m)};$$

in both cases leading to

$$\widehat{N}^c = \frac{1}{M}\sum_{m=1}^{M} \widehat{N}_m^c.$$

An analogous unconditional estimator of λ_N replaces $w_i(\theta_m)$ and $\overline{w}_i(\theta_m)$ in $\widehat{N}_m^c$ with $E[w_i]$.

Remark 3 *Implicit in Example 1 is that the area, A, of the region being surveyed need not be known to estimate λ, and in particular λ_N. However, A must be known in order to* *estimate N as in the above example, and to estimate τ more generally, at least for the implementation of LIS we have described. Realistically, the boundaries of $\mathscr{A}$ must be known, even if its area, A, must be determined later.*

Depending on the application, $w_i(\theta)$, the width of the particle perpendicular to the actual orientation of the transect, may be easier to measure than $E[w_i]$, the particle width averaged over all possible orientations. However, for connected particles and when $\theta_m \sim U[0, \pi]$, $E[w_i]$ is calculable as $E[w_i] = c_i/\pi$, where c_i is the circumference of the convex hull enclosing P_i. This circumference can be measured by wrapping a tape measure around the outer edge of P_i as projected onto the plane $\mathscr{A}$ (the length of the dashed line surrounding each particle in Fig. 1). To foresters this is the familiar result that a random calipering of tree diameter is identical, on average, to a measurement with tape of tree girth, divided by π (Matérn, 1956). Moreover, with a needle-shaped particle whose length from end to end, l_i, substantially exceeds its "diameter", then $c_i = 2l_i$ to a suitable degree of approximation. Hence

$$\widehat{\lambda}_m^u = \frac{\pi}{2L} \sum_{i \in \mathscr{P}} \frac{t_{im} y_i}{l_i}, \tag{6}$$

as has appeared in Marshall *et al.* (2000) and elsewhere in the forestry literature.

Remark 4 *Many investigators likely would opt to use the conditional estimator when all transects are oriented in a direction perpendicular to some baseline or some gradient evident in the distribution of P_i, as in Warren and Olsen (1964). The conditional estimator may be preferred, also, in the situation where the common orientation angle, $\tilde{\theta}$, is a realization of a $U[0, \pi]$ random selection. However for the latter case $\widehat{\lambda}^u$ would be appropriate, also. Conversely, even when the θ_m are selected uniformly at random separately for each transect, one might opt to use the conditional rather than the unconditional estimator of λ if the requisite field measurements were easier to garner. A choice between $\widehat{\lambda}^c$ and $\widehat{\lambda}^u$ may also be made on the grounds of which is more precise. Although this may be difficult to discern in any particular situation, one is free to select that estimator with the smaller estimated sampling variance without risk of insinuating bias.*

4. Estimation of mean size

For general y_i, an estimator of mean size, μ, conditional on the orientations of the sample transects is

$$\widehat{\mu}^c = \frac{\widehat{\lambda}^c}{\widehat{\lambda}_N^c} = \frac{A\widehat{\lambda}^c}{\widehat{N}^c}, \tag{7}$$

where $\widehat{\lambda}^c$ is defined in (1) and $\widehat{\lambda}_N^c$ is developed in the first example. An unconditional estimator of μ is given by

$$\widehat{\mu}^u = \frac{\widehat{\lambda}^u}{\widehat{\lambda}_N^u} = \frac{A\widehat{\lambda}^u}{\widehat{N}^u}. \tag{8}$$

Both (7) and (8) are design-biased estimators of μ. The design-bias of both is negligible for large M with the sampling design described above.

Remark 5 *We note that even in the unlikely situation where the size of the population, N, is known, (7) is likely to be a more precise estimator of μ than $A\widehat{\lambda}^c/N$, for reasons explained by Särndal et al. (1992, pages 180–81). Similarly, using an estimate of N in (8) rather than N will usually provide a more precise unconditional estimator of μ.*

The naive estimator of μ given by $\widehat{\lambda}^c$ (or $\widehat{\lambda}^u$) divided by the total number of particles in the sample is distinctly ill-advised: not only is its bias likely to be appreciable, it is not even a consistent estimator of μ, sensu Cochran (1977, page 21).

5. Use of an auxiliary variate to obviate measurement of Y

While the estimators (4) and (5) given above are design-unbiased and somewhat familiar by virtue of their simplification in certain applications, they can be improved upon by the introduction of an auxiliary variate, say $q_i(\theta_m)$, whose measure on P_i at the m-th transect may depend on the orientation of the transect. As demonstrated by Kaiser (1983), a judicious choice of $q_i(\theta_m)$ may obviate the need to measure y_i on the sampled P_i. This feature has not been widely recognized nor has its advantage been widely appreciated by ecologists. Kaiser's conditional estimator is

$$\widehat{\lambda}^c_m = \frac{1}{A}\sum_{i\in\mathcal{P}} \frac{t_{im}q_i(\theta_m)y_i}{E[t_{im}q_i(\theta_m)|\theta_m]}, \tag{9}$$

and the unconditional estimator is

$$\widehat{\lambda}^u_m = \frac{1}{A}\sum_{i\in\mathcal{P}} \frac{t_{im}q_i(\theta_m)y_i}{E[t_{im}q_i(\theta_m)]}. \tag{10}$$

An example will illustrate the usefulness of the auxiliary variate.

Example 2 *We suppose that P_i is a plant, more specifically the canopy of a plant projected orthogonally onto the ground. The characteristic of interest, y_i, is its canopy area, a_i, projected onto $\mathcal{A}$. Thus, $\lambda = 1/A\sum_{i\in\mathcal{P}} a_i$, known variously as cover, coverage, and composition. Let $q_i(\theta_m)$ be the length of intersection, say $I_i(\theta_m)$, of P_i with the line in $\mathcal{A}$ which contains the m-th straight-line transect (n.b., $I_i(\theta_m)$ is the length of interception by the transect when P_i is completely intersected; however, the tortuous wording in the definition of $I_i(\theta_m)$ is meant to emphasize that in the case of a partial interception, $I_i(\theta_m)$ is longer than just the interception length.) In this case, the product random variable, $t_{im}q_i(\theta_m)$ has conditional expected value*

$$E[t_{im}q_i(\theta_m)|\theta_m] = \frac{La_i}{A},$$

as derived by Lucas and Seber (1977). Therefore a conditionally unbiased estimator of cover from the m-th straight-line transect is

$$\widehat{\lambda}_m^c = \frac{1}{A}\sum_{i\in\mathcal{S}_m}\frac{I_i(\theta_m)a_i}{La_i/A} = \frac{1}{L}\sum_{i\in\mathcal{S}_m} I_i(\theta_m).$$

For an ell-shaped transect, let $q_i(\theta_m) = I_i(\theta_m)$ if P_i is intersected by the segment of the transect oriented in the θ_m direction; or $q_i(\theta_m) = I_i(\theta_m^{\perp})$ if P_i is intersected by the segment of the transect oriented in the $\theta_m^{\perp}$ direction; or $q_i(\theta_m) = \bar{I}_i(\theta_m) = \left(I_i(\theta_m) + I_i(\theta_m^{\perp})\right)/2$ whenever both segments of the transect intersect P_i. With q_i so defined, $E[t_{im}q_i(\theta_m)|\theta_m] = La_i/A$, for ell-shaped transects, too (see Appendix A). Thus (9) remains design-unbiased and can be simplified to

$$\widehat{\lambda}_m^c = \frac{1}{L}\sum_{i\in\mathcal{S}_m} t_{im}q_i(\theta_m).$$

Because $E[t_{im}q_i(\theta_m)|\theta_m]$ evidently does not depend on θ_m, the unconditional expected value of $t_{im}q_i(\theta_m)$ is La_i/A, also. Thus $\widehat{\lambda}_m^u = \widehat{\lambda}_m^c$ in this case. Hence the auxiliary variate enables unbiased estimation of canopy cover both conditionally on transect orientation, and unconditionally, without the need to measure or devise an approximate measure of the projected canopy area of any individual plant canopy.

The above can be extended to the situation where y_i is the area of a gap in forest cover (*vide* Runkle 1985, 1992; Battles *et al.*, 1996), or area of a body of water on a map. Kaiser (1983) exemplifies other applications of LIS using an auxiliary variate to aid estimation of λ when using straight-line transects. These all can be extended to accommodate ell-shaped transects, too.

6. Dealing with multiple intersections

In Fig. 1 is shown a forked particle that might, for example, represent a piece of CWD; an irregularly shaped particle that might represent a lake or a gap in forest cover; and a long, narrow, $\mathcal{S}$-shaped particle that might represent a stream, a road, or a branch-like piece of CWD. We consider the case where a straight-line transect or a segment of an ell-shaped transect intersects P_i more than once.

For pieces of forked CWD, Marshall *et al.* (2000, page 12) suggested that the two tines of the fork be treated as separate pieces when a single transect intersects both tines. For multiply-intersected $\mathcal{S}$-shaped and $\mathcal{U}$-shaped pieces, they suggest that each "piece is actually treated as ... separate pieces", one for each intersection. There is an inherent ambiguity to this practice, however, that is troublesome: one wonders whether a forked particle is to be treated as two distinct particles in the event that a straight-line transect or a segment of an ell-shaped transect crosses both tines, but only as a single particle when crossed but once beneath the fork. It seems perverse to let the manner in which particles are intersected implicitly define the number, N, of particles in the population $\mathcal{P}$, as N then becomes a random variable. Moreover, for a particle that lacks a clear forking structure, such as the $\mathcal{S}$-shaped particle or the irregularly-shaped particle of Fig. 1, its subdivision into multiple particles would be problematic, because the boundaries separating the subdivisions are not clearly suggested by the shape or structure of the particle. There is an ambiguity, too, in how to assign the portion of the forked piece in Fig. 1a that is below the fork; e.g., if the particle length l_i is to be measured when P_i is crossed by the transect, a protocol must be established that stipulates whether this length include that part of P_i below the fork only when the larger of the two forks is intersected, only the smaller, or both.

We think that connected particles ought not be subdivided further, irrespective of how they may be intersected by the line transect: each discrete but connected particle constitutes one element of $\mathscr{P}$. Furthermore, we think that population descriptive parameters such as λ and μ be estimated in a fashion that does not rely on the artifice of treating each particle as multiple identical particles corresponding to multiple intersections. In direct contrast to DeVries (1979, page 44), we advise, for straight-line transects, that $t_{im} = 1$ in (4) and (5) irrespective of the number of intersections of the m-th transect with P_i; and for ell-shaped transects, that $t_{im} = 1$ or 2 (as explained in Section 3), also irrespective of the number of intersections of each segment of the transect with P_i. With this protocol (4) remains a conditional estimator that is design-unbiased for estimation of λ, and (5) remains unconditionally unbiased. By implication, both (4) and (5) will biasedly estimate λ if multiple intersections of P_i are treated as separate, independent intersections. For sake of clarity, we emphasize that intersections of P_i by two or more transects are to be treated and recorded as separate, independent probabilistic events.

Notwithstanding the above admonition, it might be useful on occasion to let the auxiliary variate, $q_i(\theta_m)$, be the number of intersections of P_i with the m-th transect, as illustrated by the following example.

Example 3 *Let y_i be the boundary length of P_i, so that τ is the total length of perimeter of all particles in $\mathscr{A}$. Let $q_i(\theta_m)$ be the number of times that the boundary of P_i is intersected by the line in $\mathscr{A}$ containing the mth transect. Note that for each of the particles displayed in Fig. 1, $q_i = 4$. As shown by Kaiser (1983, Equation (3.8))*

$$\widehat{\lambda}^u_{ym} = \frac{\pi}{2L} \sum_{i \in \mathscr{S}_m} q_i(\theta_m), \tag{11}$$

is an unconditionally unbiased estimator of λ. Here, the judicious choice of $q_i(\theta_m)$ obviates the need for the more costly measurement of y_i itself. One useful application of this result occurs when particles are fibers or roads within $\mathscr{A}$ having the characteristic that their width is negligible compared to their length. In this case $y_i = 2l_i$, where l_i is the length of the road within $\mathscr{A}$. Thus, $\widehat{\lambda}^u_{ym}/2$ unbiasedly estimates the length of road per unit area in $\mathscr{A}$ (Matérn, 1964).

We caution against reading to much into Example 3. In particular, it is not an endorsement to always use the auxiliary variate, q_i, to record the multiplicity of intersections. It is a useful device for this particular application (estimating total boundary or road length), but it may well be counterproductive for other applications.

As an alternative to (11) to estimate total length of road, one could consider $\widehat{\lambda}^u_m = \pi/L \sum_{i \in \mathscr{S}_m} y_i/c_i$ for both straight-line and ell-shaped transects. For the forked piece in Fig. 1, c_i is the length of the dashed line enclosing the particle. With this estimator, a special case of (5), multiple intersections by a single transect are not counted as more than one intersection; moreover, y_i must be measured.

7. Monte Carlo approach to LIS

The derivation of estimators for LIS also can be approached from the theory of Monte Carlo integration. Valentine *et al.* (2001) distinguished two Monte Carlo approaches to LIS: a line method, which is an application of importance sampling (*vide* Gregoire *et al.*, 1993), and a point method, which is a two-stage application of importance sampling and crude Monte Carlo. The point method is actually simpler, since the two-stage selection of the sample point reduces to choosing coordinates (x, z) at random in $\mathscr{A}$. In effect, the method is a continuous analog of simple random sampling: each sample point is selected with probability density $1/A$ and the attribute density of y (i.e., the amount of y per unit land area) is measured at that point.

Since y is an attribute of discrete particles, its distribution over the land may be spotty, in which case the density of y at a sampling point might have a large probability of being nil over much of the landscape. Consequently, the Monte Carlo sampling would be inefficient. Moreover, if one or more particles do happen to sit atop a sample point, the field measurement of the attribute density of y over that point may be problematic, if not impossible. This inefficiency is alleviated and the mensuration becomes straightforward if we assume that the attribute, y_i, is spread evenly over a region $\mathscr{A}_i(\mathscr{A}_i \in \mathscr{A})$ with area e_i. If any part of $\mathscr{A}_i$ overlaps the sample point, then the attribute density, y_i/e_i, is measured. In the Monte Carlo approach to LIS, the size and shape of $\mathscr{A}_i$ is matched to the length of the transect line, whose role is to serve as an instrument to indicate when the sample point, (x, z), is within $\mathscr{A}_i$.

For example, let us suppose that a straight transect line, centered at the mth sample point, has orientation θ_m and length L, as before. Let us consider the protocol where partial intersections of a particle are accepted at one end of the transect line and ignored at the other. Then $\mathscr{A}_i$ is equivalent to the inclusion zone of the particle as shown in Fig. 2a, except that the shaded region on, say, the left side is omitted, and the shaded region of the right side is retained. An alternative protocol was used suggested by Valentine *et al.* (2001) for Monte Carlo LIS, i.e., accept a partial intersection if the intersecting end of transect line cuts at least half across the particle. Under either protocol, the area of $\mathscr{A}_i$ is $e_i = L\omega_i(\theta_m)$.

Total y in $\mathscr{A}$, τ_y, is estimated unbiasedly from the tally at the m-th point by

$$\widehat{\tau}_{y_m} = A \sum_{i \in \mathscr{S}_m} \frac{y_i}{e_i}. \tag{12}$$

The estimator of the density of y per unit land area is simply:

$$\widehat{\lambda}_m = \frac{\widehat{\tau}_{y_m}}{A} = \sum_{i \in \mathscr{S}_m} \frac{y_i}{e_i}. \tag{13}$$

Under either stated protocol, this density estimator is equivalent to (4), the conditional estimator.

8. Discussion

Sampling designs which use fixed-area rectangular quadrats are commonplace in ecological investigations. The orientation of these quadrats is customarily fixed in advance of sampling. We are unaware of any estimator of ecological characteristics that is unconditional in the sense that it accounts for all possible orientations of the quadrats. This well-established practice, therefore, provides an implicit argument for conditional estimation in LIS, as presented here. The force of this analogy ought not be exaggerated, however: the unconditional estimator may be more practical and suitable in many situations.

The areas within $\mathscr{A}$ which determine the selection probabilities of particles when using ell-shaped transects differ in size and shape from what they are when using straight-line transects. Thus, to preserve design-unbiasedness, either conditionally or unconditionally on transect orientation, a modification to Kaiser's estimators is required. This apparently has not been recognized heretofore, as the applications using ell-shaped transects that we have read appear to regard the ell-shaped transect simply as two straight-line transects. Awaiting further investigation is the modification needed when using triangular transects, as in Delisle *et al.* (1988) and Y-shaped transects used in the Forest Inventory and Analysis program of the U.S. Forest Service.

Finally, it strikes us that the statistical advantage of using segmented (ell-shaped or triangular) rather than straight-line transects is more apocryphal than objectively established. If this observation is correct, then another area for field research is suggested.

Acknowledgments

The authors wish to thank Jeffrey H. Gove and two anonymous referees whose careful review of the manuscript led to its substantial improvement.

Appendix A

Let $\oslash_{im}$ signify the joint event that the segment of the ell-shaped transect in the θ_m direction intersects P_i and the segment of the ell-shaped transect in the $\theta_m^{\perp}$ direction does not. Similarly, let $\oslash_{im}^{\perp}$ signify that the segment of the ell-shaped transect in the $\theta_m^{\perp}$ direction intersects P_i and the segment of the ell-shaped transect in the θ_m direction does not. For the situation where only one segment of an ell-shaped transect intersects P_i, one gets the result that

$$\begin{aligned}
E[t_{im}q_i(\theta_m)|\theta_m, t_{im}=1] &= E[I_i(\theta_m)|\theta_m, \oslash_{im}]\mathrm{Prob}(\oslash_{im}|\theta_m) \\
&\quad + E[I_i(\theta_m^{\perp})|\theta_m, \oslash_{im}^{\perp}]\mathrm{Prob}(\oslash_{im}^{\perp}|\theta_m) \\
&= \left[\frac{a_i}{w_i(\theta_m)}\left(\frac{Lw_i(\theta_m)}{2} - \psi_i(\theta_m, \theta_m^{\perp})\right)\right. \\
&\quad \left. + \frac{a_i}{w_i(\theta_m^{\perp})}\left(\frac{Lw_i(\theta_m^{\perp})}{2} - \psi_i(\theta_m, \theta_m^{\perp})\right)\right] / A \\
&= \frac{La_i}{A} - \frac{\psi_i(\theta_m, \theta_m^{\perp})a_i}{2A}\left(\frac{1}{w_i(\theta_m)} + \frac{1}{w_i(\theta_m^{\perp})}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
E[t_{im}q_i(\theta_m)|\theta_m] &= E[t_{im}q_i(\theta_m)|\theta_m, t_{im}=1] + E[t_{im}q_i(\theta_m)|\theta_m, t_{im}=2] \\
&= E[I_i(\theta_m)|\theta_m, t_{im}=1]\mathrm{Prob}(t_{im}=1|\theta_m) \\
&\quad + E\left[\frac{I_i(\theta_m)+I_i(\theta_m^{\perp})}{2}\middle|\theta_m, t_{im}=2\right]\mathrm{Prob}(t_{im}=2|\theta_m) \\
&= \frac{La_i}{A} - \frac{\psi_i(\theta_m,\theta_m^{\perp})a_i}{2A}\left(\frac{1}{w_i(\theta_m)}+\frac{1}{w_i(\theta_m^{\perp})}\right) \\
&\quad + \frac{a_i/w_i(\theta_m)+a_i/w_i(\theta_m^{\perp})}{2}\left(\frac{\psi_i(\theta_m,\theta_m^{\perp})}{A}\right) \\
&= \frac{La_i}{A} - \frac{\psi_i(\theta_m,\theta_m^{\perp})a_i}{2A}\left(\frac{1}{w_i(\theta_m)}+\frac{1}{w_i(\theta_m^{\perp})}\right) \\
&\quad + \frac{\psi_i(\theta_m,\theta_m^{\perp})a_i}{2A}\left(\frac{1}{w_i(\theta_m)}+\frac{1}{w_i(\theta_m^{\perp})}\right) \\
&= \frac{La_i}{A}.
\end{aligned}$$

Appendix B

Symbol table

a_i	the cover of P_i projected onto the region $\mathscr{A}$
A	the area of $\mathscr{A}$
b_i	the length of the boundary of P_i projected onto $\mathscr{A}$
c_i	the circumference of the convex hull enclosing P_i
$I_i(\theta_m)$	the length of intersection with P_i of the line in $\mathscr{A}$ containing the line transect
l_i	the length of a needle-shaped particle
L	transect length; in the case of an ell-shaped transect, each segment has length $L/2$
M	the number of transects in the sample
N	the number of particles in the population
P_i	the i-th particle in the population $\mathscr{P}$
$\mathscr{P}$	the population of particles for which an estimate of λ or μ is sought
$q_i(\theta_m)$	the auxiliary variate measured on P_i, whose value may depend on the orientation of the transect
t_{im}	a variate indicating whether P_i is intersected by the m-th transect, and for ell-shaped transects, an indicator of whether P_i is intersected by one or both segments
θ_m	the orientation angle of the m-th transect; for ell-shaped transects, θ_m is the direction of that segment of the m-th transect that serves to establish its orientation
$\theta_m^{\perp}$	the angle perpendicular to the orientation of the m-th transect
$w_i(\theta_m)$	the maximal width of P_i perpendicular to θ_m
$w_i(\theta_m^{\perp})$	the maximal width of P_i perpendicular to $\theta_m^{\perp}$
τ	the population total amount of the attribute y
λ	the density of y on $\mathscr{A} = \tau/A$

$\widehat{\lambda}_m^c$	the conditional (on line orientation) estimator of λ from the m-th transect
$\widehat{\lambda}^c$	the conditional estimator of λ computed as the average $\widehat{\lambda}_m^c$
$\widehat{\lambda}_m^u$	the conditional (on line orientation) estimator of λ from the m-th transect
$\widehat{\lambda}^u$	the conditional estimator of λ computed as the average $\widehat{\lambda}_m^c$
μ	the average value of y per population element $= \tau/N$
$\widehat{\mu}_m^c$	the conditional (on line orientation) estimator of μ from the m-th transect
$\widehat{\mu}^c$	the conditional estimator of μ computed as the average $\widehat{\lambda}_m^c$
$\widehat{\mu}_m^u$	the conditional (on line orientation) estimator of μ from the m-th transect
$\widehat{\mu}^u$	the conditional estimator of μ computed as the average $\widehat{\mu}_m^c$

References

Battles, J.J., Dushoff, J.G., and Fahey, T.J. (1996) Line intersect sampling of forest canopy gaps. *Forest Science*, **42**, 131–8.

Beckmann, P. (1971) *A History of π*, St. Martin's Press, New York.

Buck, R.C. (1965) *Advanced Calculus*, McGraw-Hill Book Company, New York, p. 527.

Canfield, R.H. (1941) Application of the line interception method in sampling range vegetation. *Journal of Forestry*, **39**, 34–40.

Cochran, W.G. (1977) *Sampling Techniques*, Wiley, New York.

Delisle, P.D., Woodard, M.P., Titus, S.J., and Johnson, A.F. (1988) Sample size and variability of fuel weight estimates in natural stands of lodgepole pine. *Canadian Journal of Forest Research*, **18**, 649–52.

deVries, P.G. and Van Eijnsbergen, A.C. (1973) *Line Intersect Sampling Over Populations of Arbitrarily Shaped Elements*, Mededelingen Lanbrouwhogeschool, Wageningen, The Netherlands, pp. 73–19.

deVries, P.G. (1979) Line intersect sampling: statistical theory, applications, and suggestions for extended use in ecological inventory. In *Sampling Biological Populations*, R.M. Cormack, G.P. Patil, and D.S. Robson (eds), International Co-operative Publishing House, Fairland, Maryland, pp. 1–70.

Dorazio, R.M. (1999) Design-based and model-based inference in surveys of freshwater mollusks. *Journal of the North American Benthological Society*, **18**, 118–31.

Eberhardt, L.L. (1978) Transect methods for population studies. *Journal of Wildlife Management*, **42**, 1–31.

Gregoire, T.G., Valentine, H.T., and Furnival, G.M. (1993) Estimation of bole surface area and bark volume with Monte Carlo methods. *Biometrics*, **49**, 653–60.

Gregoire, T.G. and Monkevich, N.S. (1994) The reflection method of line intercept sampling to eliminate boundary bias. *Environmental and Ecological Statistics*, 1, 219–26.

Gregoire, T.G. (1998) Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research*, **28**, 1429–47.

Hazard, J.W. and Pickford, S.G. (1986) Simulation studies on line intersect sampling of forest residue, part II. *Forest Science*, **32**, 447–70.

Hedayat, A.S. and Sinha, B.K. (1991) *Design and Inference in Finite Population Sampling*, Wiley, New York.

Kaiser, L. (1983) Unbiased estimation in line-interception sampling. *Biometrics*, **39**, 965–76.

Lucas, H.A. and Seber, G.A.F. (1977) Estimating coverage and particle density with the line intercept method. *Biometrika*, **64**, 618–22.

Marshall, P.L., Davis, G., and LeMay, V.M. (2000) Using line intersect sampling for coarse woody debris. Tecnical Report TR-003, Research Section, Vancouver Forest Region, British Columbia Ministry of Forests.

Matérn, B. (1956) On the geometry of the cross-section of a stem. *Meddelanden Från Statens Skogforskningsinstitut*, Band 46 Nr (11), pp. 28.
Matérn, B. (1964) A method of estimating the total length of roads by means of a line survey. *Studia Forestalia Suecica*, **185**, 68–70.
McIntyre, G.A. (1953) Estimation of plant density using line transects. *Journal of Ecology*, **41**, 319–30.
Pielou, E.C. (1975) Line intersect sampling. In *Encyclopedia of Statistical Sciences*, S. Kotz and N.L. Johnson (eds.), **5**, Wiley, New York, pp. 70–74.
Ringvall, A. and G. Ståhl. (1999) Field aspects of line intersect sampling for assessing coarse woody debris. *Forest Ecology and Management*, **119**, 163–70.
Runkle, J. (1985) Comparison of methods for determining fraction of land area in treefall gaps. *Forest Science*, **31**, 15–19.
Runkle, J. (1992) Guidelines and sample protocol for sampling forest gaps. U.S.D.A. Forest Service PNW–GTR–283. Portland, Oregon: U.S.D.A. Forest Service, Pacific Northwest Research Station, p. 44.
Särndal, C-E., Swensson, B., and Wretman, J. (1992) *Model Assisted Survey Sampling*, Springer-Verlag, New York.
Valentine, H.T., Gove, J.H., and Gregoire, T.G. (2001) Monte Carlo approaches to sampling forested tracts with lines or points. *Canadian Journal of Forest Research*, **31**, 1410–24.
Warren, W.G. and Olsen, P.F. (1964) A line intersect technique for assessing logging waste. *Forest Science*, **10**, 267–76.

Biographical sketches

Timothy G. Gregoire is on the faculty of the School of Forestry and Environmental Studies at Yale University, where he is the J.P. Weyerhaeuser, Jr. Professor of Forest Management. His research interests include the development of sampling strategies and estimators for forestry and ecological resources, as well as the modeling of longitudinal and spatially correlated data. He is a recipient of the Distinguished Achievement Medal awarded by the Statistics and Environment Section of the American Statistical Association, and he is an elected Fellow of the ASA.

Harry Valentine is a research scientist with the USDA Forest Service, Northeastern Research Station, in Durham, New Hampshire. He was educated at Rutgers and Yale, where he studied forestry, mensuration, and statistical ecology. The first half of his career was devoted to the analysis and mathematical modeling of forest insect and disease problems; the second half has been devoted, in part, to the advancement of specialized sampling methods for estimating attributes of forest ecosystems. He also has developed and analyzed detailed mathematical models of carbon allocation in individual trees and whole forest stands.