

Design unbiased estimation in line intersect sampling using segmented transects

DAVID L.R. AFFLECK,^{1*} TIMOTHY G. GREGOIRE¹
and HARRY T. VALENTINE²

¹*School of Forestry and Environmental Studies, Yale University, New Haven, CT, 06511*

²*USDA Forest Service, Northeastern Research Station, Durham, NH, 03824*

Received February 2004; Revised August 2004

In many applications of line intersect sampling, transects consist of multiple, connected segments in a prescribed configuration. The relationship between the transect configuration and the selection probability of a population element is illustrated and a consistent sampling protocol, applicable to populations composed of arbitrarily shaped elements, is proposed. It is shown that this protocol obviates the arbitrary practice of treating multiple intersections of a single particle as independent probabilistic events and preserves the design-unbiasedness of Kaiser's (1983, *Biometrics* 39, 965–976) conditional and unconditional estimators, suitably generalized to segmented transect designs. The relative efficiency and utility of segmented transect designs are also discussed from a fixed population perspective.

Keywords: coarse woody debris, design-unbiasedness, multiple intersections, transect configuration

1352-8505 © 2005  Springer Science + Business Media, Inc.

1. Introduction

For discretely distributed populations, line intersect sampling (LIS) is a form of unequal probability sampling in which population elements crossed by a line transect are selected into the sample. Since its introduction to ecologists by Canfield (1941), LIS has been extensively applied in scientific and management surveys of vegetative cover and abundance, logging residue, and forest fuel loading, as well as mapping applications. The recent and widespread suspicion of a decline in biodiversity coupled with a need to assess carbon stocks driven by the Kyoto Protocol have contributed to an unprecedented effort to quantify and monitor coarse woody debris

*Corresponding author.

1352-8505 © 2005  Springer Science + Business Media, Inc.

(CWD) and downed woody material (DWM). LIS and variants thereof have been used extensively for this purpose. In many of these applications of LIS, the transects have consisted of multiple, connected segments with a fixed angular relation between segments. Examples include the Vegetation Resources Inventory of British Columbia, Canada, in which L-shaped transects are implemented; the Forest Inventory and Analysis program of the U.S. Forest Service and the National Forest Inventory of Switzerland, which utilize Y-shaped transects; the Canadian National Forest Inventory, which prescribes + -shaped transects; and the National Inventory of Landscapes in Sweden program, which has established square transects. In related applications, triangular transects have found use in assessments of forest fuels (Delisle *et al.*, 1988).

Our concern in this article is design-unbiased estimation when using transects with $K > 1$ connected segments, such as those described above. Specifically we focus on the necessary field measurements that must be procured in order to ensure design-unbiased estimation when LIS is applied to discretely distributed populations. Our concern is motivated by the apparently widespread failure to appreciate that estimators that are unbiased when using straight line ($K = 1$) transects are not unbiased, in general, when using multi-segmented transects. That is, the practice of treating LIS with transects of $K > 1$ segments as equivalent to LIS with K single segment transects generally will result in biased estimation of population parameters. The practical importance of this matter is underscored by the adoption of multiply segmented transects by many public agencies, leading to many thousands of samples from such transects being collected on an ongoing annual basis. It seems prudent, therefore, to put estimation on the basis of these samples on a sound statistical footing.

The results reported here extend those of Gregoire and Valentine (2003) by considering more intricately shaped transects than the $K = 2$, L-shaped transects of concern to those authors. Moreover, we deal more comprehensively with the issues presented when a transect intersects a population element in multiple places. In turn, this is related to the proper measurement of non-convex-shaped elements that are only partially intersected by the transect. Finally, we establish in this article a distinction between radial and polygonal transects in LIS – a distinction which has quite consequential implications both for field measurement and efforts to deal with edge effect.

Following Kaiser (1983) we take a design-based approach to inference following LIS, as articulated *inter alia* by Särndal *et al.* (1992) and Gregoire (1982). Briefly stated, the reference distribution from which the statistical properties of estimators are derived under the design-based paradigm comprises the universe of estimates induced by the random location, and possibly orientation, of transects permissible under the sampling design. The population of discrete elements, called particles henceforth, is regarded as fixed, in the sense that the spatial pattern with which particles are dispersed over the landscape (or other medium being sampled) is considered to be non-stochastic. In particular, an assumption of complete spatial randomness of particles is neither necessary nor relevant for purposes of design-based inference. Attributes of particles, such as their presence or size, likewise are considered to be fixed, and inference is utterly free of any assumption that attribute values are the realization of a stochastic process. Randomness enters into the

inferential procedure solely by the probabilistic inclusion of particles into the sample as governed by the sampling design. As we mention later, the literature on LIS historically has been ambiguous in regard to the presumed inferential basis, the early treatments by Canfield (1941) and Eberhardt (1978) notwithstanding.

2. Notation

Define the population of interest as the collection $\mathcal{P} = \{P_1, P_2, \dots, P_N\}$ of discrete particles distributed over the horizontal region $\mathcal{A}$. Each particle must be connected in the sense that any two points in P_i can be joined by a path contained entirely in P_i (see Stewart, 1995, p. 910). Essentially, this means that disjoint objects must be considered distinct particles. Apart from this requirement, however, particles can have any shape. In particular, particles can be forked or contain interior hollow regions. Nothing about the spatial distribution of $\mathcal{P}$ on $\mathcal{A}$ is assumed: particles may be spread and oriented in any manner over $\mathcal{A}$ and may overlap in their projection onto $\mathcal{A}$.

Let y_i be a fixed characteristic of P_i that in no way depends on whether or how that particle is intersected by a transect. The characteristic y_i can be a binary indicator of a particular attribute of P_i or simply of its presence on $\mathcal{A}$, but more generally y_i will be taken to reflect a real-valued, measurable attribute of P_i (e.g., particle volume). Interest typically lies in estimating the aggregate quantity

$$\tau = \sum_{i \in \mathcal{P}} y_i,$$

or some function thereof. Other descriptive parameters of interest may include the density of $\mathcal{P}$ on $\mathcal{A}$:

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

where A is the area of $\mathcal{A}$; and the mean y over $\mathcal{P}$:

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

In the following, we consider estimators of τ and λ . Estimation of μ is discussed by Gregoire and Valentine (2003).

The sample design consists of M replicated transects on $\mathcal{A}$, each of total length L . Individual transects are made up of K connected directed line segments of equal length L/K . That is, the k th segment of the m th transect has a defined starting point, referenced by the coordinate pair (x_{mk}, z_{mk}) . The segments of a transect can be arranged into radial or polygonal configurations. Radial transects consist of one or more segments directed outwards from a common vertex, as in the case of straight-line, L-shaped, or Y-shaped transects (Fig. 1). Polygonal transects are made up of three or more segments forming a closed figure, such as a triangle or square. More complex transect configurations, consisting of both radial and polygonal features, have also been employed in field applications. For example, one transect configuration prescribed by the U.S. Forest Service for forest fuels inventory consists of six segments

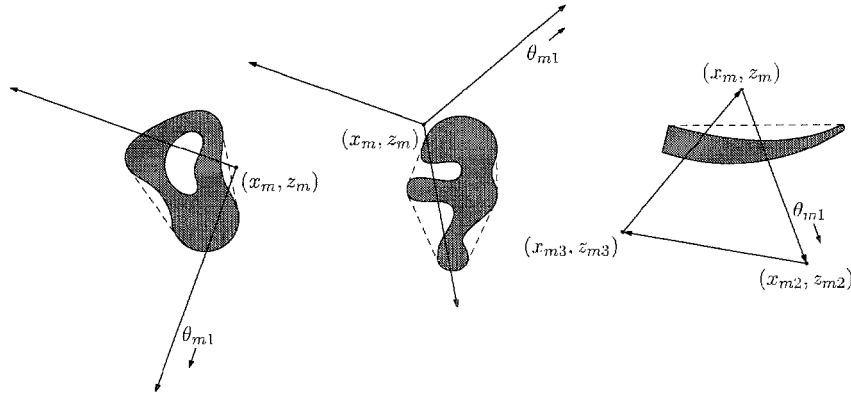


Figure 1. Example intersections of non-convex particle by segmented transects. The convex hulls of the particles are indicated by dashed lines.

in a hexagonal arrangement with a seventh segment linking the hexagon to its centroid. The sampling protocol and estimators described below apply to all such transect configurations, but for expository purposes we consider only designs under which all M transects share a common radial or polygonal geometry.

The location of the m th transect is referenced by the coordinate pair (x_m, z_m) on the horizontal plane of $\mathcal{A}$. On a radial transect this point location refers to the transect vertex, and thus corresponds to the starting point of each segment (i.e., $(x_m, z_m) = (x_{mk}, z_{mk})$ for all K segments). For specificity, we adopt the protocol that the segments of a polygonal transect are sequentially established in a counter-clockwise progression. The location of a polygonal transect can be associated with the starting point of the initial segment (i.e., $(x_m, z_m) = (x_{m1}, z_{m1})$). Transect locations are presumed to be selected independently and uniformly at random in $\mathcal{A}$.

Relative to some reference direction, say $\theta = 0$, let $\theta_m = [\theta_{m1}, \theta_{m2}, \dots, \theta_{mK}]$ denote the vector of orientations of the K segments of the m th transect. The orientation of each segment ranges over 2π radians, with $\theta_{m1} \in [0, 2\pi)$. For the radial and polygonal transects considered below, θ_m is completely determined once θ_{m1} is selected. For example, given θ_{m1} the orientations of the three segments of a Y-shaped or triangular transect are $\theta_m = [\theta_{m1}, \theta_{m1} + \frac{2\pi}{3}, \theta_{m1} + \frac{4\pi}{3}]$. The orientation of the reference segment (i.e., θ_{m1}) can be selected at random from a defined distribution or chosen arbitrarily in advance of sampling.

As noted by Kaiser (1983), estimation of population parameters can proceed conditionally on the transect orientations used or it can proceed unconditionally by jointly accounting for randomness in the transect locations as well as their orientations. Conditional on the orientations of the M transects, a design unbiased estimator of τ is

$$\hat{\tau}^c = \frac{1}{M} \sum_{m=1}^M \hat{\tau}_m^c,$$

where $\hat{\tau}_m^c$ is a conditionally unbiased estimator of τ based on the m th transect alone (i.e., $\mathbb{E}[\hat{\tau}_m^c | \theta_m] = \tau$). Similarly, but appropriate only for designs wherein θ_{m1} is selected uniformly at random from the interval $[0, 2\pi)$, an unconditionally design unbiased estimator of τ is

$$\hat{\tau}^u = \frac{1}{M} \sum_{m=1}^M \hat{\tau}_m^u,$$

where $\hat{\tau}_m^u$ is an unbiased estimator of τ based solely on the m th transect. Conditional and unconditional estimators of λ (and μ) based on replicated transects are derived in a similar fashion. Formulae for $\hat{\tau}_m^c$ and $\hat{\tau}_m^u$ are presented in subsequent sections.

From a design-oriented approach to inference, the variances of the above estimators are functions of the fixed, and presumably unknown, distribution of $\mathcal{P}$ on $\mathcal{A}$ through the joint inclusion probabilities of distinct pairs of particles. When transects are independently replicated over $\mathcal{A}$, $\hat{\tau}^c$ is asymptotically normally distributed (Barabesi and Fattorini, 1998) with variance that is design-unbiasedly estimated by

$$\widehat{\text{var}}(\hat{\tau}^c) = \frac{1}{M(M-1)} \sum_{m=1}^M (\hat{\tau}_m^c - \hat{\tau}^c)^2. \quad (1)$$

Similarly,

$$\widehat{\text{var}}(\hat{\tau}^u) = \frac{1}{M(M-1)} \sum_{m=1}^M (\hat{\tau}_m^u - \hat{\tau}^u)^2 \quad (2)$$

provides a design-unbiased estimator of the variance of $\hat{\tau}^u$. Again, variance estimators for conditional and unconditional estimators of λ are obtained by direct substitution. For large M , $(1-\alpha)100\%$ confidence intervals for τ can be constructed by

$$\hat{\tau}^c \pm z_{\frac{\alpha}{2}} \sqrt{\widehat{\text{var}}(\hat{\tau}^c)}$$

and

$$\hat{\tau}^u \pm z_{\frac{\alpha}{2}} \sqrt{\widehat{\text{var}}(\hat{\tau}^u)},$$

where $z_{\frac{\alpha}{2}}$ is the $\frac{\alpha}{2}$ quantile of the standard normal distribution. For moderate M , confidence intervals can be constructed using the quantiles of Student's t distribution or from the percentiles of an empirical distribution of bootstrap estimates for τ (or λ).

In the following, we assume that edge effects are accounted for such that they do not impart a design bias into the estimators of population descriptive parameters. For designs employing radial transects, edge conditions can be handled by adaptations of the methods described in Kaiser (1983) or Gregoire and Monkevich (1994). Mitigation of edge effects under polygonal transect designs is more complicated as one or more entire segments can be located exterior to $\mathcal{A}$. A more comprehensive consideration of methods to deal with edge effects when using segmented transects in LIS is beyond the scope of this manuscript. Indeed, to simplify further discussion, we assert that the sampling protocol presented below applies strictly to particles in the interior of $\mathcal{A}$.

3. Intersection

A particle is selected into the sample if its projection onto $\mathcal{A}$ is intersected by a transect. When sampling populations composed of convex particles, sampling protocols follow from straightforward and intuitive definitions of complete and partial intersections. Let B_i denote the boundary of the vertical projection of P_i on $\mathcal{A}$. If the vertical projection of P_i is convex, then a segment can intersect B_i at most twice. McIntyre (1953), who considered populations composed strictly of convex particles, thus defined a partial intersection as one in which a segment crosses B_i once, and a complete intersection as one in which B_i is crossed twice. Kaiser (1983) adopted these definitions and applied them to populations composed of arbitrarily shaped particles, as have subsequent authors (e.g., Gregoire and Valentine, 2003). Yet these definitions are insufficient when applied to connected particles of arbitrary shape, because a segment can span multiple lobes, or cavities, of a non-convex particle (Fig. 1). When a segment can cross the particle boundary more than twice, more exacting definitions of complete and partial intersection are required.

We consider P_i to be completely intersected by a segment only if the segment intersects B_i as many times as a coincident line of infinite length. In Fig. 2, the boundary of the forked particle may be crossed by a single segment as many as four times. However, only in Fig. 2a do the segments completely intersect P_i : no linear extension of these segments would result in another intersection of B_i . More formally, an intersection is complete if the number of times B_i is cut by a transect segment is equal to the total projection (see Kendall and Moran, 1963, p. 59) of B_i onto a line orthogonal to the segment over the point (x_{mk}, z_{mk}) . We consider any other type of intersection as partial. Furthermore, we differentiate between two types of partial intersections on the basis of the location of the starting point of the segment relative to B_i . Partial intersections that involve the starting point of a segment occur if (x_{mk}, z_{mk}) is located between the intersections of a coincident line of infinite length and B_i (e.g., Fig. 2b). That is, if an extension of the segment in the direction $\theta_{mk} + \pi$ (i.e., in the opposite direction) intersects B_i . As shown in Fig. 2b, this does necessarily imply that (x_{mk}, z_{mk}) is located within the vertical projection of P_i . Partial intersections that do not involve the starting point occur when (x_{mk}, z_{mk}) is located anterior to B_i with respect to the direction of the segment (Fig. 2c).

We propose a sampling protocol for any transect configuration and for populations of arbitrarily shaped particles based on the above definitions. Specifically, a particle is selected into the sample by the m th transect if it is completely or, with one restriction, partially intersected by at least one segment. Selection by partial intersection is restricted to the case where the intersection does not involve the starting point of the segment involved. While this protocol may appear abstract in definition, it is straightforward in implementation: P_i is selected by the k th segment of the m th transect if the segment crosses a boundary of P_i and if no portion of P_i occurs behind the starting point (i.e., in the direction $\theta_{mk} + \pi$). Under the protocol, any partial intersections that involve the starting point of a segment (e.g., Fig. 2b), including those that could occur if an entire segment were to fall inside a particle, are ignored.

This sampling protocol can be effectively summarized by consideration of a particle's inclusion field, i.e., the locus of points in $\mathcal{A}$ in which the origin of the transect,

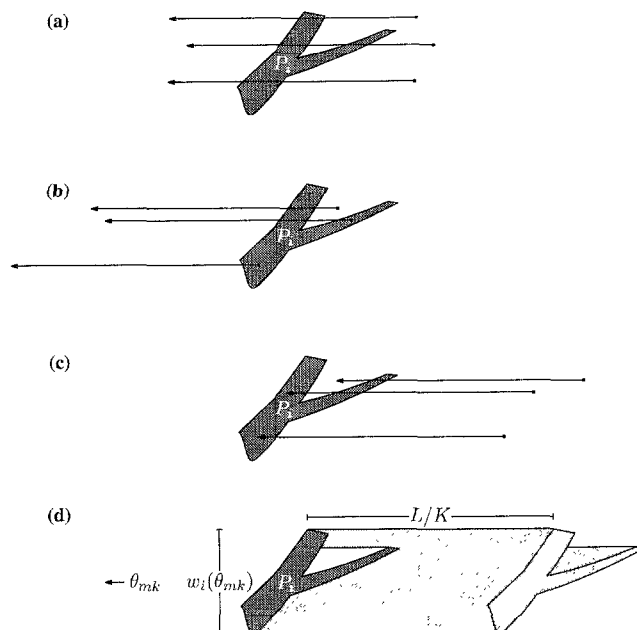


Figure 2. (a) Complete intersection; (b) partial intersection involving the starting point; (c) partial intersection not involving the starting point; (d) the inclusion field of the particle P_i for a single-segment transect of fixed orientation.

(x_m, z_m) , must be located in order to intersect P_i and thereby select it into the sample. As is made evident in the next section, the size and shape of the inclusion field has a crucial bearing on the probability with which P_i will be intersected by one or more segments, either completely or partially under the protocol. There is never a need to demarcate the inclusion field of a particle when implementing LIS, however.

For a single-segment transect with fixed orientation θ_{mk} , the forked particle considered above will be selected into the sample only if (x_m, z_m) falls in the lightly shaded region of Fig. 2d. For transects with $K > 1$, however, the shape and area of the inclusion field will vary with the shape of the particle and the configuration of the transect, as illustrated in Figs. 3 and 4.

Applying the sampling protocol to a radial transect design, all partial intersections involving the transect vertex are ignored. However, each segment of a radial transect generates a subset of the inclusion field, and these subsets can overlap. The generation of an inclusion field for a Y-shaped transect with fixed orientation θ_{m1} is shown in Fig. 3: P_i will be selected once if (x_m, z_m) falls in any of the lightly shaded areas and twice if (x_m, z_m) falls in the darker shaded areas. With other types of radial transects, the inclusion field may contain regions where the corresponding particle is selected up to K times. For example, van Wagner (1968) suggested the use of a radial transect with $K=3$ and $\theta_m = [\theta_{m1}, \theta_{m1} + \frac{\pi}{3}, \theta_{m1} + \frac{2\pi}{3}]$; a particle close to the vertex of

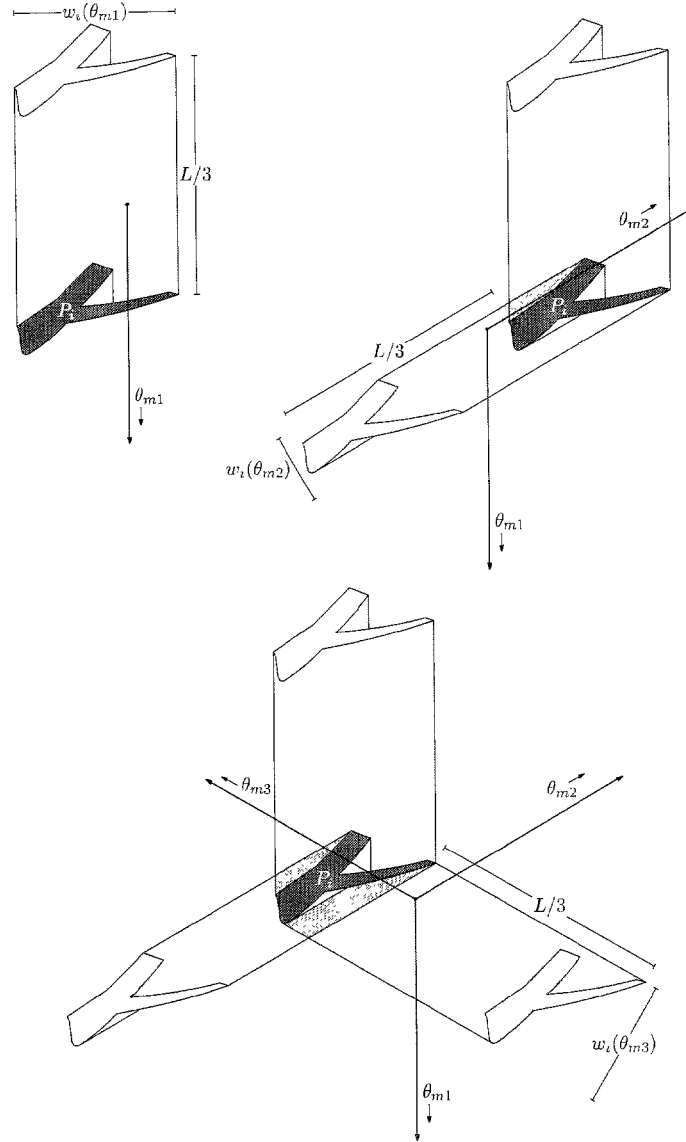


Figure 3. Generation of an inclusion field for a Y-shaped transect; the particle P_i is selected into the sample once (twice) if the transect is located in the light (dark) shaded regions.

this transect could be spanned by all three segments. Under the protocol we have established, a particle cannot be selected by a radial transect if (x_m, z_m) falls within the vertical projection of P_i on $\mathcal{A}$.

Under a polygonal transect design, provided the transect does not fall entirely within the vertical projection of P_i , any type of intersection will lead to selection because a partial intersection involving the starting point of one segment is also a partial intersection not involving the starting point of the previous segment. This is illustrated for a triangular transect of fixed orientation in Fig. 4, but applies to any type of polygonal transect. Note that for reasonably large L/K (relative to the size of particles under consideration), the configuration of a polygonal transect will generally preclude the selection of a particle by more than two segments, though up to K selections are possible. In practice, visual assessment of the shape of the particle will

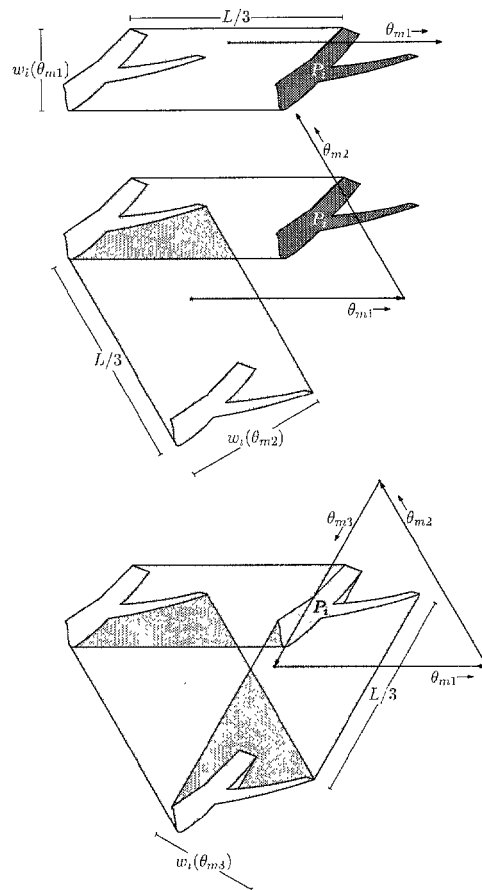


Figure 4. Generation of an inclusion field for a triangular transect; the particle P_i is selected into the sample once (twice) if the transect is located in the light (dark) shaded regions.

generally suffice to determine whether an intersection meets the standards established by the protocol and, if so, for which segment(s) of the transect.

4. Estimation

The sampling protocol developed in the previous section eliminates any need to subdivide connected particles, thereby permitting unambiguous identification of particles and their respective inclusion fields under any segmented transect design. This eliminates the rather arbitrary practice (see Waddell, 2002) of treating multiple intersections of a single particle as being equivalent to single intersections of multiple particles.

Let $t_{imk} = 1$ if P_i is selected into the sample by the k th segment of the m th transect and let $t_{imk} = 0$ otherwise; also, let $t_{im\cdot} = \sum_k t_{imk}$. It is evident from Fig. 2d that

$$\mathbb{P}(t_{imk} = 1 \mid \theta_{mk}) = \frac{Lw_i(\theta_{mk})}{AK},$$

where $w_i(\theta_{mk})$ is the width of P_i perpendicular to the k th segment, i.e., the maximum distance between lines parallel to the segment and tangent to B_i . Also, regardless of the transect configuration,

$$\mathbb{E}[t_{im\cdot} \mid \theta_m] = \sum_{k=1}^K \mathbb{E}[t_{imk} \mid \theta_{mk}] = \frac{L\bar{w}_i(\theta_m)}{A}, \quad (3)$$

where $\bar{w}_i(\theta_m) = \sum_k w_i(\theta_{mk})/K$. The simplicity of these results derive directly from the restriction on partial intersections established in the sampling protocol.

A design-unbiased estimator of τ conditional on the transect orientation can be obtained on the basis of (3). Specifically, the conditional estimator given by Gregoire and Valentine (2003) can be extended to any segmented transect design as

$$\hat{\tau}_m^c = \sum_{i \in \mathcal{P}} \frac{t_{im\cdot} y_i}{\mathbb{E}[t_{im\cdot} \mid \theta_m]} = \frac{A}{L} \sum_{i \in L_m} \frac{t_{im\cdot} y_i}{\bar{w}_i(\theta_m)}, \quad (4)$$

where the second summation is taken over all particles selected by the m th transect. This estimator is design-unbiased for a given transect orientation, provided θ_m is selected in advance of sampling or in some objective manner. A conditionally unbiased estimator of λ is obtained directly from (4) as $\hat{\lambda}_m^c = \hat{\tau}_m^c/A$. Note that for $K=1$, the linear homogeneous estimator (see Cassel *et al.*, 1977, p. 22) given by (4) reduces to the more familiar Horvitz–Thompson (HT) estimator. When $K>1$, however, the practical difficulties involved in measuring the area of an inclusion field will prevent a determination of the inclusion probability of P_i , thereby precluding the use of the HT estimator.

Alternatives to $\hat{\tau}_m^c$ that do not condition on the transect orientation can be considered when θ_{m1} is selected at random from a defined distribution. In particular, if θ_{mk} is selected uniformly at random from the interval $[0, 2\pi)$ then $\mathbb{E}[w_i(\theta_{mk})] = (\mathbb{E})[\bar{w}_i(\theta_m)] = c_i/\pi$, where c_i is the perimeter of the convex hull of the vertical projection of P_i (Kendall and Moran, 1963, p. 57). With this result, we can

also extend Gregoire and Valentine's (2003) unconditional estimator to any segmented transect design as

$$\hat{\tau}_m^u = \sum_{i \in \mathcal{D}} \frac{t_{im} \cdot y_i}{\mathbb{E}[t_{im}]} = \frac{A\pi}{L} \sum_{i \in L_m} \frac{t_{im} \cdot y_i}{c_i}. \quad (5)$$

The corresponding unconditional estimator of λ is given by $\hat{\lambda}_m^u = \hat{\tau}_m^u / A$. Both $\hat{\tau}_m^u$ and $\hat{\lambda}_m^u$ are design-unbiased when $\theta_{m1} \sim U[0, 2\pi)$, regardless of whether θ_{m1} is selected independently for each transect or whether a single orientation is randomly selected and used for all transects.

When transects have been randomly oriented, one can opt to use either the conditional or the unconditional estimator, and the choice can be based on the estimated variances from (1) and (2). However, it may be prudent to consider the practical implications of this choice prior to sampling. In particular, $\hat{\tau}_m^c$ requires measurements of the perpendicular width of P_i in every direction spanned by the transect regardless of which segments actually intersect P_i , whereas $\hat{\tau}_m^u$ requires only measurement of c_i . For particles that lie approximately flat on $\mathcal{A}$, such as fragments of CWD, it may be easier to obtain c_i (e.g., by measuring the length of rope wrapped taut about P_i) than $\bar{w}_i(\theta_m)$. On the other hand, as long as the number of segments is not too great, it may be faster and more accurate to estimate $\bar{w}_i(\theta_m)$ when the projection of P_i on $\mathcal{A}$ is not as easily delineated, as when particles consist of tree canopies (e.g., O'Brien, 1989), or canopy gaps (Runkle, 1992).

5. Estimation using an auxiliary variable

As described above, in certain applications one or other of the conditional and unconditional estimators given by (4) and (5), respectively, may reduce the mensurational effort required in the field. However, as shown by Kaiser (1983), even larger gains in practical efficiency can be obtained by the use of an auxiliary random variable, say $q_i(\theta_m)$. By judicious choice of $q_i(\theta_m)$ one can obviate the need to measure y_i as well as $\bar{w}_i(\theta_m)$ or c_i . This can be advantageous when the characteristic of interest is difficult to measure or its measurement is subject to considerable error. Generalized to segmented transect designs, Kaiser's conditional and unconditional estimators of τ are, respectively,

$$\hat{\tau}_m^c = \sum_{i \in \mathcal{D}} \frac{t_{im} \cdot q_i(\theta_m) y_i}{\mathbb{E}[t_{im} q_i(\theta_m) | \theta_m]}, \quad (6)$$

and

$$\hat{\tau}_m^u = \sum_{i \in \mathcal{D}} \frac{t_{im} \cdot q_i(\theta_m) y_i}{\mathbb{E}[t_{im} q_i(\theta_m)]}, \quad (7)$$

where

$$q_i(\theta_m) = \sum_k \frac{t_{imk} v_i(\theta_{mk})}{t_{im}}, \quad (8)$$

and $v_i(\theta_{mk})$ is a random variable that depends on the manner in which P_i is intersected by the k th segment of the m th transect. The practical value of these estimators is best illustrated through an example.

Example. Let P_i represent a particle of CWD with volume y_i and suppose the parameter of interest is the total particle volume on the tract $\mathcal{A}$. Denote by $v_i(\theta_{mk})$ the cross-sectional area of P_i in the vertical plane containing the k th segment of the m th transect (Fig. 5). Practical methods for obtaining cross-sectional areas of CWD and other solids are described by Valentine *et al.* (2001). With $v_i(\theta_{mk})$ in (8) so defined, it can be shown (see Appendix A) that

$$\mathbb{E}[t_{im} q_i(\theta_m) \mid \theta_m] = \mathbb{E}[t_{im} q_i(\theta_m)] = \frac{L y_i}{A}. \quad (9)$$

Substituting this result into (6) and (7) yields

$$\hat{\tau}_m^c = \hat{\tau}_m^u = \frac{A}{L} \sum_{i \in L_m} \sum_{k=1}^K t_{imk} v_i(\theta_{mk}),$$

which is an estimator of the total volume on $\mathcal{A}$ that requires only the measurement of the cross-sectional areas of the particles selected by each segment. Moreover, this estimator is design-unbiased regardless of whether transect orientations are selected arbitrarily or at random. The cross-sectional area of P_i must be measured in the vertical plane of each segment by which it is selected, but measurements of $\bar{w}_i(\theta_m)$ and c_i are not required.

The estimator derived in the above example can be viewed as a two-stage estimator of τ , with $v_i(\theta_{mk})$ being the volume of a random slice of P_i of infinitesimal width. In many instances a similar form for $v_i(\theta_{mk})$ will be suggested by a geometric conception of the problem (see examples in Kaiser, 1983; Gregoire and Valentine, 2003), which in turn may be facilitated by interpretation of LIS as an application of Monte Carlo integration (see Valentine *et al.*, 2001). In such cases, the variance of the estimator will reflect the variability in the characteristic of interest both among and within the individual particles on $\mathcal{A}$. Thus, it is possible that practical efficiencies gained through the use of an auxiliary variable may be offset by reduced precision. As noted above however, the use of the auxiliary variable can not only obviate the need to measure y_i but also the need to measure $\bar{w}_i(\theta_m)$ or c_i . Besides the ensuing savings in time, this could also eliminate potential

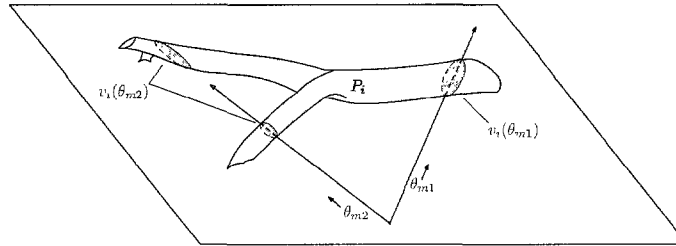


Figure 5. Intersection of a forked particle by two segments of a transect; $v_i(\theta_{mk})$ is the total cross-sectional area of P_i in the vertical plane containing the k th segment.

errors associated with the measurement of these dimensions of the vertical projection of P_i when P_i is a complex three-dimensional entity. Considerably more field study, similar to that of Ringvall and Ståhl (1999), is needed before the degradation of precision in estimation resulting from nearly unavoidable measurement error is properly appreciated.

6. Discussion

The practice of treating multiple intersections of a single connected particle as singular intersections of multiple particles (cf. DeVries, 1973; McRae *et al.*, 1979; Marshall *et al.*, 2000; Waddell, 2002) renders identifiability of particles impossible: a forked particle intersected once is considered a single particle, yet when intersected twice each time is considered a particle. The sampling protocol described in Section 3 avoids this difficulty altogether. By relying on connectedness as the defining characteristic of a population particle, as in Kaiser (1983), it unambiguously establishes how to treat multiple intersections of a single particle: if intersection occurs multiple times along one segment of a transect, the particle is selected at most once; if intersections occur along multiple segments, the particle can be selected by each segment.

The impetus to adopt transects with $K > 1$ segments originally arose as an effort to mitigate “orientation bias” (Warren and Olsen, 1964; van Wagner, 1968; DeVries, 1979). When basing inference on a presumption of particles whose locations are the realization of a Poisson process and whose orientations are $U[0, 2\pi]$ realizations, it was felt that transects with segments with prescribed but alternating orientations would reduce to an acceptable level the error that resulted from violations of this spatial model.

The advantages of a design-based inferential approach are twofold. It provides a more utilitarian basis for inference, inasmuch as biological phenomena rarely (ever?) exhibit complete spatial randomness. Also it permits inference to proceed conditionally upon transects that are oriented purposefully to take advantage of particles in $\mathcal{P}$ whose orientations generally are in the same direction. To judge from the published literature, those that have chosen to apply LIS to discretely distributed populations have often failed to appreciate these advantages: much of the rationale for multiply segmented transects has implicitly conveyed the hope that “if we don’t cross it in this direction, we might in the other direction.”

The irrelevance of particle orientation to unconditional estimation and of its utility to conditional estimation in the design-based framework begs the question of whether multiply segmented transects should be used. Surely their use could be justified if the precision with which τ can be estimated is greater with segmented transects than it is with straight line transects. This precision is determined by the distribution of $\mathcal{P}$ over $\mathcal{A}$ and the transect configuration through the inclusion probabilities of individual particles and the joint inclusion probabilities of distinct pairs of particles. For fixed L , multiply segmented transects make for a more local search around the sample point relative to straight line transects. As such, estimates of τ from segmented transects are likely to be more variable when particles exhibit a clumped spatial structure, as discussed by Marshall *et al.* (2000;

see also Delisle *et al.*, 1988). One reviewer of this manuscript maintained that segmented transect designs were very efficient for sampling populations composed of linear particles. Yet to the best of our knowledge, no analytical, empirical, or simulation study has established such a superiority over straight line transects for any type of discretely distributed population, and hence we assert that this issue remains as much an unanswered question as it was when first raised by Kaiser (1983).

It seems likely, also, that the burden of field measurements is increased with segmented transects, if for no other reason than the need to orient $K > 1$ transects of total length L rather than just a single transect of identical length. As discussed in Section 3, radial segments will also require greater attention to distinguish partial intersections than is required with straight line transects. Straight line transects may also be preferred in terms of the relative ease with which edge effects can be handled, though this is subject of ongoing research. When adopting more complex transect configurations, the potential for errors of implementation, as discussed by Ringvall and Ståhl (1999), should be given utmost consideration.

Our reservations concerning segmented transects in LIS notwithstanding, they are in widespread use. Design-unbiased estimation of population parameters in LIS with segmented transects can be ensured by adopting a sampling protocol that reflects the modified selection probabilities of particles under the design, as we have derived here.

Appendix A

Assertion (9) in the example of Section 5 follows from the fact that $v_i(\theta_{mk}) \equiv v_i(\theta_{mk}, l_{mk})$, where l_{mk} is the location of the sampling plane along the perpendicular projection of P_i and $l_{mk} | t_{imk} = 1 \sim U(0, w_i(\theta_{mk}))$. Therefore,

$$\begin{aligned}
 \mathbb{E}[t_{im} \cdot q_i(\theta_m) | \theta_m] &= \sum_k \mathbb{E}[t_{imk} v_i(\theta_{mk}) | \theta_{mk}] \\
 &= \sum_k \mathbb{E}[\mathbb{E}[t_{imk} v_i(\theta_{mk}) | t_{imk}, \theta_{mk}] | \theta_{mk}] \\
 &= \sum_k \mathbb{E}[\mathbb{E}[v_i(\theta_{mk}) | t_{imk} = 1, \theta_{mk}] \\
 &\quad \times \mathbb{P}(t_{imk} = 1 | \theta_{mk}) | \theta_{mk}] \\
 &= \sum_k \mathbb{E}\left[\frac{y_i}{w_i(\theta_{mk})} \frac{L w_i(\theta_{mk})}{KA} \middle| \theta_{mk}\right] \\
 &= \sum_k \frac{y_i L}{KA} \\
 &= \frac{L y_i}{A}.
 \end{aligned} \tag{10}$$

Since result (10) does not depend on θ_m , it is evident that

$$\mathbb{E}[t_{im} \cdot q_i(\theta_m)] = \mathbb{E}[t_{im} \cdot q_i(\theta_m) \mid \theta_m] = \frac{Ly_i}{A}.$$

Acknowledgments

This research was supported by funds from the USDA Forest Service, Northeastern Research Station, RWU-4104, through a cooperative agreement with the School of Forestry and Environmental Studies at Yale University. The authors also thank four anonymous reviewers for their insightful comments that improved the quality of the manuscript.

References

- Barabesi, L. and Fattorini, L. (1998) The use of replicated plot, line and point sampling for estimating species abundance and ecological diversity. *Environmental and Ecological Statistics*, **5**, 353–70.
- Canfield, R.H. (1941) Application of the line interception method in sampling range vegetation. *Journal of Forestry*, **39**, 388–94.
- Cassel, C.M., Särndal, C.E., and Wretman, J.H. (1977) *Foundations of Inference in Survey Sampling*, Wiley, New York.
- Delisle, G.P., Woodward, P.M., 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. (1973) A general theory on line intersect sampling with application to logging residue inventory. Mededelingen Landbouwhogeschool 73-11, Wageningen, The Netherlands.
- DeVries, P.G. (1979) Line intersect sampling statistical theory applications, and suggestions for extended use in ecological inventory, In *Sampling Biological Populations*, R. Cormack, G. Patil and D. Robson (eds.), International Co-operative Publishing House, Fairland, MD, pp. 1–70.
- Eberhardt, L.L. (1978) Transect methods for population studies. *Journal of Wildlife Management*, **42**, 1–31.
- Gregoire, T.G. (1982) The unbiasedness of the mirage correction procedure for boundary overlap. *Forest Science*, **28**, 504–8.
- 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. and Valentine, H.T. (2003) Line intersect sampling: Ell-shaped transects and multiple intersections. *Environmental and Ecological Statistics*, **10**, 263–79.
- Kaiser, L. (1983) Unbiased estimation in line-intercept sampling. *Biometrics*, **39**, 965–76.
- Kendall, M.G. and Moran, P.A.P. (1963) *Geometrical Probability*, Hafner, New York.
- Marshall, P.L., Davis G., and LeMay, V.M. (2000) Using line intersect sampling for coarse woody debris. Technical Report TR-003, Research Section, Vancouver Forest Region, B.C. Ministry of Forests.

- McIntyre, G.A. (1953) Estimation of plant density using line transects. *Journal of Ecology*, **41**, 320–30.
- McRae, D.J., Alexander, M.E., and Stocks, B.J. (1979) *Measurement and description of fuels and fire behavior on prescribed burns: a handbook*, Information Report O-X-287, Canadian Forestry Service, Great Lakes Forest Research Centre, Sault Ste. Marie, ON.
- O'Brien, R.A. (1989) Comparison of overstory canopy cover estimates on forest survey plots. Research Paper INT-417, USDA Forest Service, Intermountain Research Station, Ogden, UT.
- Ringvall, A. and Ståhl, G. (1999) Field aspects of line intersect sampling for assessing coarse woody debris. *Forest Ecology and Management*, **119**, 163–70.
- Runkle, J.R. (1992) Guidelines and sample protocol for sampling forest gaps. General Technical Report PNW-GTR-283, USDA Forest Service, Pacific Northwest Research Station, Portland, OR.
- Särndal, C.E., Swensson, B., and Wretman, J.H. (1992) *Model Assisted Survey Sampling*, Springer-Verlag, New York.
- Stewart, J. (1995) *Multivariable Calculus*, 3rd edn. Brooks/Cole, Pacific Grove, CA.
- 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.
- Van Wagner, C.E. (1968) The line intersect method in forest fuel sampling. *Forest Science*, **14**, 20–6.
- Waddell, K.L. (2002) Sampling coarse woody debris for multiple attributes in extensive resource inventories. *Ecological Indicators*, **1**, 139–153.
- Warren, W.G. and Olsen, P.F. (1964) A line intersect technique for assessing logging waste. *Forest Science*, **10**, 267–76.

Biographical sketches

David L.R. Affleck is a Ph.D. Student in the School of Forestry and Environmental Studies at Yale University. He holds a B.Sc. in forestry and a M.Sc. in forest biometrics from the University of British Columbia, Canada. Specific research interests include forest inventory and environmental sampling methods, and the development of statistical models for ecological data.

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 modelling 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 T. 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 modelling 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.