

A comprehensive comparison of perpendicular distance sampling methods for sampling downed coarse woody debris

Jeffrey H. Gove^{1*}, Mark J. Ducey², Harry T. Valentine¹ and Michael S. Williams³

¹USDA Forest Service, Northern Research Station, 271 Mast Road, Durham, NH 03824, USA; ²Department of Natural Resources and the Environment, James Hall, University of New Hampshire, Durham, NH 03824, USA; ³Mathematical Statistician, Risk Assessment Division, Office of Public Health Science, Food Safety Inspection Service, 2150 Centre Ave, Building D, Room 3207, Fort Collins, CO 80526, USA

*Corresponding author Telephone: +1(603)8687667; Fax: +1(603)8687604; jgove@fs.fed.us, jhgove@unh.edu

Received 30 September 2011

Many new methods for sampling down coarse woody debris have been proposed in the last dozen or so years. One of the most promising in terms of field application, perpendicular distance sampling (PDS), has several variants that have been progressively introduced in the literature. In this study, we provide an overview of the different PDS variants and comprehensive comparison of their associated estimator performance. Additionally, two new variants and their estimators are introduced to the existing pool. Simulations are used to determine the efficiency of competing protocols and associated estimators. The simulation results corroborate the theoretical unbiasedness of the estimators for PDS in all forms. In addition, while there is some difference in variance among the protocols, the overall difference is small, and depends on the attribute to be estimated. Therefore, the choice of which PDS variant to use in field applications will depend on the field conditions and attributes to be estimated.

Introduction

The estimation of attributes within the different components of forest ecosystems continues to be an important subject of great interest to both scientists and practitioners alike. The standing tree component of forests has been the recipient of much effort in advancing sampling designs to estimate volume, biomass, diameter distributions and related quantities, for the obvious reason that there is economic value associated with the standing crop. The down wood component of forests constitutes a biomass pool that is important for other reasons, including its value for wildlife habitat, its importance as a fuel source in fire-prone ecosystems and its ability to store carbon for moderate time periods. Historically, not as much effort was devoted to the down coarse woody debris (CWD) component of the forest and only a few methods were available for sampling CWD such as line intersect sampling (LIS)¹ and fixed-area plots. However, the past dozen years have seen a proliferation of competitive new methods that provide a number of choices that are optimized for several important attributes and hence make inventorying the deadwood pool easier and more efficient in general.

Most of the new methods that have been developed recently are based on probability proportional to size sampling theory and are optimized for a chosen size attribute, allowing for the design-unbiased estimation of that design attribute by only counting the number of pieces of down CWD (hereafter simply logs) that are included in the sample at a given location.

These methods often have simple related strategies for the estimation of other log attributes as well. Similar to LIS, the attribute of choice is often the log length. Transect relascope sampling,² distance-limited sampling (DLS)³ and the so-called 'sausage sampling' protocol for fixed-area circular plots⁴ are examples of new methods developed to sample with probability proportional to length. A point-based equivalent to transect relascope sampling, point relascope sampling, samples logs with probability proportional to length-square.⁵ Both length and length-square are closely correlated with log volume, and thus these methods offer alternatives for the estimation of volume or closely related attributes such as woody biomass and carbon content. However, if volume-related attributes are the main goal of the survey, then perpendicular distance sampling (PDS),⁶ which samples logs with probability proportional to volume is a useful alternative, and has been specifically recommended for the estimation of carbon content precisely because of this property.⁷ With slight modification, PDS can be viewed as a probability proportional to surface or coverage area method as well.^{8,9} Another closely related method, line intersect distance sampling (LIDS), combines the strengths of PDS for volume estimation, with the transect protocol of LIS,¹⁰ resulting in a method that samples larger logs more frequently than LIS alone, for example.¹¹ Finally, methods have been developed that allow the selection of logs using a prism as in standing tree inventories. These methods are again useful for volume and related attribute estimation.^{12,13}

Of all the new methods mentioned for sampling CWD, those based on PDS appear to have received the most attention in terms of extensions and field testing. This is perhaps due to the simplicity of the method as originally envisioned in the field applications for the estimation of volume. Under PDS, a count of the number of logs that meet the criterion for being included in the sample on a given point provides an estimate of volume for that point. Additionally, PDS was found to perform better than LIS for the estimation of aggregate volume in simulation studies,⁶ and was competitive in field studies¹⁴; these results provided support for the general theory and its application. The original theory, however, was constrained to the estimation of volume, and this was an encumbrance to the application of the method when estimates of other attributes were desired. Similarly, it was realized early on that large logs could exist far from the sample point and still be selected, requiring an extended search effort to make sure no logs were overlooked on each sample point. These potential problems, as well as others (e.g. handling logs of various shapes), have now been effectively dealt with through clarification and extensions to PDS in the literature, and will be described in more detail in what follows.

The result of this active research work is that several variants of PDS now exist in the form of protocols, and these protocols have extended the method far beyond the original conception. The purpose of this paper is threefold. First, because PDS now exists in several forms, it is useful to synthesize the salient points concerning the different protocols for a complete understanding of the method. Secondly, two new sampling protocols are added here to more fully complete the family of estimators that has become PDS. Finally, a set of simulations is presented to facilitate the comparison of the different PDS variants. This type of comprehensive comparison should serve to both clarify and codify PDS into a more easily apprehended method for sampling down CWD. Note that only point-based protocols are considered herein – Affleck^{10,11} should be consulted for comparisons of the LIDS variant to PDS. A list of the symbols used in this paper along with their meanings can be found in Table 1.

PDS methods

Preliminaries

All variants of PDS are areal sampling methods that are based on sampling logs with probability proportional to some attribute of size. The areal component comes from the idea that each log in the population on a given tract, \mathcal{A} , has an inclusion zone associated with it. The inclusion zone is simply that area projected onto the tract datum within which a random sample point could fall and select the log into the sample record. The area associated with a log's inclusion zone can be calculated and depends on some attribute of the log, in addition to design parameters inherent in the sampling method. In the case of PDS, all variants employ an expansion factor (\mathcal{K}) that inflates the attribute of interest for the i th log to a larger area, a_i , the inclusion zone area, based on some measure of log taper: e.g. radial or cross-sectional taper. Thus, the inclusion zones are proportional to log taper on either side of the log, with proportionality constant \mathcal{K} . This yields a bilaterally symmetric inclusion zone along each log's needle

(the pith on straight logs, see de Vries¹⁵ and Williams *et al.*¹⁶ for more details on branched or crooked logs).

One design-based criterion that is common to all of the PDS variants is the concept of the selection attribute used to determine the inclusion zone of individual logs. Logs can be selected with probability proportional to volume (PDS_V), surface area (PDS_S) or coverage area (PDS_C),^{6,9,14} determining the selection protocol. On any given inventory, one would generally fix the selection protocol for all sample points so that, for example, the entire survey would be conducted with PDS_V . As part of the survey design, the PDS expansion factor, \mathcal{K} , would also be fixed in advance. This is similar to other methods of sampling, from fixed-area plots, where the plot radius would normally be fixed for the entire survey, or sampling standing trees with a constant basal area factor for all points.

Table 2 presents the necessary information to distinguish the three selection-based variants of PDS. For example, when sampling with probability proportional to volume, the inclusion zone width is proportional to log cross-sectional area, and is determined by expanding the cross-sectional area about both sides of the log by a factor of $D_l = \mathcal{K}x(l)$ and integrating. As l varies from 0 to L_i for the log, D_l determines the perimeter of the inclusion zone, or the maximum search distance perpendicular to the needle for each side of the log. Integrating cross-sectional area, $x(l)$, over the entire log length yields the log volume (V), so integrating $2D_l$ yields $a_i = 2\mathcal{K}V$. Therefore, the inclusion zone area is proportional to volume, making PDS_V a probability proportional to volume sampling method where each log selected contributes $\mathcal{F} = (10\,000/2\mathcal{K})\text{m}^3\text{ha}^{-1}$ to the per unit area estimate. Under the PDS_V selection protocol then, a count of the logs on a sample point yields an estimate of the volume of logs on the tract for that point. This follows because when the attribute of interest on a log is volume, $y_i = V_i$, and the estimator reduces to $|A_i|/2\mathcal{K}$ for the i th log. Similar interpretations for both PDS_S and PDS_C can be gleaned from Table 2 – more details are provided in the papers cited earlier.

Inclusion zones for the three selection-based protocols for PDS are compared in Figure 1. Because the functions of log diameter presented in Table 2 are in metres, it is easy to see that PDS_V will have the smallest maximal search distance and inclusion zone area for logs whose largest diameter is less than 1.27 m. For logs whose diameter is larger than 1.27 m, coverage area will generally have the smallest inclusion area of the three selection protocols. Inclusion zone widths for surface area and volume do not intersect until log diameters reach 4 m. Note, however, that $c(l) = \pi d(l)$ in general, so that the inclusion zone width for PDS_S is simply a scaled version of PDS_C .

Traditional PDS

The original concept behind PDS was for the design-unbiased estimation of volume: a simple count of the number of logs on each sample point was all that was required.⁶ Other quantities were initially deemed to be non-estimable in field application under the PDS_V method. The reason for this is that, for any of the selection protocols given in Table 2, one must know the true value of the selection attribute (e.g. volume), which appears in the denominator of the estimator, to derive unbiased estimates whenever y_i

Table 1 Symbol table

Symbol	Meaning
ρ	Bulk density (mass per unit volume) – in general, this can be a function of l
ψ	Biomass to carbon conversion (mass carbon per mass wood) – in general, this can be a function of l
π	Universal constant (unsubscripted); inclusion probability (subscripted)
\mathcal{A}	A physical tract of land
$ \mathcal{A} $	Area of tract \mathcal{A} (m ²)
a	Inclusion zone area (m ²)
\mathcal{B}	Woody biomass (appropriate mass units)
C_a	Carbon content (appropriate units)
C	Coverage area (m ²)
C	Coverage area (m ² ha ⁻¹)
$c(l)$	Log circumference (m) at length l
$D_{\perp}(l)$	Perpendicular distance from the log needle to the sample point at length l as determined by the sample point
D_l	Distance search limit (m)
D_{\max}	Maximum search distance (m)
$d(l)$	Log diameter (m) at length l
d_b	Log large end diameter (m) at $d(0)$
d_L	Log small end diameter (m) at $d(L)$
\mathcal{F}	Per unit area volume, surface area or coverage area factors (see Table 2)
\mathcal{K}	PDS expansion factor (for units, see Table 2)
l	Length (m) along the log, $0 \leq l \leq L$
L	Total log length (m)
m	Number of sample points
m_e	Number of sample points covered by inclusion zones
N	Number of logs in the synthetic population used in simulation
n_j	Number of sampled logs recorded at the j th sample point
n_{j1}, n_{j2}	Number of logs recorded for the DLS and PDS portions of DLPDS, respectively, on the j th sample point where $n_j = n_{j1} + n_{j2}$
r	Taper equation parameter expressing the log form such that $0 < r < 2$ is neiloid, $r = 2$ is a cone and $r > 2$ is paraboloid
S	Surface area (m ²)
S	Surface area (m ² ha ⁻¹)
$x(l)$	Log cross-sectional area (m ²) at length l
y	Some attribute measured on a log
Y	The aggregate total over \mathcal{A} for a given attribute y
V	Volume (m ³)
V	Volume (m ³ ha ⁻¹)

Table 2 PDS quantities for each of the selection-based protocols

Selection protocol	a_i	D_l	Estimator ¹	Selection integral	\mathcal{K} units	\mathcal{F}^2
PDS _V	$2\mathcal{K}V$	$\mathcal{K}x(l)$	$\sum_{i=1}^{n_j} \frac{ \mathcal{A} y_i}{2\mathcal{K}V_i}$	$V = \int_0^L x(l) dl$	m ⁻¹	$\frac{10\,000}{2\mathcal{K}}$
PDS _S	$2\mathcal{K}S$	$\mathcal{K}c(l)$	$\sum_{i=1}^{n_j} \frac{ \mathcal{A} y_i}{2\mathcal{K}S_i}$	$S = \int_0^L c(l) dl$	[dimensionless]	$\frac{10\,000}{2\mathcal{K}}$
PDS _C	$2\mathcal{K}C$	$\mathcal{K}d(l)$	$\sum_{i=1}^{n_j} \frac{ \mathcal{A} y_i}{2\mathcal{K}C_i}$	$C = \int_0^L d(l) dl$	[dimensionless]	$\frac{10\,000}{2\mathcal{K}}$

¹ For the j th point.

² In the same units as the selection attribute ha⁻¹.

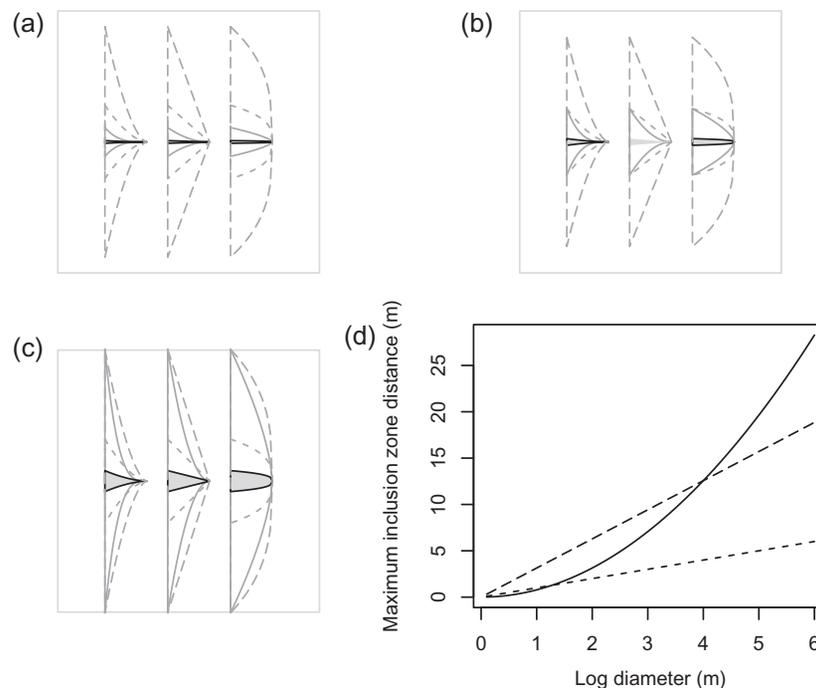


Figure 1 A comparison of inclusion zones for PDS when the selection attribute is volume (solid), surface area (long dash) and coverage area (short dash). The logs vary in geometric shape (taper) from neiloid (left), to conic (middle) and paraboloid (right) in each frame, with butt diameters of (a) 50 cm, (b) 127 cm and (c) 400 cm. (d) The inclusion zone width relations over a larger diameter range with $K = 1$.

is not equivalent to that attribute. Employing models to estimate these quantities imparts a bias in the estimate of unknown magnitude and is therefore not recommended. To help circumvent these peculiarities, *ad hoc* methods were proposed. For example, alternative companion sampling methods like circular plots were suggested to estimate other quantities such as log density.⁹

This limitation implies that PDS is useful only for estimating the attribute of interest, regardless of the selection protocol. A solution to this dilemma that allowed the estimation of any quantity under PDS that could be written in an integral form was proposed by Ducey *et al.*⁸ The solution proposed by the authors (see also Gregoire and Valentine¹⁷) involves employing a crude Monte Carlo (CMC) approach¹⁸ to the estimation of the selection attribute integral and the quantity of interest, although the estimator can also be derived from continuous Horvitz–Thompson (HT) theory.¹⁹ For example, if interest is in the estimation of some attribute y , then it can be shown that all that is required is to develop a judiciously chosen companion integral relation for this quantity in the form $y = \int g(l) dl$. Given this relation, the following estimator of the total for the j th sample point can be employed for each of the PDS selection variants

$$\hat{Y}_j = \frac{|A|}{2K} \sum_{i=1}^{n_j} \frac{g_i(l)}{f_i(l)} \quad (1)$$

This estimator follows because the selection attribute of volume, surface area and coverage area can similarly always be written

in integral form, $\int f(l) dl$, which can also be estimated by CMC. Here we denote the quantity $g(l)/f(l)$ in (1) as the Monte Carlo ‘attribute ratio’. The required relationships for each selection protocol are given in Table 3. For example, the CMC estimate for volume derives from cross-sectional area measured perpendicular to the sample point; therefore, $f(l) = x(l)$. Now if log density is to be estimated, then $g(l) = L^{-1}$, resulting in the estimator fraction given in Table 3 for this attribute under PDS_v. In this case, both the cross-sectional measurement and the log length, L , must be recorded. Other relations are similarly developed for each attribute to be estimated under the different selection protocols of PDS; details and proof of unbiasedness are given in Ducey *et al.*⁸ Notably, it is easy to see that under each of the selection variants, the selection attribute cancels where $g(l) \equiv f(l)$ in the attribute ratio in (1), and the estimator reduces to the traditional PDS form. We will refer to this method as ‘omnibus’ PDS (OPDS) in what follows, while the original form, where only the selection attribute is readily estimable will be termed ‘canonical’ PDS.

Both canonical and omnibus PDS can be used on all attributes listed in Table 3 in simulations, because the selection attribute can be calculated exactly under each selection protocol. However, only omnibus PDS imparts the ability to use each of the three selection variants to estimate any attribute in a field survey as well. It should be noted that the inclusion zone for a given log is *exactly* the same under both canonical and omnibus PDS. What differs is the field measurement protocol in each case, and the resulting estimators under omnibus for the allowable attributes of interest.

Table 3 Integral quantities that can be easily estimated under CMC, and associated attribute ratios for use in the omnibus protocols

Attribute	$g(l)$	Integral	Attribute ratio ¹		
			PDS _V	PDS _S	PDS _C
Density	$\frac{1}{L}$	$\int_0^L \frac{1}{L} dl = 1$	$\frac{L^{-1}}{x(l)}$	$\frac{L^{-1}}{c(l)}$	$\frac{L^{-1}}{d(l)}$
Length	1	$\int_0^L dl = L$	$\frac{1}{x(l)}$	$\frac{1}{c(l)}$	$\frac{1}{d(l)}$
Surface area ²	$c(l)$	$\int_0^L c(l) dl \approx S$	$\frac{c(l)}{x(l)}$	$\frac{c(l)}{c(l)}$	$\frac{c(l)}{d(l)}$
Coverage area	$d(l)$ ³	$\int_0^L d(l) dl = C$	$\frac{d(l)}{x(l)}$	$\frac{d(l)}{c(l)}$	$\frac{d(l)}{d(l)}$
Volume	$x(l)$	$\int_0^L x(l) dl = V$	$\frac{x(l)}{x(l)}$	$\frac{x(l)}{c(l)}$	$\frac{x(l)}{d(l)}$
Biomass	$\rho x(l)$	$\rho \int_0^L x(l) dl = B$	$\frac{\rho x(l)}{x(l)}$	$\frac{\rho x(l)}{c(l)}$	$\frac{\rho x(l)}{d(l)}$
Carbon	$\psi \rho x(l)$	$\psi \rho \int_0^L x(l) dl = C_a$	$\frac{\psi \rho x(l)}{x(l)}$	$\frac{\psi \rho x(l)}{c(l)}$	$\frac{\psi \rho x(l)}{d(l)}$

¹The Monte Carlo ratio $g(l)/f(l)$ in (1).

²This integral is approximate, but very close – compare with Table 5.

³Diameter (m) parallel to the horizontal plane of the ground.

Distance-limited PDS

In the previous section, we showed how omnibus PDS can be used to overcome the shortcomings of canonical PDS with regard to estimating a wide array of useful attributes on the population of down CWD. However, a second perceived shortcoming that afflicts both canonical and omnibus PDS is the exaggerated search distances that can accompany certain combinations of \mathcal{K} for large-diameter logs. This is especially true when sampling with probability proportional to surface area under PDS_S, where large search distances can result, even for small and moderate diameter logs (Figure 1). Recently, Ducey *et al.*¹⁴ introduced distance-limited PDS (DLPDS) as a way to minimize missed logs due to non-detection by restricting the search distance for logs in the area of larger diameters. Consequently, the maximum width of the inclusion zone on either side of a log's needle is constrained by the limiting distance.

DLPDS can most easily be envisioned as a combination of two sampling methods: PDS and DLS.³ The DLS method simply samples a log when the perpendicular distance from the sample point to the log's needle ($D_{\perp}(l)$) is less than some distance limit D_1 forming a rectangular inclusion zone encompassing each log. Two field measurement protocols were suggested for DLS. The first is based on HT theory and the normal field measurements that would be taken in an inventory for most areal sampling methods, like fixed-area plots. In this protocol, y_j is normally a direct measurement or estimate of the quantity to be estimated on the log; for example, an estimate of volume from Smalian's formula might suffice for y_j in the case of volume estimation. The second DLS measurement protocol is based on CMC and utilizes the same

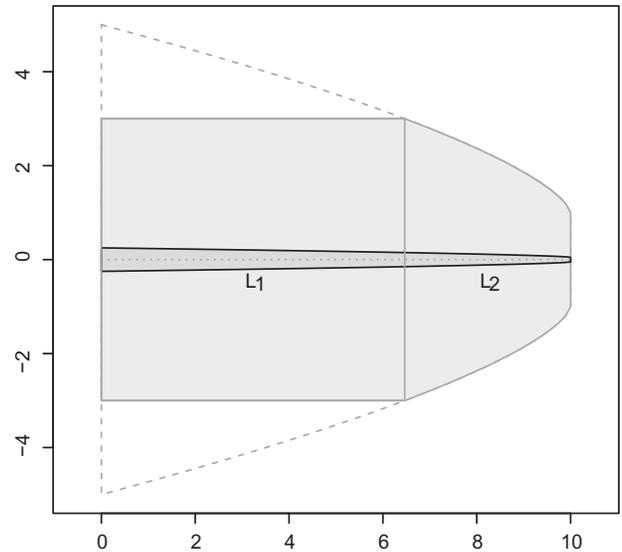


Figure 2 An inclusion zone for DLPDS (shaded) using the selection protocol PDS_C with $D_1 = 3$ m and $\mathcal{K} = 10$. The log has dimensions $L = 10$ m, $d_b = 0.5$ m and $d_u = 0.1$ m shown with 'needle' (dotted). The transition diameter is $d(L_1) = 0.3$ m with $L_1 = 6.46$ and $L_2 = 3.54$. The inclusion zone outline for PDS_C is also shown for comparison (dashed).

concept of a random measurement on $g(l)$ given in Table 3; for example, a design-unbiased estimate of volume is established by measuring the cross-sectional area on the log perpendicular to the sample point. Therefore, the CMC protocol for DLS is closely related to omnibus PDS in terms of field measurements. In general, CMC is asymptotically unbiased leading to a design-unbiased estimator for DLS, which Gove *et al.*³ corroborated via simulation.

DLPDS is illustrated in Figure 2. The rectangular DLS portion of the shaded inclusion zone area has width D_1 , while its length (L_1) depends on the properties of the log (i.e. the log length and taper) and \mathcal{K} . The rest of the shaded area comprises the PDS component of the log having length L_2 , such that $L = L_1 + L_2$. A given log can be all PDS, all DLS or a combination of both, again depending on the log dimension and the design parameters of the survey, which now include both \mathcal{K} and D_1 . It is perhaps simplest to consider the limiting or transition diameter $d(L_1)$, if any, as dividing the log into two separate logs, which are then treated according to their respective protocols. The limiting diameter can be easily deduced for each of the PDS selection protocols using the relationships for D_1 given in Table 2.

The estimators for DLPDS are given in Table 4. Ducey *et al.*¹⁴ introduced DLPDS as a design-unbiased method for estimating volume or surface area, and this approach also extends to the estimation of the coverage area. Their estimator was presented in the form of a hybrid approach (HDLPDS) where the PDS component of the log (if any) is sampled with canonical PDS to take advantage of design-unbiasedness, whereas the DLS portion of the log (again, if any) is sampled with the DLS CMC protocol. However, there are two other useful protocols that are based on the concepts already discussed: these are again referred to as canonical (DLPDS) and omnibus (ODLPDS). The former parallels canonical PDS, augmented with the basic HT-based DLS field

Table 4 DLPDS estimators on the j th plot for each of the selection-based protocols. In each case, the leftmost component corresponds to the DLS estimator and the rightmost to the PDS estimator

Selection protocol	Canonical estimator	Omnibus estimator	Hybrid estimator
PDS _V	$\frac{ \mathcal{A} }{2D_l} \sum_{i=1}^{\eta_{j1}} \frac{y_i}{L_i} + \frac{ \mathcal{A} }{2\mathcal{K}} \sum_{i=1}^{\eta_{j2}} \frac{y_i}{V_i}$	$\frac{ \mathcal{A} }{2D_l} \sum_{i=1}^{\eta_{j1}} g_i(l) + \frac{ \mathcal{A} }{2\mathcal{K}} \sum_{i=1}^{\eta_{j2}} \frac{g_i(l)}{x_i(l)}$	$\frac{ \mathcal{A} }{2D_l} \sum_{i=1}^{\eta_{j1}} g_i(l) + \frac{ \mathcal{A} }{2\mathcal{K}} \sum_{i=1}^{\eta_{j2}} \frac{y_i}{V_i}$
PDS _S	$\frac{ \mathcal{A} }{2D_l} \sum_{i=1}^{\eta_{j1}} \frac{y_i}{L_i} + \frac{ \mathcal{A} }{2\mathcal{K}} \sum_{i=1}^{\eta_{j2}} \frac{y_i}{S_i}$	$\frac{ \mathcal{A} }{2D_l} \sum_{i=1}^{\eta_{j1}} g_i(l) + \frac{ \mathcal{A} }{2\mathcal{K}} \sum_{i=1}^{\eta_{j2}} \frac{g_i(l)}{c_i(l)}$	$\frac{ \mathcal{A} }{2D_l} \sum_{i=1}^{\eta_{j1}} g_i(l) + \frac{ \mathcal{A} }{2\mathcal{K}} \sum_{i=1}^{\eta_{j2}} \frac{y_i}{S_i}$
PDS _C	$\frac{ \mathcal{A} }{2D_l} \sum_{i=1}^{\eta_{j1}} \frac{y_i}{L_i} + \frac{ \mathcal{A} }{2\mathcal{K}} \sum_{i=1}^{\eta_{j2}} \frac{y_i}{C_i}$	$\frac{ \mathcal{A} }{2D_l} \sum_{i=1}^{\eta_{j1}} g_i(l) + \frac{ \mathcal{A} }{2\mathcal{K}} \sum_{i=1}^{\eta_{j2}} \frac{g_i(l)}{d_i(l)}$	$\frac{ \mathcal{A} }{2D_l} \sum_{i=1}^{\eta_{j1}} g_i(l) + \frac{ \mathcal{A} }{2\mathcal{K}} \sum_{i=1}^{\eta_{j2}} \frac{y_i}{C_i}$

η_{j1} and η_{j2} are the number of logs sampled by DLS and PDS, respectively, on the j th plot. $d_i(l)$ is the diameter (m) parallel to the horizontal plane of the ground.

measurement protocol. The latter parallels omnibus PDS, with the CMC field protocol used for DLS as well. The canonical DLPDS protocol is useful only in simulation studies for any attribute other than the design attribute, for the same reasons as PDS, since we must know the true value of the design attribute to derive unbiased estimates. This is also true of the hybrid approach for the estimation of quantities other than the design attribute. However, under omnibus DLPDS, we have a set of estimators that will provide design-unbiased estimates of any attribute through the associated random quantity $g(l)$, found in Table 3. The combination of Monte Carlo-based estimators under omnibus DLPDS provides a simple, consistent protocol useful for field application. Of course, a hybrid protocol that is the complement of the one presented here can also be envisioned, but its utility is questionable, so it will not be discussed further.

It is important to note that the estimators given in Table 4 are for a full count of logs on the j th sample point, where $\eta_j = \eta_{j1} + \eta_{j2}$ logs are sampled. The DLS component of the estimator applies to the η_{j1} logs that were selected under this protocol on the j th sample point. Similarly, the PDS component of the estimator applies to the η_{j2} logs that were selected under the appropriate PDS protocol on the point. Note that either quantity can also be a zero tally, in which case no logs were selected under the respective protocol. Finally, as in canonical and omnibus PDS, the inclusion zone for a given log is the same for all of the DLPDS protocols for fixed D_l , \mathcal{K} and selection protocol. The differentiation arises from the field measurement protocol and associated estimator adopted for estimation.

Simulation methods

Williams^{20,21} formalized the sampling surface approach to simulation for the comparison of areal sampling estimators. The sampling surface approach tessellates the tract \mathcal{A} into square grid cells of chosen resolution, where the centre of each cell is treated as a sample point location. The tract, \mathcal{A} , is then populated with down logs whose inclusion zones are known and can also be mapped. A buffer is used to ensure that all inclusion zones fall within the tract, eliminating any potential bias due to boundary overlap. A sampling surface is developed as follows for each protocol and attribute. The grid of all cells within \mathcal{A} constitutes

an exhaustive sample of size m , which depends on the grid cell resolution. At each grid cell, $j = 1, \dots, m$, a sample is effectively taken including the $\eta_j = 0, 1, 2, \dots$ logs whose mapped inclusion zones overlap the cell centre point. The appropriate estimators (Tables 2 and 4) are applied to each of the ‘in’ logs on the j th grid cell, yielding an estimate \hat{Y}_j for that cell. The variable height sampling surface is effectively built up by summation of the individual log attribute estimates in this manner for each of the m cells.

The variance of each estimator is directly related to the unevenness of the sampling surface and is approximated for a given resolution by

$$\text{Var}(\hat{Y}) = \frac{1}{m-1} \sum_{j=1}^m (\hat{Y}_j - \hat{Y})^2 \quad (2)$$

where the summation over all m sample points equates to summation over all m cells in the rectangular grid. Other statistics such as the surface mean (\hat{Y}), which provides the simulation estimate, can be similarly calculated. In particular, an estimate of the surface standard deviation is $\text{SD}(\hat{Y}) = \sqrt{\text{Var}(\hat{Y})}$. The closeness of the approximations to the true population attribute values depends on the grid cell resolution. As the resolution increases (grid cell size decreases), the number of sample points, m , increases yielding a better approximation. There will always be a small bias that is recorded in the simulations because one can never sample every point ($m \rightarrow \infty$). In general, this small recorded bias is an artefact of the grid cell resolution, not the sampling methods themselves. Past simulation studies using grid cell resolutions on the order of one-quarter to one-half metre have proven quite reasonable,^{3,4,6,13,22} therefore half metre resolution is used here. Finally, the sampling surface approach has been implemented for the sampling methods and attributes discussed here, as well as several others, in the ‘sampSurf’ package²³ for the R statistical language.²⁴

A simple taper equation²⁵ was used to construct the synthetic log population for the simulations. This equation is shown in Table 5, along with closed-form solutions to the corresponding volume, and coverage area equations, while surface area must

Table 5 Taper equation and associated equations used in the simulations

Attribute	Equation
Log taper (m)	$d(l) = d_u + (d_b - d_u) \left(\frac{L-l}{L} \right)^{2/r}$
Volume (m ³)	$V(l) = \frac{\pi}{4} \left[d_u^2 l + L(d_b - d_u)^2 \frac{r}{r+4} \left(1 - \left(1 - \frac{l}{L} \right)^{(r+4)/r} \right) + 2Ld_u(d_b - d_u) \frac{r}{r+2} \left(1 - \left(1 - \frac{l}{L} \right)^{(r+2)/r} \right) \right]$
Surface area (m ²)	$S = \pi \int_0^L d(l) \sqrt{1 + \frac{d'(l)^2}{4}} dl$ $d'(l) = -2 \frac{(d_b - d_u)(L-l)^{2/(r-1)}}{rL^{2/r}}$
Coverage area (m ²)	$C(l) = \frac{1}{(r+2)L^{2/r}} \left[(r+2)d_u l^{2/r} + (L-l)^{2/r} ((d_u - d_b)rL + (d_b - d_u)lr) + (d_b - d_u)rL^{(r+2)/r} \right]$

$d'(l)$ is the derivative term for the surface area integral – all diameters are in the same units as length.

be numerically integrated. Log forms can vary depending on the shape parameter from neiloid ($0 < r \leq 2$), through conical ($r = 2$) to parabolic ($r > 2$). A population of $N = 100$ logs was generated from this taper model with log forms in the range of $r \in [1, 10]$. Log dimensions were determined as random uniform deviates in the following ranges: $D_b \in [8, 40]$ cm, $D_u \in [0, 0.9] \times D_b$ and $L \in [1, 10]$ m. Each log was randomly placed within a tract of $|A| = 1$ ha with a random orientation angle $\phi_i \in [0, 2\pi]$, $i = 1, \dots, N$. The tract was minimally buffered such that all inclusion zones fit within the tract.

The synthetic log population that was used in the simulations and their associated inclusion zones is depicted in Figure 3. The survey design parameters K_V , K_C and D_l were chosen in an attempt to equalize the inclusion zone areas between PDS_V and PDS_C to facilitate the comparison with regard to sampling effort (m_e). In general, the goal of equality is elusive, however, and it is only possible to approximate as the inclusion zones scale differently with diameter (Table 2) and log form. PDS_S was not considered in the simulations as it is a scaled version of PDS_C, as mentioned earlier. Sixty-eight per cent of the logs in the synthetic population had truncated inclusion zones under distance-limited PDS_V (Figure 3), while 58% were affected under PDS_C (not shown). For this log population and this particular set of design parameters, the maximum search distance for both PDS selection protocols was not large, with a maximum of 6.1 and 4.7 m for PDS_V and PDS_C, respectively. This implied a maximum reduction in search distance under DLPDS of 3.1 and 1.7 m for the two selection protocols, respectively. Other simulations with differing design parameters were also run on this same population of logs to demonstrate the difference in efficiency under the various DLPDS protocols.

Results

It is worth reiterating that the omnibus methods are the only protocols that do not require an exact measurement of the selection

attribute for design-unbiased estimation of any attribute listed in Table 3 in field applications. In the simulation approach, these attributes are known exactly for each log, so it is possible to estimate any quantity desired under all of the different protocols, not just the omnibus variants. In addition, not all attributes were considered in the simulations; biomass and carbon are scaled versions of volume, and so they were not considered. Similarly, surface area is a scaled version of coverage area when the diameter is taken parallel to the horizontal plane of the ground, so it also was not considered.

Table 6 presents the results of the basic simulations under the PDS_V protocol. The most obvious result to be gleaned is that all of the protocols for all attributes do prove to be unbiased in accordance with theory. The small (largely less than 1%) bias recorded is a manifestation of the sampling (grid cell) resolution and would be recorded in any simulation approach, including pure Monte Carlo. The ‘bias’ of the magnitudes reported, therefore, is a ‘computational’ or ‘apparent’ bias and is due to the numerical precision of the simulation of the sampling surface and not to any inherent bias in the estimators – which are all unbiased in theory. For example, the small positive bias recorded in the estimation of density under the OPDS and ODLPDS protocols has a logical explanation and should not be misconstrued to indicate an underlying bias in the estimators. Referring to Table 3, it can be seen that each of the three omnibus protocols for density has an L^{-1} component associated with the estimator. As noted by Ducey *et al.*,⁸ this in combination with small cross-sectional area can cause an inflation in the estimator for very short logs and evidently contributes to this small increase in bias over the other estimates. Both omnibus estimators for density also have higher variance. This is directly related to small cross-sectional area in sample locations near the tip of short logs, yielding a maximum sampling surface value that is more than three times that for the canonical and hybrid protocols. The inflation of length estimates under the omnibus protocols is even more pronounced in terms of the surface maximum, because cross-sectional area goes to zero on logs that

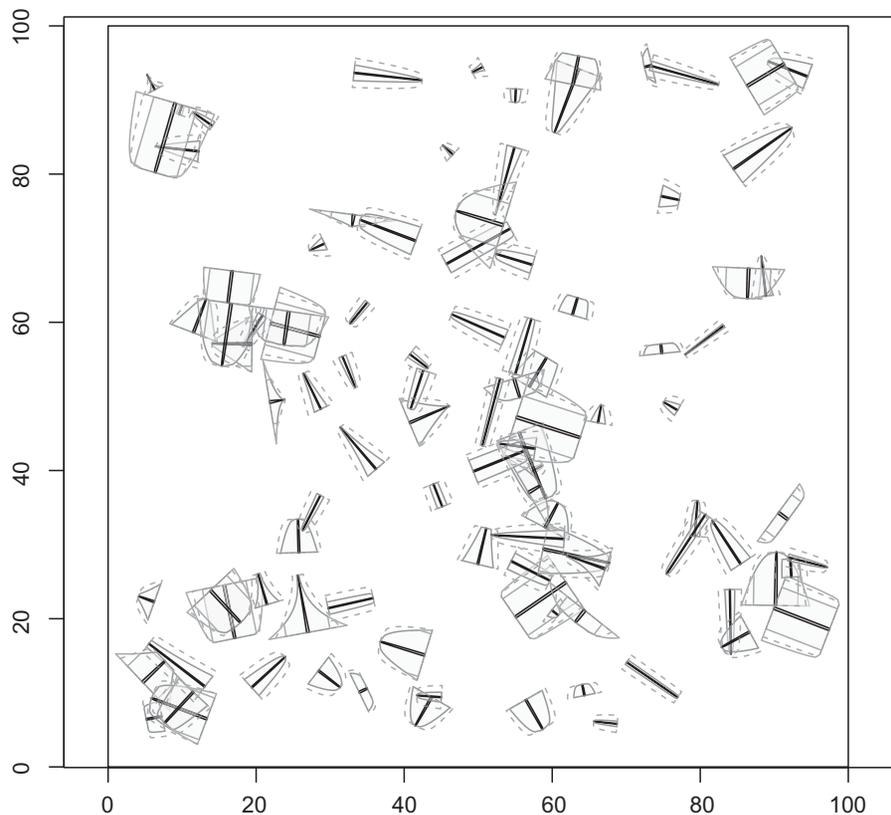


Figure 3 Synthetic log population of $N = 100$ logs used in the simulation experiments with inclusion zones for DLPDS (shaded) using selection protocol PDS_V with $D_l = 3$ m and $\mathcal{K}_V = 50$ m^{-1} . The inclusion zone outline for both canonical PDS_V (solid) and PDS_C (dashed, $\mathcal{K}_C = 12$) are also shown for comparison.

taper to the tip, and is not scaled by length as in the case of the density estimators, so it is independent of the log length. Again, this results in an increase in variance over the canonical and hybrid methods, but it does not result in any appreciable increase in apparent bias. The other results conform to what would be expected by the form of their estimators, which appear to be well behaved in general. For example, both canonical and omnibus estimates of volume reduce to a count of the logs under the PDS_V selection protocol, and so produce identical statistics, with surface maximum an integer factor of the $\mathcal{F}_V = 100$ based on a maximum of $n_j = 4$ overlapping inclusion zones in this simulation.

The simulation results for PDS_C are presented in Table 7. Again, under this selection protocol all the methods are unbiased. Additionally, in all cases, the variances are quite consistent among the methods. Both OPDS and ODLPDS record spikes in the surface maximum when estimating length. Referring to Table 3, this can be attributed to having log diameter in the denominator of the estimator: recording a small diameter near the end of a piece that tapers to the tip will inflate the estimate for the corresponding sampling grid point. Similar, but somewhat less severe inflation is recorded for density estimates under these two protocols where the addition of log length in the denominator attenuates the inflation due to small diameter. Under the PDS_C selection protocol, a count

of the number of logs yields an estimate of the coverage area so again, canonical and omnibus produce the same statistics. In this case, $\mathcal{F}_C = 416.67$ m^2 ha^{-1} , so the intersection of inclusion zones for four points again yields a maximum surface estimate of $n_j \mathcal{F}_C$.

A few general trends may be established for the variance (via SD) among the different protocols. Canonical and omnibus PDS always have smaller variance than the distance-limited protocols for both volume and coverage area, regardless of the selection protocol. When considering log density and length, the omnibus-based protocols (OPDS and ODLPDS) have generally higher variance than the others, with the exception of OPDS for density under PDS_C . This is perhaps due to the fact that log diameter scales linearly compared with cross-sectional area in the estimator denominator. In all cases, ODLPDS has the highest variance of the five estimators, regardless of the selection protocol or attribute being estimated. This may result from both components of the estimator being Monte Carlo based, which could contribute more variability to the estimate. Still considering the relative magnitude of the variances in each case, the increases are not large and the efficiency of the different protocols in terms of variance is surprisingly comparable.

Although the observations on variance noted above are indeed true for this particular set of design parameters, they will change

Table 6 Results of the simulations for a standard synthetic log population of size $N = 100$ logs under PDS_V. Sampling design constants were: $D_l = 3$ m and $\mathcal{K} = 50$ m⁻¹

Attribute/protocol	Population total	Sampling surface				
		Estimate	% Bias	SD	Max	m_e ¹
Volume (m³)						
PDS	18.31	18.29	-0.12	46.7	400	6049
OPDS	18.31	18.29	-0.12	46.7	400	6049
DLPDS	18.31	18.27	-0.21	50.4	360	5340
ODLPDS	18.31	18.27	-0.21	50.7	380	5340
HDLPDS	18.31	18.27	-0.21	50.7	380	5340
Length (m)						
PDS	525.42	526.84	0.26	1802.4	29 583	
OPDS	525.42	527.68	0.43	2049.4	115 079	
DLPDS	525.42	527.04	0.31	1812.7	29 582	
ODLPDS	525.42	528.51	0.59	2060.7	115 078	
HDLPDS	525.42	527.04	0.31	1812.7	29 582	
Density						
PDS	100	100.81	0.81	437.5	12 500	
OPDS	100	101.16	1.16	509.4	33 175	
DLPDS	100	100.86	-0.86	457.2	12 500	
ODLPDS	100	101.27	1.27	524.0	33 176	
HDLPDS	100	100.86	-0.86	457.2	12 500	
Coverage area (m²)						
PDS	100.66	100.64	-0.02	270.5	2544	
OPDS	100.66	100.63	-0.03	274.1	3828	
DLPDS	100.66	100.62	-0.04	278.6	2649	
ODLPDS	100.66	100.64	-0.02	282.8	3828	
HDLPDS	100.66	100.62	-0.05	278.9	2649	

¹Values for m_e , the number of grid cell centres covered by inclusion zones, are the same for each attribute.

under the distance-limited protocols as the D_l design parameter changes. As can be noted from Tables 6 and 7, there is a reduction of sampling effort (effective sample size, m_e) of 700 and 500 sample points (grid cells) under the distance-limited protocols for PDS_V and PDS_C, respectively, in this particular synthetic log population. As D_l approaches the maximum perpendicular distance, D_{max} , encountered on the largest log under canonical PDS, the sampling effort will approach that of PDS, since $L_1 \rightarrow 0$ when this happens (Figure 2). A subset of $N = 50$ logs were used for a broader set of simulations illustrating the effect of increasing D_l . The results for the standard deviation are shown in Figure 4 for $D_l \in \{2, 4, 6, 8\}$ m; the maximum width of the inclusion zones in this log population is $D_{max} = 12.2$ and 9.1 m, for PDS_V and PDS_C, respectively. The observations for canonical and omnibus PDS made above also hold for this new set of simulations because the inclusion zones are independent of D_l . Note particularly that OPDS has higher variance than canonical PDS for both length and density as described above and, of course, they coincide for the selection protocol attribute under PDS_V and PDS_C. In all cases, the variance for ODLPDS converges to that of OPDS. Similarly, the variances for DLPDS and HDLPDS converge to PDS. This can be verified in Table 4 by noting that as $D_l \rightarrow D_{max}$, $L_1 \rightarrow 0$ and thus $n_{j1} \rightarrow 0$, so this component of the estimator vanishes, leaving only the canonical component for each selection protocol. The

same observation can be made for ODLPDS which reduces to the OPDS estimator.

What appears to be an aberration in the results for the variance when estimating density under PDS_C, shows up in all three of the distance-limited estimators in Figure 4. The reason for the spike is that this particular distance limit produces several logs whose PDS component length is 3 cm or less, with the shortest being 0.7 cm. Referring to Figure 2 and Tables 3 and 4, the omnibus estimator has length, L_2 , in the denominator, and as this goes to zero, the estimator inflates. Likewise, under both DLPDS and HDLPDS, the computed coverage area for L_2 component of the log is very small (the smallest was on the order of 0.002 m²), inflating the estimator accordingly. These produce estimates (maximum surface values) for the particular grid cell point on the order of 60 000 m² ha⁻¹, under each of the three distance-limited protocols. This result is to be expected occasionally in simulations when a certain combination of distance limit, log length and shape, and PDS selection protocol combine to form a very small log length for either component of the three estimators in Table 4 and will inflate the estimate for density in this case.

Finally, with regard to these simulations, the convergence of the effective sample size (m_e) is presented in Figure 5 for the DLPDS variants, allowing comparison of the number of sample

Table 7 Results of the simulations for a standard synthetic log population of size $N = 100$ logs under PDS_C. Sampling design constants were: $D_l = 3$ m and $\mathcal{K} = 12$

Attribute/protocol	Population total	Sampling surface				
		Estimate	% Bias	SD	Max	m_e ¹
Volume (m ³)						
PDS	18.31	18.26	-0.31	43.2	314	7809
OPDS	18.31	18.26	-0.27	43.7	326	7809
DLPDS	18.31	18.24	-0.42	47.0	368	7309
ODLPDS	18.31	18.26	-0.28	47.6	397	7309
HDLPDS	18.31	18.25	-0.34	47.5	404	7309
Length (m)						
PDS	525.42	524.12	-0.25	1220.7	9926	
OPDS	525.42	524.28	-0.22	1268.7	50 316	
DLPDS	525.42	524.91	-0.10	1236.8	9926	
ODLPDS	525.42	525.06	-0.07	1286.1	50 316	
HDLPDS	525.42	524.91	-0.10	1236.8	9926	
Density						
PDS	100	99.82	-0.18	279.4	3603	
OPDS	100	100.05	0.05	304.2	21 595	
DLPDS	100	100.00	-0.01	316.4	8973	
ODLPDS	100	100.22	0.22	336.9	21 595	
HDLPDS	100	100.00	-0.01	316.4	8973	
Coverage area (m ²)						
PDS	100.66	100.40	-0.26	221.9	1667	
OPDS	100.66	100.40	-0.26	221.9	1667	
DLPDS	100.66	100.43	-0.23	231.9	1858	
ODLPDS	100.66	100.45	-0.21	232.5	1930	
HDLPDS	100.66	100.45	-0.21	232.5	1930	

¹Values for m_e , the number of grid cell centres covered by inclusion zones, are the same for each attribute.

points for each different distance limit (D_l). Note again that for a fixed set of design parameters, all DLPDS variants share the same inclusion zone for each log, as do both canonical and omnibus PDS. Under both selection protocols, the sample sizes converge non-linearly from approximately 44% of the full sample size at $D_l = 2$ m. This kind of convergence parallels the non-linear shape of the inclusion zones as $D_l \rightarrow D_{\max}$ for the largest diameter log in the population.

Confidence interval estimation

The sampling distributions of the estimators given herein play a role in determining how many samples are necessary in a field inventory for the nominal coverage rate to apply under normal theory confidence intervals. It is important to look at this aspect of the estimators because as Affleck¹⁰ showed, the sampling distributions can be far from Gaussian for variants of PDS studied there, leading him to question the appropriateness of normal theory intervals for these estimators. However, the overarching question is not whether normal theory intervals apply, because the central limit theorem covers the distribution of sample means under repeated random sampling designs^{26,27}; rather the main question is what sampling intensities might be necessary to attain near nominal coverage in the field applications. This is not a

trivial question to answer, because each combination of selection protocol, sampling protocol and attribute to be estimated has its own estimator and related sampling distribution; moreover, the empirical sampling distributions will differ in practice for each log population surveyed.

The sampling distribution of the estimators can be approximated for a given PDS sampling protocol under both PDS selection protocols and, for each attribute, by the empirical distribution of the estimates over all grid cells for each simulation run in Tables 6 and 7. Specifically, the zero-truncated distribution is used here because, while zero values are reasonable estimates (no samples at a given point), their number is dependent on the tract size, resolution and sample design parameters \mathcal{K}_V , \mathcal{K}_C and, for distance-limited variants, D_l . The resulting sampling distributions for the $N = 100$ log simulations reported here range from negative exponential to positively skewed and sometimes multimodal, similar to those described by Affleck.¹⁰ As noted earlier, the distributions for the selection attribute for each of the PDS_V and PDS_C selection protocols is integral valued (at the respective \mathcal{F}_V and \mathcal{F}_C levels) as well.

The results of a small Monte Carlo study on each of the estimators (selection protocol, sampling protocol and attribute combinations) is presented in Figure 6. These results, like those in Tables 6 and 7 on which they are based, are conditional on the synthetic log population used and will vary somewhat for

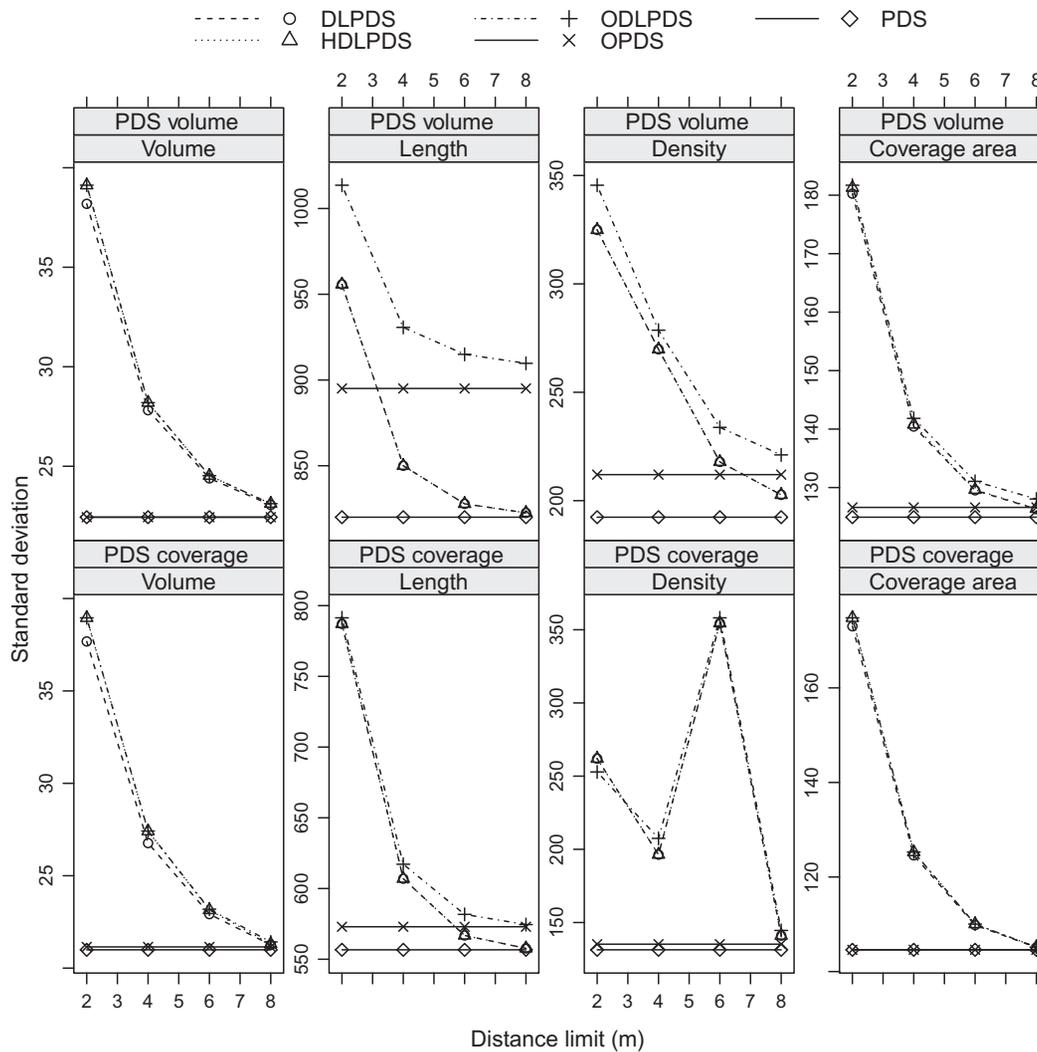


Figure 4 Comparison of standard deviations for a 50-log subset of the population in Figure 3 for all PDS variants. Both selection protocols, PDS_V ($\mathcal{K}_V = 100 \text{ m}^{-1}$) and PDS_C ($\mathcal{K}_C = 24$), are given for each attribute at different distance limits.

different populations. Therefore, we adopt a somewhat arbitrary rule that observed the coverage rates over approximately 92% will be adequate for nominal 95% coverage in other populations in lieu with a more comprehensive simulation study. In addition, we caution here, as elsewhere, that because the effective sample size cannot be equated between the two selection protocols due to the differing non-linear inclusion zone perimeters, it is unwise to make comparisons between the results for PDS_V and PDS_C . The variance of the estimators is directly related to the effective sample size in each population, with more sample points leading to smaller variance. For any of the protocols, increasing the inclusion area increases the effective sample size, decreases the variance and so smooths the empirical sampling distributions; this was demonstrated in Figure 4, and is true in general.

Under PDS_V and for all of the sampling protocols, it appears that a sample size of approximately 20–30 could be used to estimate volume and coverage area in terms of attaining nominal confidence interval coverage. Between 30 and 50 sample points

might be reasonable for the estimation of length under PDS, DLPDS and HDLPDS, whereas approximately 100 would be necessary for OPDS and ODLPDS. Density appears quite problematic for all of the estimators, requiring in the neighbourhood of 250 sample points for near-nominal coverage. PDS_C performed somewhat better, but again, this could be due to the larger overall effective sample size composing the sampling distributions in each case. All methods approach nominal coverage for volume estimation with sample sizes as small as 20. Again, sample sizes between 20 and 30 could be used for the estimation of the coverage area for all methods, with the possible exception that 30 should be used for ODLPDS. When estimating length, between 20 and 30 sample points could be used for close to nominal coverage. Finally, the estimation of density requires approximately 100 samples under all of the protocols.

The estimator coverage results are somewhat variable based on these simulations with the estimation of volume and coverage area requiring fewer sample points in general than length

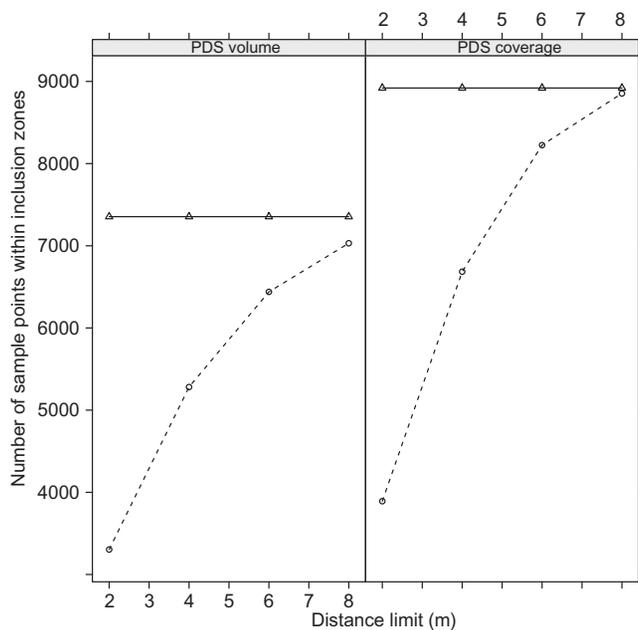


Figure 5 The effective sample size, m_e , for a 50-log subset of the population in Figure 3 with increasing distance limit for PDS (solid) and DLPDS (dashed), under selection protocols PDS_V ($\mathcal{K}_V=100 \text{ m}^{-1}$) and PDS_C ($\mathcal{K}_C = 24$).

and density. One obvious solution would be to choose the most variable attribute that is to be estimated and to determine the overall sample size for the inventory based on this. Another more equitable approach would be to use the smaller sample size for all attributes and supplement sample points for the estimation of length or density. Schemes like these have often been proposed with points established along a cruise line intermediate to the main sample points in sampling for standing trees, for example. Nevertheless, the main point of this exercise was to illustrate that all of the combinations of estimators do indeed converge to the nominal levels, regardless of how skewed or multimodal the underlying sampling distribution, even though the convergence rates differ. Therefore, normal theory confidence intervals are appropriate, true to theory, for all variants of PDS sampling given a sufficient sample size. However, a sufficiently large sample size for attributes like length and density may be too large for many down wood inventory applications, especially given that the natural variability of downed log populations may be greater than that observed in the simulations. In such cases, the use of sample sizes that are too small will lead to erroneous inferences when normal theory intervals are applied. Hence, for practical field applications, our results are in accord with Affleck,¹⁰ who conjectured that even bootstrap methods may fail to reach nominal coverage rates in some situations. Clearly, this is an area deserving further study in general; and again, we caution that the results will change for other log populations and combinations of design parameters.

Discussion

In this paper, we have shown that PDS can be presented in many forms. Design-unbiased estimators can be derived based

on classical HT estimation, ostensibly requiring the traditional suite of measurements such as length and volume, for estimation. However, the presence of the selection attribute (volume, surface area, coverage area) in the estimator (Tables 2 and 4) limits this form to the estimation of the design attribute only. The reason for this is that the presence of, for example, volume requires the exact measurement of this attribute for true design-unbiasedness of the estimator. Because this is normally impractical, the canonical PDS and DLPDS estimators should not be used for the estimation of other attributes; nor should HDLPDS, because the PDS component of this estimator is based on canonical PDS. It may be tempting to substitute a model estimate for the true volume (e.g. from Smalian's formula) but this imparts an unknown model bias to the estimate.

An alternative form of estimator can be derived from the application of Monte Carlo integration to the solution of the integrals in Table 3. The Monte Carlo connection stems from the measurement of the required quantities at a point on the log's needle that is perpendicular to a randomly chosen sample point, and is thus a Monte Carlo (random) selection of the measure of interest on the log. The omnibus protocols (OPDS and ODLPDS) use this method exclusively, and therefore can be applied to any attribute in a field inventory because there is no requirement for the knowledge of the true volume, surface area or coverage area of the logs selected. It is well known that the methods based on Monte Carlo integration strategies in general are unbiased by the Law of Large Numbers,²⁸ and therefore, the omnibus methods can also be shown to be unbiased,^{3,8} a fact that has been corroborated by our simulation results. As the results of this study have shown, the omnibus estimators tend to have slightly higher variance than the canonical estimators, and can require more samples for adequate confidence interval coverage rates. However, these results can be partially offset by judicious choice of the PDS design parameter \mathcal{K} .

It was noted earlier that it is possible under the various distance-limited protocols for the log length of the PDS component (L_2) to become very small. When this occurs, the estimator for density inflates under all three methods. This can also happen if the DLS component (L_1) becomes small, and will likewise affect all three estimators for density. However, it will also inflate the DLS portion of the canonical estimator for all attributes (Table 4). The main concern for field application, however, is the omnibus estimator for reasons mentioned previously. The inflated estimates resulting from very short logs are indeed true estimates and their precise enumeration under simulation is required to show unbiasedness. However, in field applications of ODLPDS, this phenomenon is less likely to be encountered, not because it is indeed unlikely, but because it is unlikely to be detected. The problem really arises from the lack of precision in measurements. First, there will be a judgment error in determining exactly where the perpendicular point on the log's needle is such that a log with tiny L_2 may not even be tallied. However, if tallied, the simple inconsistencies in taper due to the presence or absence of bark, partial deflation due to decay and numerous other factors may be a challenge in determining where the transition from distance-limited (L_1) to PDS (L_2) occurs when the perpendicular point lies at the end of the log. It is unlikely that a log length on the order of a few centimetres would even be able to be determined under these conditions. The consequence is that the log would most

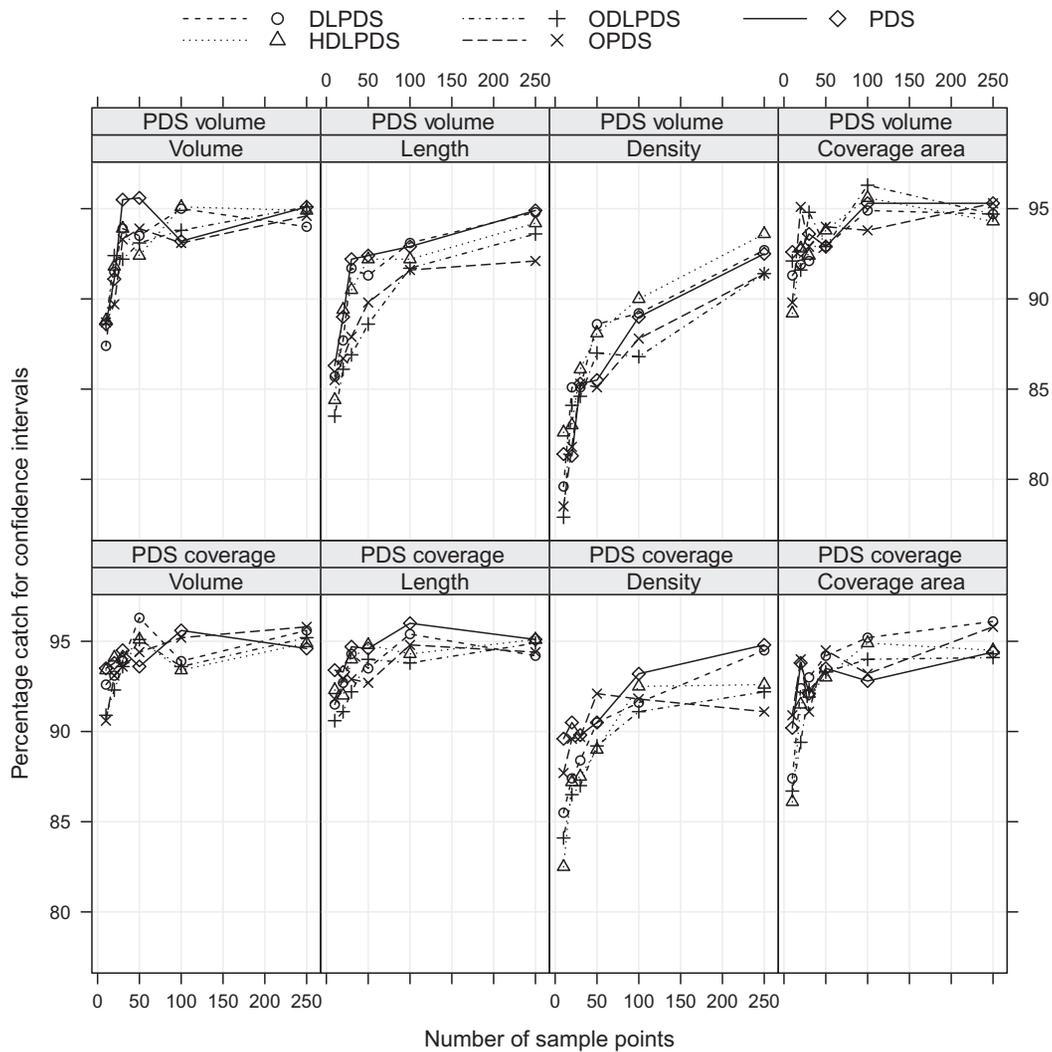


Figure 6 Approximate coverage percentages of the population mean for normal theory confidence intervals. The results are based on drawing 1000 repeated samples with replacement from the respective zero-truncated sampling (surface) distributions for the results in Tables 6 and 7. Samples of size 10, 20, 30, 50, 100 and 250 were drawn. Both selection protocols, PDS_V ($K_V=100 \text{ m}^{-1}$) and PDS_C ($K_C=24$), are given for each attribute at different sampling intensities.

likely be sampled as all DLS (i.e. when the L_2 component is very short). This is now a combined measurement and classification error that will theoretically impart a bias in the estimate, because the inflated estimate that would normally arise will be precluded. In actuality, of course, the measurement so taken will be unbiased, it will simply be for an incorrect classification of the perpendicular measurement taken. Errors of classification between the DLS and PDS portions of the log under DLS are less of a concern at other points along the log because the estimator behaves reasonably in these cases, and they will undoubtedly cancel over a large population of log measurements. Therefore, while very short log lengths, L_2 , are theoretically a potential source of bias in field application, in reality, its effect on the final estimate will be minimal.

In general, there are several relationships among the estimators that can be gleaned from Tables 3 and 4 and will always

hold. The canonical and omnibus PDS protocols will always give the same estimate for the selection attribute (e.g. volume for PDS_V). This is true for ODLPDS and HDLPDS as well, because the PDS components of the estimators both reduce to a count of the number of logs, while the DLS components are identical under these two protocols. Canonical DLPDS and HDLPDS will always yield identical estimates when sampling for length or log density. The PDS component of these two estimators is identical regardless of the attribute under consideration, so the DLS component is what determines the difference in the estimators and the resultant protocol. For length, the DLS component of these estimators reduces to a count of the number of logs, since $y_i = L_i$ and $g_i(l) = 1$. When estimating log density, the DLS components both reduce to $(|A|/2D_l) \sum (1/L_i)$ since $y_i = 1$ and $g_i(l) = 1/L_i$. All of these relationships may be verified in the estimators and are borne out by the simulation results in Tables 6 and 7.

We have mentioned three PDS sample selection protocols (volume, surface area and coverage area) and have contrasted two in this paper. What may not be so apparent is that canonical PDS is actually a very general method for sampling attributes on down logs (and perhaps other particles as well) that can be envisioned as an integral along the limits of log length. Any of the integrals in Table 3 could be used to derive an estimator for that attribute in the sense of PDS_V or PDS_C . For example, the estimator for log carbon content is simply a scaled version of PDS_V (assuming constant bulk density, ρ , within a log), and is therefore trivial to define. A less obvious example is that of DLS used in the distance-limited estimators shown here and in Ducey *et al.*¹⁴ DLS is an estimator of the aggregate length based on the integral in Table 3, and again can be derived in both HT and CMC forms³; equating D_l to \mathcal{K} provides the link. The exception is the estimation of density, where the solution to the integral is not based on any attribute of the log, and therefore does not appropriately fit the theory. In each of these cases where the selection attribute is used to define the estimator, a simple count of the number of logs will provide an estimate of the aggregate attribute of interest when properly expanded.

Earlier it was mentioned that normally one selection protocol would be chosen with associated fixed design parameters \mathcal{K} (and D_l under DLPDS) and implemented on all plots in the inventory. Here we loosen those guidelines somewhat and note that it is indeed possible to use each of the three selection protocols (PDS_V , PDS_C and PDS_S) plus DLS from each sample point to arrive at design-unbiased estimates of volume, coverage area, surface area and length from a simple count of the number of 'in' logs on the point. In this case, each method would have its own associated design parameter (\mathcal{K}_V , \mathcal{K}_C , \mathcal{K}_S and D_l) for the attribute of interest. Obviously, depending on how these are chosen, a given log could be sampled under one selection protocol (say PDS_V) and not selected under one or more of the others (Figure 1). While this type of design is indeed possible, note that keeping track of the different design parameters, each with associated limiting distance tests for borderline logs, could make this design confusing and hence error prone in the field. Unless absolute minimal variance estimation is required, one of the omnibus variants should be preferred to this approach in general.

All of the PDS protocols ultimately rely on the establishment of a log's needle, whether the log is straight, crooked or branched. Methods for dealing with these issues are discussed in detail elsewhere.^{16,29} On a more general level, the estimation of log volume can be a potential source of bias in the estimation of aggregate volume. Methods that rely on a model such as Smalian's or similar³⁰ involve assumptions based on the model, and ultimately impart a bias to the estimate. Discounting for deflation due to decay becomes problematic under such designs. The omnibus methods alleviate this problem by not relying on a volume model with its inherent source of error. However, they do still require an accurate measurement of diameter or cross-sectional area at the perpendicular point on the log, which can equally be affected by deflation and the like. Under PDS, the random measurement point provides the ability to sub-sample for estimates of decay or estimate bulk density at this point, which can be used to unbiasedly estimate attributes such as volume or carbon.⁷ The random measurement point is integral to the PDS

design protocols and therefore provides a designed-based solution to these potential problems without resorting to the use of a measurement model.

Finally, some form of boundary correction method is required for logs whose inclusion zones are intersected by the tract boundary. Williams and Gove⁶ describe a solution based on the boundary reflection method that is implemented in the field using the walk-through procedure about a log's needle.^{5,31} The method has been shown to be unbiased and is straightforward and simple to implement in practice. Furthermore, the method works for all of the PDS protocols discussed in this paper and also applies to DLS as well.

Summary and conclusions

At its most fundamental level, PDS is a single sampling method based on the idea that an integral quantity associated with a log attribute can be transformed to an associated estimator. When the different sample selection, field measurement and distance protocols are added to this, PDS is expressed as a family of sampling protocols and associated estimators. All of the estimators in the family are design unbiased, but not all are necessarily appropriate for practical implementation in the field. The seminal PDS design provides for the estimation of aggregate log volume on a forested tract from counts of logs at sample points. This canonical design has evolved to provide for the estimation of aggregate amount of other design attributes, namely, coverage area, surface area and length. By the application of CMC, the canonical formulations transform to omnibus formulations, which easily accommodate the estimation of aggregate log attributes that differ from the design attribute; e.g., log number or density is also easily estimated under the omnibus formulation. The precision of omnibus and canonical designs is identical whenever the design attribute and the measured attribute are the same. The canonical formulations are more precise when the measured and design attributes differ; however, the measurement of design attributes for canonical estimation such as volume, surface area or coverage area are impractical if not impossible, rendering the estimation of other attributes infeasible in field surveys. On the other hand, the requisite measurements for omnibus estimation are easily obtained, and any observed deficiencies in precision can be easily offset by a modest increase in the number of sample points, a judicious choice of \mathcal{K} , or both. Distance limitation increases the practicality of omnibus PDS in situations where logs run large in diameter or where sight lines are truncated by vegetation, topography or other obstructions.

The results of the simulations corroborate the unbiasedness of the different variants, but do not clearly point to any one of the variants as being superior to the others with regard to efficiency over the suite of estimable attributes tested. For example, it was noted that the estimators for log length and density under PDS_V show what might be construed as significant differences in the simulated standard deviations (Table 6). In reality, however, only the omnibus methods are useful for the estimation of these two attributes in field application, and the differences between these estimates are trivial in the context of the simulation. A related finding that should not be ignored is with regard

to the applicability of normal theory confidence intervals to small surveys using small or moderate sample sizes. Our findings here, for density and length especially, affirm those of Affleck¹⁰ and point to the need for more extensive future research in this area, not only for PDS, but also other probability proportional to size methods as well.

The flexibility of having multiple PDS variants with similar performance to choose from may prove to be its biggest strength, because the variety of the different variants allows PDS to be tailored to the field conditions and attributes of interest by choosing the most appropriate variant or variants for the situation. Finally, as noted earlier, the wider PDS family also includes the LIDS protocol, which has proved to be an efficient method in initial trials.^{10,11} LIDS combines LIS and PDS and is yet another option that should be afforded serious consideration where surveys based on lines rather than points may be more desirable.

Acknowledgements

The authors thank the two reviewers, Dr David L. R. Affleck and Dr Kim Iles, and the Editor, Dr Gary Kerr, for their very helpful and insightful comments on the manuscript.

Conflict of interest statement

None declared.

References

- Warren, W.G. and Olsen, P.F. 1964 A line intersect method for assessing logging waste. *For. Sci.* **3**, 267–276.
- Ståhl, G. 1998 Transect relascope sampling—a method for the quantification of coarse woody debris. *For. Sci.* **44**, 58–63.
- Gove, J.H., Ducey, M.J. and Valentine, H.T. 2012 A distance limited method for sampling downed coarse woody debris. *For. Ecol. Manage.* **282**, 53–62.
- Gove, J.H. and Van Deusen, P.C. 2011 On fixed area plot sampling for downed coarse woody debris. *Forestry* **84**, 109–117.
- Gove, J.H., Ringvall, A., Ståhl, G. and Ducey, M.J. 1999 Point relascope sampling of downed coarse woody debris. *Can. J. For. Res.* **29**, 1718–1726.
- Williams, M.S. and Gove, J.H. 2003 Perpendicular distance sampling: an alternative method for sampling downed coarse woody debris. *Can. J. For. Res.* **33**, 1564–1579.
- Valentine, H.T., Gove, J.H., Ducey, M.J., Gregoire, T.G. and Williams, M.S. 2008 Estimating the carbon in coarse woody debris with perpendicular distance sampling. In *Field Measurements for Forest Carbon Monitoring: A Landscape-Scale Approach*. C.M. Hoover (ed.). Springer, New York, pp. 73–87.
- Ducey, M.J., Williams, M.S., Gove, J.H. and Valentine, H.T. 2008 Simultaneous unbiased estimates of multiple downed wood attributes in perpendicular distance sampling. *Can. J. For. Res.* **38**, 2044–2051.
- Williams, M.S., Ducey, M.J. and Gove, J.H. 2005 Assessing surface area of coarse woody debris with line intersect and perpendicular distance sampling. *Can. J. For. Res.* **35**, 949–960.
- Affleck, D.L.R. 2008 A line intersect distance sampling strategy for downed wood inventory. *Can. J. For. Res.* **38**, 2262–2273.
- Affleck, D.L.R. 2010 On the efficiency of line intersect distance sampling. *Can. J. For. Res.* **40**, 1086–1094.
- Bebber, D.P. and Thomas, S.C. 2003 Prism sweeps for coarse woody debris. *Can. J. For. Res.* **33**, 1737–1743.
- Ståhl, G., Gove, J.H., Williams, M.S. and Ducey, M.J. 2010 Critical length sampling: a method to estimate the volume of downed coarse woody debris. *Eur. J. For. Res.* **129**, 993–1000.
- Ducey, M.J., Williams, M.S., Gove, J.H., Roberge, S. and Kenning, R.S. 2012 Distance limited perpendicular distance sampling for coarse woody material: theory and field results. *Forestry* **86**, 119–128.
- de Vries, P.G. 1986 *Sampling Theory for Forest Inventory*. Springer, Berlin, Heidelberg, New York, p. 249.
- Williams, M.S., Valentine, H.T., Gove, J.H. and Ducey, M.J. 2005 Additional results for perpendicular distance sampling. *Can. J. For. Res.* **35**, 961–966.
- Gregoire, T.G. and Valentine, H.T. 2008 *Sampling Strategies for Natural Resources and the Environment*. Chapman and Hall/CRC, New York, p. 341.
- Rubinstein, R.Y. and Kroese, D.P. 2008 *Simulation and the Monte Carlo Method*. 2nd edn. Wiley, Chichester, p. 27.
- Cordy, C.B. 1993 An extension to the Horvitz–Thompson theorem to point sampling from a continuous population. *Statist. Probab. Lett.* **18**, 353–362.
- Williams, M.S. 2001 New approach to areal sampling in ecological surveys. *For. Ecol. Manage.* **154**, 11–22.
- Williams, M.S. 2001 Nonuniform random sampling: an alternative method of variance reduction for forest surveys. *Can. J. For. Res.* **31**, 2080–2088.
- Gove, J.H., Williams, M.S., Ståhl, G. and Ducey, M.J. 2005 Critical point relascope sampling for unbiased volume estimation of downed coarse woody debris. *Forestry* **78**, 417–431.
- Gove, J.H. 2011 sampSurf: sampling surface simulation. <https://r-forge.r-project.org/projects/sampsurf/>.
- R Development Core Team. 2011 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>. ISBN 3-900051-07-0.
- Van Deusen, P. 1990 Critical height versus importance sampling for log volume: does critical height prevail? *For. Sci.* **36**, 930–938.
- Barrett, J.P. and Goldsmith, L. 1976 When is n sufficiently large? *Am. Statist.* **30**, 67–68.
- Barrett, J.P. and Nutt, M.E. 1979 *Survey Sampling in the Environmental Sciences: a Computer Approach*. COMPRESS, Inc., Wentworth, NH, USA, p. 38.
- Suess, E.A. and Trumbo, B.E. 2010 *Introduction to Probability Simulation and Gibbs Sampling with R*. Springer, Berlin, Heidelberg, New York, p. 67.
- Valentine, H.T., Gove, J.H. and Gregoire, T.G. 2001 Monte Carlo approaches to sampling forested tracts with lines or points. *Can. J. For. Res.* **31**, 1410–1424.
- Fraver, S., Ringvall, A. and Jonsson, B.G. 2007 Refining volume estimates of down woody debris. *Can. J. For. Res.* **37**, 627–633.
- Ducey, M.J., Gove, J.H. and Valentine, H.T. 2004 A walkthrough solution to the boundary overlap problem. *For. Sci.* **50**, 427–435.