

Article

Some Refinements on the Comparison of Areal Sampling Methods via Simulation

Jeffrey H. Gove

USDA Forest Service, Northern Research Station, 271 Mast Road, Durham, NH 03824, USA; jgove@fs.fed.us;
Tel.: +1-603-868-7667

Received: 15 August 2017; Accepted: 10 October 2017; Published: 16 October 2017

Abstract: The design of forest inventories and development of new sampling methods useful in such inventories normally have a two-fold target of design unbiasedness and minimum variance in mind. Many considerations such as costs go into the choices of sampling method for operational and other levels of inventory. However, the variance in terms of meeting a specified level of precision is always among the most important criteria. Similarly, in designing new sampling methods, one always seeks to decrease the variance of the new method compared to existing methods. This paper provides a review of some graphical methods that may prove useful in these endeavors. In addition, in the case of the comparison of variances between new and existing methods, it introduces the use of wavelet filtering to decompose the sampling variance associated with the estimators under consideration into scale-based components of variance. This yields an analysis of variance of sorts regarding how the methods compare over different distance/area classes. The graphical tools are also shown to be applicable to the wavelet decomposition. These graphical tools may prove useful in summarizing the results for inventory design, while the wavelet results may prove helpful as we begin to look at sampling designs more in light of spatial processes for a given population of trees or downed coarse woody debris.

Keywords: forest sampling; smith plot; Taylor diagram; wavelets

1. Introduction

There are numerous extant ways to inventory standing and downed components of forest ecosystems. Of course, the methods chosen and the intensity of sampling depend on many factors, not the least of which are the objectives of the cruise coupled with financial and manpower constraints. A ‘pre-cruise’ (or perhaps a database of inventory information from similar stands) is often recommended to help judge the intensity from a variability perspective, and refinements such as stratification can be used in this context to help reduce the number of sample points or lines. Often, however, the choice of method for actually selecting the individuals of interest on a sample unit may be made simply because of familiarity with a particular method, favoring it over other available alternatives. These are all interesting, practical and useful points that require being addressed when undertaking an inventory, and the aforementioned list is by no means complete. For example, the choice of a sampling method can be enhanced by targeting the attribute of greatest interest in the inventory and using a method that is either directly ‘optimized’ for that attribute or one that is correlated with the target quantity.

Another very useful tool that can be employed to aid in decisions concerning the choice of sampling method and, potentially, intensity is simulation. Synthetic populations on a variety of scales can be simulated to mimic conditions of the target population (e.g., [1]). This is most helpful in the event that something about the structure of the target population is known, such as from reconnaissance information or an existing database, as noted above. The method that yields the

smallest variance would be the most desirable in general. Alternatively, well-designed simulated populations can be subsampled in a Monte Carlo experiment using differing sample sizes to judge the efficacy of different areal sampling methods based on confidence interval capture rates; when the sample size approaches the nominal coverage rate, we have an approximate estimate for the required sample size (e.g., [2,3]). However, sample size alone is not the only factor that influences the outcome of such experiments in terms of precision. It has been known for decades that the higher the sampling effort (where sampling effort is defined in terms of average plot size), the lower the variance, this was evidently first demonstrated empirically by Smith [4] and is often referred to as ‘Smith’s law.’ Due to this relationship, the naive application of simulation or even on-the-ground test cruises for the comparison of a potential pool of sampling methods can lead to an erroneous conclusion if the sampling effort is not considered. In field studies, the best way to do this is arguably using time as a surrogate to cost. An example of a relative efficiency measure that has been useful in such studies is given in Ducey et al. [5] (and the references therein; derivation in their appendix), while Rubinstein and Kroese [6] (p. 122) demonstrate its usefulness in other applications. This latter method is not as useful in simulations, unless there is a model derived for each method that would allow the prediction of time taken on a sample unit based on the structure of the population. Such data are often either incomplete or completely lacking for this application, as it requires time studies for each sampling strategy and forest type (perhaps even extending to structural components within the type) considered. However, time or effort models can be developed with appropriate data for both fixed and variable radius sampling schemes (e.g., [7,8]). Recently, for example, Lynch [9,10] incorporated Zeide’s [7] model into the determination of optimal plot size. Where applicable data exist, these approaches could be readily included in simulation comparisons.

The previous motivation is only one part of the story on the use of simulation for the comparison of different sampling methods. A second application encompasses the development of new methods for sampling different forest components (e.g., standing live and/or dead trees or downed woody debris). Foresters have developed numerous methods for sampling the different vegetative components of interest, with much early work also in ecology, and a flurry of new methods in the intervening decades, many building on the foundations of the early innovative work in our field. A very succinct list would include the early works of Robertson [11] (fixed-area plot sampling), Bitterlich [12] and Grosenbaugh [13] (horizontal point or angle gauge sampling, HPS) with specialization for volume estimation by Kitamura [14] (critical height sampling, CHS) for standing trees and Warren and Olsen [15] (line intersect sampling, LIS) for downed coarse woody debris (CWD). The work of plant biologists and ecologists include Cottam and Curtis [16] (nearest neighbor and similar methods) and early contributions on LIS by McIntyre [17]. More recent (and intervening years) sampling methods that have built on these efforts include both extensions of all of the above methods and a number of new methods for both standing trees (including snags) or downed woody material (see, e.g., [18,19] for detailed treatments of the most recent methods). Briefly, when a new areal sampling method is developed, the creators begin with an analytic proof of unbiasedness, often following on the precepts of Palley and Horwitz [20]. Variances can also be derived as these authors do, but variance estimates change depending on the criteria previously mentioned among others. In addition, areal sampling designs result in quantities like areas of overlapping inclusion zones (see Section 2.2) that are part of the true estimator variance, but are impossible to calculate in the field, and non-trivial even via simulation; we instead resort to tractable unbiased variance estimators. Therefore, it is often useful to design simulation experiments that will aid in comparing new methods against well-known and -used methods for the purpose of comparing the resultant precision of the two, by way of some measure of variance. For a given population, sampling effort, and set of sampling methods, the latter with the smallest variance is often deemed the ‘best.’

It might be argued that the use of the variance alone is enough for most purposes. Reducing the variance is certainly our ultimate goal in judging the efficiency of different methods (as noted above, perhaps coupled with timing information when available), with the minimum variance method usually

chosen as the best from a simulation perspective or a strictly statistical comparison. However, if the minimum variance method is associated with significantly higher costs due to increased sampling effort, then the purely statistical perspective is not enough in practice. In this study, we take a different approach as our interest lies in trying to develop a set of tools that allow us to look more deeply into the relationship between sampling methods as judged by different components or properties of the variance through simulation. To do this, we will look at existing techniques that allow one to compare methods by the following:

1. Sampling effort: using plot size as a surrogate for time (cost) with regard to effort;
2. Spatial juxtaposition: looking at the covariance or, equivalently, correlation between the mapped simulated estimates of the methods under comparison; and
3. Variance decomposition by distance (area), which allows a finer look into the process that is under consideration in areal sampling, both via independent inclusion zones and those of overlapping objects.

Each of the techniques in the above list use the associated variances for different sampling methods on a given population as the basis for providing more depth to the analysis of the simulation results. The above list comprises the objectives of this study with the intent to provide a more useful set of tools for comparing sampling methods that provide either similar information to what is already in use, but in a simpler manner, or more details through methods that are not often applied to areal sampling simulation. A simulated population of a small northern hardwood stand [21] is used to illustrate the techniques in detail using four different sampling methods as an example.

2. Methods, Applications and Results

2.1. Simulation Methods

In general, any well-designed areal sampling simulation program can be used for simple comparison of variances and for some of the methods discussed in the following sections. However, as will be made clear shortly, some of the other techniques described here for comparing sampling methods require structured results in the form of raster images. *sampSurf* [1,22] is an R [23] package that can be used for the simulation of areal sampling methods commonly used in forestry. The package is somewhat unique in that it creates ‘sampling surfaces’ [24,25], which are raster images, as a result of the simulation. Briefly, an N row by M column raster tract, $\mathcal{A}$, which holds a synthetic population (either from actual field measurements or simulated) of standing trees or downed logs is created. Inclusion zones are added about the individual members of the population for the desired sampling methods to be simulated and those cells that are internal to each inclusion zone are appropriately filled with the estimate of the attribute desired for each individual. The final surface is then created by simple raster addition of estimates within overlapping inclusion zones. The center of each pixel is a sample point, which is used to determine whether the estimate registered for that pixel is within the inclusion zone boundary for a given tree or log. In practice, a resolution of $\Delta xy = 1$ m is often more than adequate, regardless of tract size, though finer resolutions have been used.

For the simulation examples used here, an internal tract size of 1.5 hectares is created with a $\Delta xy = 1$ m resolution, and an external 14 m buffer, such that $N = M = 150$ m, yielding $m = NM = 22,500$ sample points, with an overall area $|\mathcal{A}| = 2.25$ ha. The buffer contains inclusion zones that extend beyond the 1.5-ha internal tract where the trees reside as a simple unbiased method for handling boundary overlap [26]. A synthetic population of $n = 392$ trees (approximately 261 trees ha^{-1} with a basal area of approximately $18.4 \text{ m}^2 \text{ ha}^{-1}$) common to all simulations was generated using the built-in taper function with randomly assigned shape parameter $r \in [1.5, 3]$, where $0 < r < 2$ is

neiloid, $r = 2$ is a cone and $r > 2$ is paraboloid stem form; the taper function for diameter at height $0 \leq h \leq H$ is [27]:

$$d(h) = D_u + (D_b - D_u) \left(\frac{H-h}{H} \right)^{\frac{2}{r}} \quad (1)$$

where D_b is the butt diameter, with top diameter D_u at height H . Corresponding tree volumes are generated from the integral of this equation ([28] p. 8). The simulated diameter distribution followed a three-parameter Weibull distribution [29] with location, shape and scale parameters $\alpha = 10$ cm, $\gamma = 2$, $\zeta = 20$ cm, respectively; heights were generated using a metric version of the all species equation for northern hardwoods in New Hampshire given by Fast and Ducey [30]. The trees were dispersed throughout the tract area using a spatial inhibition process with an inhibition distance of 3 m ([31] p. 434). The resulting simulated stand would be representative of a small, well-stocked northern hardwoods stand based on the northern hardwoods stocking guide [21].

2.2. Areal Sampling Methods

In this paper, we will be using four different areal sampling methods for standing trees to illustrate the different procedures for method comparisons. Each of the sampling methods is design-unbiased and possesses an enlarged circular inclusion zone, $\mathcal{I}_k$, centered on the k -th tree with area a_k m². It follows from this that the inclusion probability for the k -th tree is:

$$\pi_k = \frac{a_k}{|\mathcal{A}|} \quad (2)$$

The inclusion probabilities are different for each tree in three of the four methods used, which select trees with probability proportional to size (PPS). Let z_k be a measurable attribute of interest on the k -th tree. Estimation follows that for variable probability sampling using the Horvitz–Thompson (HT) estimator ([18] p. 215, [32] p. 49) such that:

$$\hat{v}_{z_k} = \frac{z_k}{\pi_k} \quad (3)$$

is the estimator for quantity z_k . Similarly, given n_s individuals sampled on the s -th point, the estimator for that point is:

$$\hat{v}_{z_s} = \sum_{k=1}^{n_s} \hat{v}_{z_k} \quad (4)$$

Using the sampling surface approach, each grid cell receives the estimate $\hat{v}_{z_s}$ corresponding to the sample point at its center. The surface total, $\hat{v}_z$, is given as the mean over the m grid cells and is an estimate of the population total, v_z , which can be deduced from the known synthetic population of trees. The surface variance measures the roughness around the mean plane as:

$$\text{Var}(\hat{v}_{z_s}) = \frac{1}{(m-1)} \sum_{s=1}^m (\hat{v}_{z_s} - \hat{v}_z)^2 \quad (5)$$

In the simulations that follow, z_k represents total tree volume (m³) unless specifically noted otherwise.

The first sampling method, horizontal point sampling (HPS) ([12,18] p. 247, [13]), selects individuals with probability proportional to basal area, yielding inclusion probability $\pi_k = \frac{b_k}{\mathcal{F}|\mathcal{A}|}$, where $\mathcal{F}$ m² ha⁻¹ is the basal area factor (BAF) and b_k m² is the basal area for the k -th tree. The true volume is known for each tree and is calculated by the integration of (1) as noted above. The second sampling method is a simple variant of the previous method where HPS is used to select the individual trees at a given point, which are then subsampled for volume estimation using crude Monte Carlo sampling (CMC) ([18] p. 94). This method shares the exact same inclusion probability with HPS and

differs only in the calculation of individual tree volumes at each sample point. Calculation of volume for the k -th tree at the s -th sample point is determined using one subsample point along the stem as:

$$\hat{v}_{ks} = \rho(h_s)H_k \quad (6)$$

where $\rho(h_s)$ is the cross-sectional area calculated from (1) at a random height along the bole h_s drawn from the uniform attribute density $[0, H_k]$. This method is design-unbiased, but approaches unbiasedness asymptotically such that it is quite possible to accumulate a slight bias in the simulations, purely from not drawing a full range of heights in $[0, H_k]$ in the finite sample size of the simulations. Drawing more than one sampled height per tree will alleviate this as will other more complex Monte Carlo methods such as importance sampling ([33] p. 39).

The third sampling method is critical height sampling (CHS) [14,34,35]. Like the previous two areal methods, CHS shares its tree selection mechanism with HPS. In its most general form, however, CHS uses the diameter at the base (or merchantable stump) of the tree, rather than DBH, to determine the size of $\mathcal{I}_k$. In this case, CHS is design-unbiased for volume estimation. However, one can also implement CHS using a (normally) smaller inclusion zone for selection that derives from DBH just as the previous two methods. In this case, the volume below DBH is treated as a cylinder, so that the volume of the butt swell is lost. This introduces a small bias into the estimation of volume, but allows the exact same ‘footprint’ in terms of $\mathcal{I}_k$ and associated area, a_k , to be used for each method. Both alternatives are supported in the sampSurf package, but the latter will be used here.

Critical height sampling is also a form of CMC sampling. However, with CHS, the selection of a height—the critical height, h_{ks}^{chs} , for the k -th tree on sample point s —is spatially structured. It is found by determining the point along the tree stem where the angle gauge shows the tree to be exactly borderline. This produces a radially-symmetric parabolic-shaped sampling surface where volume estimates are low near the edge of the inclusion zone and increase as distance from the sample point to the stem decreases radially. Tree volume estimation [36] follows in a somewhat analogous manner to (6) where:

$$\hat{v}_{ks} = b_k h_{ks}^{\text{chs}} \quad (7)$$

Here, volume is calculated from the product of tree basal area and the spatially-structured sample height, determined randomly from the distance of the sample point to the tree. The spatial structure of this method also assures that the full range of heights $[0, H_k]$ is always sampled in a full enumeration of the sampling surface within $\mathcal{I}_k$ (within the resolution of the sampling raster). In contrast, CMC sampling is spatially unstructured: regardless of the juxtaposition of sample point relative to the tree, CMC simply draws a random height, which can be anywhere on the tree stem at any point within $\mathcal{I}_k$. This produces a spatially unstructured sampling surface that appears completely random with the potential for a small amount of bias as noted above. In both CHS and CMC, the volume estimate $\hat{v}_{ks}$ is substituted for z_k in $\hat{v}_{z_k}$ using the inclusion probability for HPS to arrive at the sample estimate for the k -th tree on the s -th sample point.

Alternatively, note that one can also formulate the volume estimator (6) for CMC in terms of drawing a random cross-section, $\rho_k \sim \text{Uniform}(0, b_k)$, and sampling the height at that cross-section, h_{ks}^{cmc} , yielding $\hat{v}_{ks} = b_k h_{ks}^{\text{cmc}}$, which makes it exactly analogous to (7). The estimate of volume using this method, when combined with HPS, excludes the volume outside a cylinder from the ground (or stump) to breast height as in CHS with reference height at DBH [36,37]; drawing $\rho_k \sim \text{Uniform}(0, \rho(0))$ alleviates this problem. However, the approach in (6) is used in sampSurf [33] and has been adopted here.

Finally, fixed-area circular plot (CPS) sampling ([19] p. 276) is the last areal sampling method used in the comparisons. The inclusion probability for all trees is $\pi_k = \frac{a_k}{|\mathcal{A}|}$, where $a_k = \pi R^2$ is constant and is equal to the plot size (where R is the plot radius in m), as this method samples with probability proportional to the tree frequency. Tree volume is known in the simulations from the integration of (1) as in HPS.

Table 1 presents the parameters used in the simulations and summarizes the sampling surface results for each. Note that the small amount of ‘bias’ that accrues for HPS, CMC and CPS is due solely to the resolution of the raster. As the resolution gets small ($\Delta xy \rightarrow 0$), this bias goes to zero. The exception is CHS, where the small (approximately 1%) bias is due to treating the stem below DBH as a cylinder as noted previously. The variance is given in Table 1 in terms of the standard deviation, and the surface maximum is noted to be different for each method and sampling effort. The sampling effort can be categorized by the BAF or plot radius for the respective methods. However, it is also reported as the effective sample size, $m_e\%$, which is a measure of the percentage of total points (raster cell centers) in the scene (including the buffer area) that are overlapped by at least one inclusion zone, thus providing a non-zero estimate for volume at the respective point (e.g., [36,38]). It is a measure of the total sampled ‘effort’ in each simulation and decreases with plot size or as the BAF increases.

Table 1. Results of the sampling surface (sampSurf) simulations for the areal sampling protocols.

Sampling Method	$\mathcal{F}$ or R	$\hat{v}_z$ (m ³)	Bias (%)	St. Dev. (m ³)	Surface Max (m ³)	m_e (%)	Average a (m ²)
HPS	3.0	214.90	0.01	157.76	725.90	79.62	216.17
HPS	5.0	214.73	−0.07	177.95	886.94	73.12	129.70
HPS	7.0	214.78	−0.05	194.19	996.01	68.33	92.64
HPS	9.0	214.88	−0.00	208.78	1115.00	64.00	72.06
CHS	3.0	212.53	−1.11	163.29	670.13	79.62	216.17
CHS	5.0	212.50	−1.12	184.47	837.39	73.12	129.70
CHS	7.0	212.50	−1.12	202.88	931.64	68.33	92.64
CHS	9.0	212.54	−1.10	220.06	1068.01	64.00	72.06
CMC	3.0	214.06	−0.38	182.95	1050.41	79.62	216.17
CMC	5.0	214.74	−0.06	214.11	1476.05	73.12	129.70
CMC	7.0	215.87	0.46	241.45	1686.63	68.33	92.64
CMC	9.0	214.80	−0.04	264.82	1931.63	64.00	72.06
CPS	8.92	214.87	−0.00	169.97	873.29	80.68	250.00
CPS	6.89	214.56	−0.15	196.03	1332.61	75.26	149.25
CPS	5.64	214.45	−0.20	222.11	1725.91	70.78	100.00
CPS	5.05	214.60	−0.13	239.79	2117.20	67.45	80.00

Note: the second column lists the BAF (m² ha^{−1}) for methods based on HPS and plot radius (m) for CPS.

The surface standard deviation (St. Dev.) is given rather than the surface variance.

Figure 1 presents the sampling surfaces for the four areal sampling methods corresponding to the first rows for each method in Table 1 (note that all perspective images have been scaled relative to surface height to facilitate viewing). Recall that the variance for each method is the variability of the individual point estimates about the plane of the mean for each surface [24,25]. The surfaces themselves, when regarded devoid of the summary computations, can be deceiving even with some degree of study. The only clear interpretation is that CMC is certainly the most variable surface. However, it is difficult to distinguish between HPS and CPS with regard to variability, because of the number of trees and respective inclusion zones that are heaped within the tract. Furthermore, CHS appears to be the most gently variable surface and could justify the interpretation of the smallest variance from a purely visual perspective. One must keep in mind when viewing these surfaces that though they appear to be all scaled the same, the maxima are different in each case (Table 1). Because these surfaces are so complex, a simple set of two-tree intersecting sampling surfaces is presented in the Supplementary Material as a more intuitive guide to their interpretation (Figures S1 and S2). In addition, a raster version of Figure 1 is also presented in Supplementary Material Figure S3. Other surfaces from Table 1 with different BAFs and plot radii appear similar, but with a smaller overall footprint (e.g., $m_e\%$).

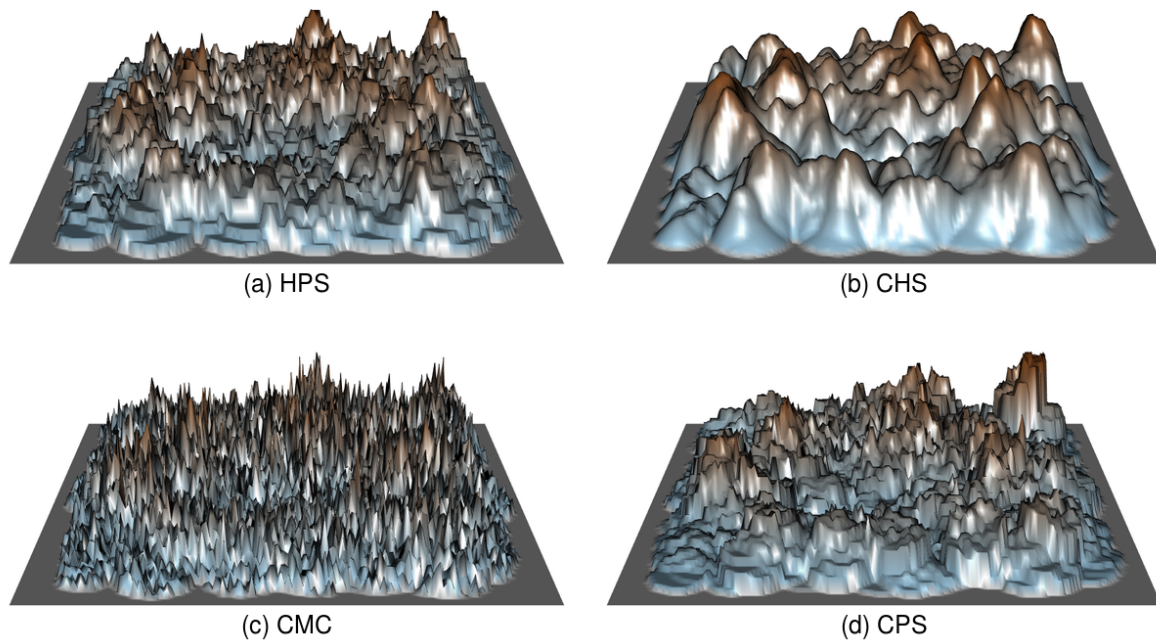


Figure 1. Perspective view of the sampling surfaces from the simulations in Table 1 illustrating the four sampling methods: (a) HPS; (b) CHS; (c) CMC, all with $\mathcal{F} = 3 \text{ m}^2 \text{ ha}^{-1}$; and (d) CPS with $R = 8.92 \text{ m}$.

2.3. Smith Plots

As previously noted, perhaps the simplest and most employed method for discriminating between different unbiased sampling methods is the comparison of variances, such that the method with the smallest variance is often deemed the best: *viz.*, most efficient. However, as mentioned earlier, one must take into consideration other factors such as cost (or its surrogate, sampling time) in order to attempt to compensate for methods that may provide lower variance, but at the expense of higher sampling effort. One way to avoid this issue is to attempt to equalize the average inclusion zone area over the methods under comparison by choosing design parameters (the BAF, plot size, etc.) for each sampling method that yield inclusion zones that are, on average, approximately the same size. While this method is useful and has been applied in several studies (e.g., [36,38]), it can be difficult to find design parameters that yield approximate conformity, especially under PPS designs that often have quite different inclusion zone shapes such as perpendicular distance sampling (PDS) [39], point relascope sampling (PRS) [40] and sausage sampling [41] for downed coarse woody debris. Indeed, when discussing this issue in the comparison of different PDS protocols, Gove et al. [38] noted that “the goal of equality is elusive, however, and it is only possible to approximate as the inclusion zones scale differently” under the differing protocols; this caveat also applies to the many other sampling methods used on standing trees and downed coarse woody debris, and a more general example might be the comparison of line-based methods to those already mentioned.

The set of simulations in Table 1 presents an example where it is rather simple to approximately equalize the sampling effort in terms of average inclusion zone area so that direct comparisons are possible. However, even this is unnecessary, because a simple graphical method based on a power law is available to compare any set of different sampling methods or protocols for a given population of standing trees or downed logs without the effort of adjusting design parameters beforehand. The method was characterized as an ‘empirical law’ for the variance when first described by Smith [4], who noted the log-log relationship between variance and plot size within experimental block designs. The general variance function is:

$$\text{Var}(\bar{a}) = V\bar{a}^{-\beta} \quad (8)$$

where $\text{Var}(\bar{a})$ is the variance in volume among plots of size $\bar{a}$, while V , a scaled variance whose definition depends on that of plot size, and $0 \leq \beta \leq 1$ the slope parameter, are both constants that can be estimated if desired. It is interesting to note that Deming [42] (p. 206), while mentioning [4], cites a set of unpublished studies by Mahalanobis evidently culminating in Mahalanobis [43] and extended in Mahalanobis [44] as the principle contribution to the variance function (8).

In the current application, interest lies simply in the general form of this relationship for graphical discernment in the comparison of different methods, rather than in parameter estimation per se. Clearly, estimation of (8) is unnecessary for comparison of the simulation results, which can be presented either in raw units or on a log-log scale, whichever is the most convenient. Using this relationship, each of the different sample methods under consideration are simulated several times on the same population using varying design parameters to give a range of average inclusion zone areas as in Table 1. A simple plot of either the raw variances by average inclusion zone area or their log-log counterpart suffices for the comparison, obviating the requirement for seeking the rather ‘elusive’ goal of equalized sampling effort. The results of applying this method to the simulations in Table 1 are presented in Figure 2.

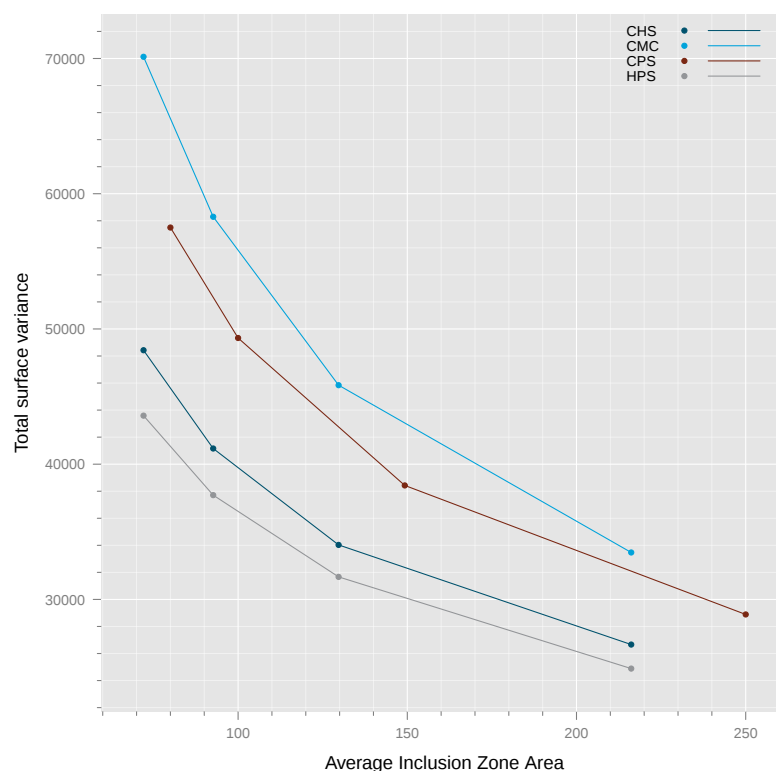


Figure 2. Smith plot of the simulations in Table 1.

The ‘Smith plot’ clearly shows the ranking of the different methods and would work equally as well were the inclusion zone areas not approximately commensurate as they are in this example. Clearly, and in accordance with theory, HPS is the most efficient method for estimating volume, followed by CHS, with CPS and CMC being the least efficient in this particular population.

This ‘law’ has found uses both in forestry and other areas of agriculture such as agronomy since its inception. A number of early studies has looked at the estimation of the slope parameter, β , in connection with the optimal plot size question. For example, Koch and Rigney [45] and Hatheway and Williams [46] studied the estimation of the slope parameter from designed experiments for determining soil heterogeneity with regard to optimal plot size. These authors have suggested a weighted approach to the variance in estimating (8) due to the correlation between neighboring plot

measurements in the dependent variable. Similarly, Swallow and Wehner [47] extended these results by proposing a generalized least squares approach to the estimation of β and its connection to optimal plot size in yields of cucumber. In a related study, Whittle [48] rigorously explored the relationships between the variance function (8) and the associated spatial covariance function. In a simulation study, Pearce [49] found that Smith's law holds over a range of plot areas that would be found in practice, though correlation did affect the relationship over larger plot areas.

Smith's [4] relationship (8) has been recognized as a useful tool in forestry where it is sometimes alternatively represented in terms of the coefficient of variation (CV) [50], rather than the variance. O'Regan and Palley [51] used Freese's [50] relationship to graphically compare HPS and CPS using the average number of trees sampled at a location rather than $\bar{a}$, much as in Figure 2. A more extensive simulation study in Arvanitis and O'Regan [52] found that adding a quadratic term to the logarithmic version of equation (8) sometimes provided a better fit than equation (8) alone, where plot size or $\mathcal{F}^{-1}$ were used for CPS and HPS, respectively. Graphically, these authors used the variance versus plot size for CPS comparisons, while CV versus average number of trees sampled at a location was used to compare HPS and CPS as in O'Regan and Palley [51] for both static and growth components. Wensel and John [53] fitted (8) using $\mathcal{F}$ for HPS, which changes the direction of the slope, but presented a simple relationship to convert the results back in terms of $\bar{a}$ for subsequent graphical comparison with CPS using the logarithm of the variance.

Freese [50] also noted that spatial pattern and plot shape can play a role in the shape of the relationship between plot size and variability. Reich and Arvanitis [54], using Freese's [50] relationship with stem density, demonstrated that the curve shape (β) can change depending on the spatial pattern. It will also change for different stand attributes of interest in simulations.

More recently, Smith's [4] variance function has been employed by Lynch [9] to determine optimal plot size (or BAF) using Lagrange multipliers in the objective function to minimize the variance subject to total fixed costs. In a companion study, Lynch [10] provided optimal cost-plus-loss solutions under an expected absolute value loss function and the variance function of Smith [4], using the cost functions developed by Zeide [7] and Gambill et al. [8] for CPS and HPS, respectively. Yang et al. [55], using the same optimization approach of Lynch [9], developed a cost function that was specific to the 'Big BAF' [56] sampling scheme. These authors used a CV versus $\mathcal{F}^{-1}$ representation for the variance relationship. These recent papers also present a more thorough coverage of the forestry literature on the Smith [4] and related Freese [50] approaches and should be consulted for more information.

The applications cited above speak to the versatility of Smith's [4] empirical law having been applied in both field (including experimental agronomy) and simulation studies, over a reasonably wide range of plant population forms; not to mention ecology and population biology where it is known as 'Taylor's law' (e.g., [57]). In the case of simulation, where not only the same population is used, but the same sampling grid, correlation will also be present in the dependent variable as a consequence of sample points being aligned with the grid centers. While it is perfectly acceptable to fit the relationship (8) (or similar) for quantitative comparison of, e.g., the slope parameter for different sampling methods or protocols, the graphical comparison of simulation results, not estimation, is emphasized here; therefore, such considerations based on estimation are unnecessary for this straightforward application, though correlation of a different sort is discussed in the following section.

2.4. Taylor Diagrams

The Smith plots are useful for judging the overall relationship between sampling methods based on their variances (standard deviations or CVs). However useful this may be, it says nothing about the structural or spatial correspondence between the techniques over the spatial field or tract. A measure of this correspondence is the simple pattern covariance or correlation between two simulations. In this sense, 'pattern' is used to emphasize that the statistics are computed between the two simulated raster representations over each grid cell, just as the variance, resulting in an evaluation of the strength of spatial correspondence between two sampling methods. Both the covariance and correlation are useful

in this sense as they can be related to the variances by defining one particular sampling method as the standard, or reference, to which the others are compared. The simple relationships that allow for this comparison are given through the variance of the difference:

$$\text{Var}(x - y) = \sigma_x^2 + \sigma_y^2 - 2\text{Cov}(x, y) \quad (9)$$

$$= \sigma_x^2 + \sigma_y^2 - 2\sigma_x\sigma_y\text{Corr}(x, y) \quad (10)$$

for two random variables x and y with corresponding variances σ_x^2 and σ_y^2 . In this application, x and y correspond to two matrix simulation results, where x will be taken as the reference method. Since HPS with $\mathcal{F} = 3 \text{ m}^2 \text{ ha}^{-1}$ (HPSbaf3) has the smallest variance in Figure 2, this sampling method will be taken as the standard to be compared against.

Taylor [58] developed a diagram that incorporates all of the information in either (9) or (10) into a type of polar plot by recognizing the similarity of these relationships to the law of cosines:

$$c^2 = a^2 + b^2 - 2ab \cos \phi \quad (11)$$

where the triangular sides are a , b and c with opposite angle ϕ . Equating (10) to (11) gives: $\text{Var}(x - y) \equiv c^2$, $\sigma_x^2 \equiv a^2$, $\sigma_y^2 \equiv b^2$ and, notably, $\mathcal{R} \equiv \cos \phi$, the correlation coefficient. Taylor's [58] insight makes it possible to compare all four of these statistics simultaneously in a simple visual diagram. This lessens the necessity for lengthy tabular summaries and facilitates the visual grouping of similar sampling methods as compared to the reference based on all four metrics. It should be noted that Taylor's [58] original application of this procedure was the analysis of different simulation model runs against observations (our reference method), and others have applied the method similarly [59,60]. As such, his presentation is in terms of centered mean squared error, which corresponds to our variance of the difference.

A Taylor diagram for the simulations in Table 1 is presented in Figure 3 (see Table S1 for the requisite statistics). The first point to note is that, due to the polar nature of the diagram, the standard deviations are displayed with respect to the circular arcs joining the two axes. To facilitate comparison, an extra arc is displayed aligning with the reference method, HPSbaf3. Thus, for example, critical height sampling with $\mathcal{F} = 3$ (CHSbaf3) is found to be radially closest to the arc of the reference method. In addition, the radial rays emanating from the origin demarcate the correlation, $\mathcal{R}$, which increases clockwise from the y -axis. Again, CHSbaf3 is quickly located as the method with the highest pattern correlation to HPSbaf3. Finally, semicircular arcs centered on the reference standard deviation denote the distance in terms of $\sqrt{\text{Var}(x - y)}$. Thus, once more, CHSbaf3 is the closest in regard to minimizing the variance of the difference with the standard.

While the analysis with respect to CHSbaf3 may be trivial, the diagram in Figure 3 facilitates comparison of other methods that would be more difficult to glean quickly from Table S1. For example, the clusters of CPSrad9, HPSbaf5 and CMCbaf3 are all close in relation to each other. Each has approximately the same pattern correlation with HPSbaf3 of approximately $\mathcal{R} = 0.87$, with CPSrad9 having the smallest standard deviation of the three, as well as the smallest variance difference with the reference method. The result may be somewhat surprising given the difference in the average inclusion zone areas between HPSbaf5 and CPSrad9 compared to the reference method (Table 1); e.g., the HPSbaf5 surface, even with smaller $\bar{a}$ for this method, correlates well with HPSbaf3. On the other hand, CPS evidently requires a larger plot size ($\bar{a}$) to match the efficiency of HPS in terms of variance, which is also discernible from the Smith plot. Visually connecting a line through all of the CPS methods on the graph shows that they get increasingly further from the standard deviation of the reference method (denoted by the arc through that point on the x -axis) as the plot radius or inclusion zone area decreases. This relationship is also discernible on the Smith plot (Figure 2) by regarding the vertical distance from the reference method to the CPS curve as plot size decreases; however, this relationship appears to be more nuanced in that figure, and it does not show the corresponding

decrease in correlations as does the Taylor diagram. Visually adding a line connecting each of the individual method sets shows the same tendency to be further from the reference method as $\bar{a}$ decreases. Not surprisingly, CMCbaf9 has the lowest correlation and largest variance difference with HPSbaf3, coupled with the highest overall variance, which agrees with Figure 2. Other clusters of sampling methods can also be discerned from the diagram, which may not be as easily detectable from Table S1; further exploration of the relationships in Figure 3 is left to the interested reader.

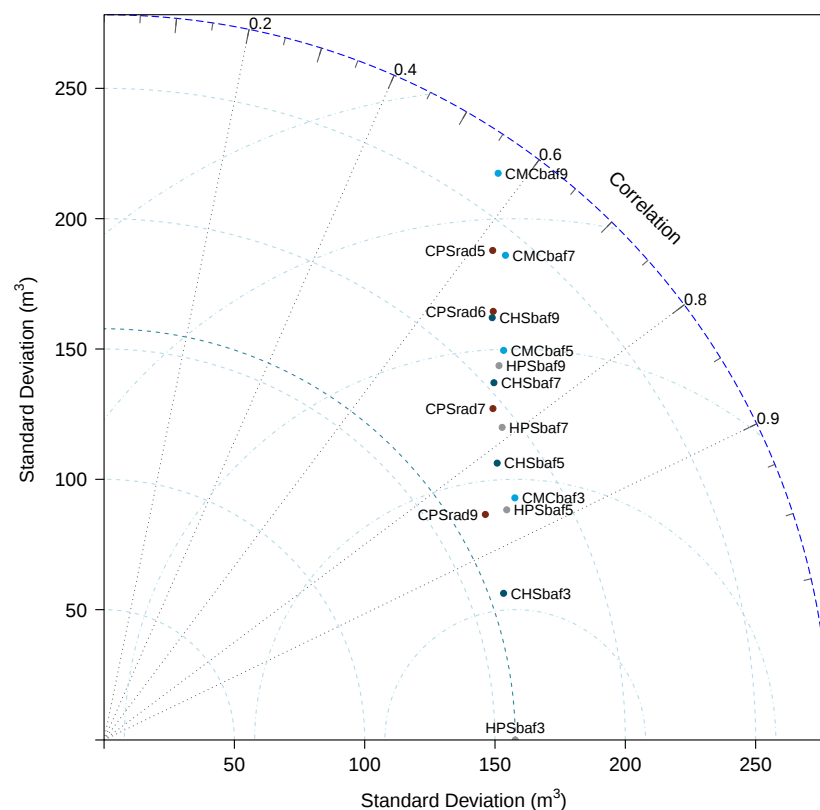


Figure 3. Taylor diagram of the simulations in Table 1, where the labels follow the convention of the standard method ‘HPSbaf3,’ which corresponds to horizontal point sampling with BAF $\mathcal{F} = 3 \text{ m}^2 \text{ ha}^{-1}$.

2.5. Wavelet Variance Decomposition

The diagnostics presented thus far for comparing sampling methods are all applied over the full simulation raster without regard to scale. However, it is possible to decompose the resultant statistics—the variance, covariance, correlation—by scale in terms of distance. In what follows, a two-dimensional maximal overlap discrete wavelet transform (MODWT) ([61] p. 159) is applied to the simulation raster results to perform this scale-based decomposition. The wavelet filter, when applied to the raster simulations, decomposes the image into wavelet coefficients at each different scale corresponding to each different level of the decomposition. The wavelet decomposition produces a new set of anisotropic images at each scale, the sum of which is an overall isotropic raster. A scaling filter is also applied resulting in a raster of scaling or smooth coefficients, whose mean is the sample mean for the simulation at the highest level (coarsest scale). Often, wavelet analysis concerns only these different sets of raw wavelet and scaling coefficients from the MODWT decomposition. However, this same set of wavelet images at the different levels can be easily transformed into wavelet variances, whose sum is the total sample variance in the overall simulation result for each sampling protocol. Likewise, the raw wavelet coefficients from image (simulation) pairs can be used to calculate covariance or correlation results at each scale. In what follows, a short review of the wavelet and scaling filters is

presented, with their relation to the wavelet variance decomposition. Subsequently, these results are applied to the simulations and interpreted using the diagnostic plots from the previous sections.

2.5.1. MODWT Basics

In general, to apply the two-dimensional wavelet decomposition, let $\mathbf{X}$ denote an N row by M column raster image indexed by (u, v) as $\mathbf{X}_{u,v}$ [62,63], corresponding to our individual raster simulation results from Table 1. The MODWT is applied to $\mathbf{X}$ at levels $j = 1, \dots, J$, with corresponding standardized scale $\tau_j = 2^{j-1}$ for the j -th-level wavelet filter and $2\tau_j = 2^j$ for the scaling filter. The τ_j are unitless (in terms of distance in pixels), but can be represented on a physical basis as $\tau'_j = \tau_j \Delta xy$, where Δxy m is the resolution of the square grid cells; as noted above, $\Delta xy = 1$ m in the simulations.

A simple Haar wavelet of length $L = 2$ is used in the analysis presented here. The wavelet filter coefficient vector, $\mathbf{h}$, for level $j = 1$ of the decomposition has coefficients $\mathbf{h}_1 = (h_{1,0}, h_{1,1})' = (-1/\sqrt{2}, 1/\sqrt{2})'$. The corresponding scaling filter coefficient vector is $\mathbf{g}_1 = (g_{1,0}, g_{1,1})' = (1/\sqrt{2}, 1/\sqrt{2})'$. The Haar filters have $L_j = (2^j - 1)(L - 1) + 1$ coefficients for the j -th level. For example, if $j = 2$, then the Haar filter has $L_j = 4$ coefficients. Therefore, the physical scale for the corresponding wavelet filter is $\tau'_2 = 2$ m. This scale gives the width of the wavelet filter corresponding to the change in distance because of the sign change in the coefficients (two negative and two positive). The scaling filter has the same number of coefficients, L_j , but the scale is $2\tau'_2 = 4$ m, because all coefficients are positive in this filter; it is a smoothing filter.

The two-dimensional Haar filters obtained from the wavelet and scaling vectors at level $j = 1$ are given by the outer (or tensor) product of the vectors as [64]:

$$\begin{aligned} \mathbf{h}\mathbf{h}' &= \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} & \mathbf{h}\mathbf{g}' &= \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \\ \mathbf{g}\mathbf{h}' &= \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} & \mathbf{g}\mathbf{g}' &= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \end{aligned} \quad (12)$$

Conceptually, these two-dimensional filters are convolved with $\mathbf{X}$ as a moving window through the scene, row-by-row, across columns. Alternatively, these matrices may be padded with zeros to be conformable with $\mathbf{X}$ [65,66]. In the moving-window approach, $\mathbf{h}\mathbf{h}'$ contrasts the diagonal elements in $\mathbf{X}$. The matrix $\mathbf{h}\mathbf{g}'$ contrasts rows of $\mathbf{X}$, producing vertical differences. Likewise, $\mathbf{g}\mathbf{h}'$ contrasts columns of the image and yielding the horizontal components of variation. Lastly, $\mathbf{g}\mathbf{g}'$ produces a matrix of scaling coefficients that corresponds to a mean smooth decomposition. Note that the wavelet matrices now correspond to changes in area rather than distance. The change in area associated with the wavelet filters is at the physical scale of $\frac{1}{2}(2\tau'_j \times 2\tau'_j) = 2\tau_j'^2$ for given level j [63]. For example, in the case of $j = 1$ in (12) above, there are $L_1^2 = 4$ coefficients, corresponding to a physical area of 4 m^2 , with change in area corresponding to 2 m^2 (two negative and two positive coefficients) in the wavelet matrices. Similarly, if $j = 2$, then $L_2^2 = 16$ (yielding 4×4 filter matrices), with $\tau'_2 = 2$ m corresponding to a change in area of 8 m^2 . In the examples that follow, the physical scale is denoted in terms of distance rather than area difference as the former is simpler and conversion can be made directly using the above formula (see also Table S2).

The matrix approach is both conceptually simple to understand and to implement; however, in application, the matrix $\mathbf{X}$ is filtered using the pyramid algorithm. The pyramid algorithm treats rows and columns as univariate time series and filters these individually in a specific manner using the wavelet filter coefficients $\mathbf{h}_1$ and scaling filter coefficients $\mathbf{g}_1$ (for details, see [63,64]). In practice, the pyramid algorithm is more complex to implement, but is computationally more efficient than the matrix approach.

In the current application, each individual simulation image $\mathbf{X}$ in Table 1 is decomposed at each level j by the wavelet filter into matrices conformable with $\mathbf{X}$ corresponding to horizontal ($\mathbf{H}_j$), vertical ($\mathbf{V}_j$) and diagonal ($\mathbf{D}_j$) wavelet coefficients; likewise, applying the scaling filter results in a scaling (smooth) matrix ($\mathbf{Z}_J$) that is retained only at level J . These correspond to the application of the two-dimensional wavelet filter matrices given above, though, again, in practice, they are computed using the pyramid algorithm. The anisotropic matrices can be summed to produce an overall isotropic decomposition as will be discussed in more detail in the variances section. This is all accomplished using the R *ssWavelets* package [67], which applies the pyramid algorithm through the methods in the R *waveslim* package [68].

Just as in forest sampling, wavelet applications require a boundary correction method in application. Several boundary correction approaches are discussed in detail by Addison [69] (Section 2.15). The *waveslim* package uses periodization, which is sometimes referred to as toroidal correction [70] (p. 152 and 166). Toroidal correction can be conceptualized as a translation of the image about all sides so that when the wavelet filter reaches the end of a row or column, it wraps back around to the beginning of the image. Another method, reflection (folding about each side to create a mirror image), has been advocated by both Lark and Webster [64] and Ma and Zhang [65], who cited potential problems with periodization in their applications. However, their applications were quite different, and the concerns they raise with periodization do not appear to be relevant in the current areal sampling application; indeed, it can be argued that periodization is superior in the current application (e.g., see the discussion in [71]) and is therefore used in this study.

2.5.2. The Two-Dimensional Wavelet Variance and Covariance

Two-dimensional wavelet variance decomposition theory and applications include Mondal and Percival [62], Geilhufe et al. [63] and Lark and Webster [64], the latter of which also includes some information on wavelet covariances when comparing two images (or, in our case, two sampling methods). A recent study using these methods is found in Ma and Zhang [65]. In these applications in wavelet analysis, one encounters two main classes of variance. The first is called the wavelet variance and only involves the wavelet components of the decomposition (no scaling components are involved). The wavelet variance is calculated at all scales for its contribution to the decomposition of the total surface variance in the simulations. The second is called the sample variance, which does indeed involve both wavelet and scaling components and therefore exists only at level J . The sample variance is essentially equivalent to the surface variance, $\text{Var}(\hat{v}_{zs})$, given in the simulations. The small difference is that the divisor in the latter is $m - 1$, while the wavelet sample variance is in terms of $m = NM$. The wavelet sample variance is sometimes referred to as the population variance in sampling (e.g., [72] p. 23), and the surface variance can be multiplied by $\frac{(m-1)}{m}$ to make the two equivalent. Finally, the concept of the energy in a time series ([61] p. 42, [73]) or image [62] is equivalent to what is more commonly known as the sum of squares.

The sample variance derives from the familiar relation (see, e.g., https://en.wikipedia.org/wiki/Algebraic_formula_for_the_variance):

$$\text{Var}(x) = E[X^2] - E[X]^2 \quad (13)$$

therefore, the total energy in the image is given by the sums of squares of all elements [62]:

$$\|\mathbf{X}\|^2 = \sum_{u=1}^N \sum_{v=1}^M x_{u,v}^2 \quad (14)$$

and substituting in the image decomposition matrices over all j gives:

$$= \sum_{j=1}^J \left(\|\mathbf{D}_j\|^2 + \|\mathbf{V}_j\|^2 + \|\mathbf{H}_j\|^2 \right) + \|\mathbf{Z}_J\|^2 \quad (15)$$

Note from the last relation that the wavelet transformation is energy preserving [62]. In two dimensions, the sample (population) variance is given by:

$$\hat{\sigma}_X^2 = \frac{1}{NM} \sum_{u=1}^N \sum_{v=1}^M (\mathbf{X}_{u,v} - \bar{X})^2 \quad (16)$$

which, in terms of (13) is

$$= \frac{1}{NM} \|\mathbf{X}\|^2 - \bar{X}^2 \quad (17)$$

and substituting (15) gives [62]

$$= \frac{1}{NM} \left(\sum_{j=1}^J \left(\|\mathbf{D}_j\|^2 + \|\mathbf{V}_j\|^2 + \|\mathbf{H}_j\|^2 \right) + \|\mathbf{Z}_J\|^2 \right) - \bar{X}^2 \quad (18)$$

$$= \bar{\mathcal{E}} - \bar{X}^2 \quad (19)$$

where $\bar{\mathcal{E}}$ is the total average energy. The penultimate relationship states that the surface variance can be decomposed as the mean sum of squares of the anisotropic wavelet decomposition matrices over all levels plus the squared smooth at level J , less the squared mean. Again, it should be noted that since the energy in the wavelet decomposition matrices is preserved through $\bar{\mathcal{E}}$, the sample variance can be decomposed by scale without loss of information.

The sample covariance can be decomposed in a similar manner to the variance. In this case, the idea is to compare two images, $\mathbf{X}$ and $\mathbf{Y}$, with the same raster extents ($N \times M$) and resolution Δxy , generated from different sampling methods on the same population of trees. Relation (13) generalizes to:

$$\text{Cov}(x, y) = E[X \cdot Y] - E[X]E[Y] \quad (20)$$

with total covariance energy—sums of cross products of all elements—given (loosely) as:

$$\|\mathbf{XY}\| = \sum_{u=0}^{N-1} \sum_{v=0}^{M-1} \mathbf{X}_{u,v} \mathbf{Y}_{u,v} \quad (21)$$

For both the variance and covariance, it is perfectly reasonable to look at anisotropic and isotropic components of the variance as images at any decomposition level, analogous to the raw wavelet coefficient results. For example, letting $\mathbf{X}^2$ denote the matrix of the squares of the individual components (u, v) , then $\frac{1}{NM} \mathbf{H}_j^2, j = 1, \dots, J$ yields the average energy images for the horizontal decomposition over all levels. For covariances, this would be written in terms of the horizontal decompositions from the two sampling methods via the Hadamard matrix product ([https://en.wikipedia.org/wiki/Hadamard_product_\(matrices\)](https://en.wikipedia.org/wiki/Hadamard_product_(matrices))). That is, if $\mathbf{H}_X$ and $\mathbf{H}_Y$ represent the horizontal decompositions for two sampling methods, then $\frac{1}{NM} (\mathbf{H}_X \circ \mathbf{H}_Y), j = 1, \dots, J$ are the

corresponding average energy contribution matrices to the overall covariance matrix between the two sampling methods. Marginal sums and totals can be calculated for each level in the decomposition as explained in Gove [71], and the correlation can be calculated from these quantities for use in comparing methods by, e.g., the Taylor diagram.

2.5.3. Distance-Based Smith Plots

A MODWT wavelet decomposition to level $J = 6$ was applied to each of the simulation results in Table 1. The isotropic wavelet variances were calculated for each level of the decomposition using the methods described above. The Smith plot discussed earlier can be used to provide a summary of the wavelet analysis for each simulation. Figure 4 presents a set of Smith plots where each panel corresponds to a level of the MODWT decomposition in terms of distance, τ'_j . These individual Smith plots, having decomposed the overall sampling surface variance by distance class, illustrate the variance relationships among sampling methods at different scales, where the same interpretation applies to these as to Figure 2.

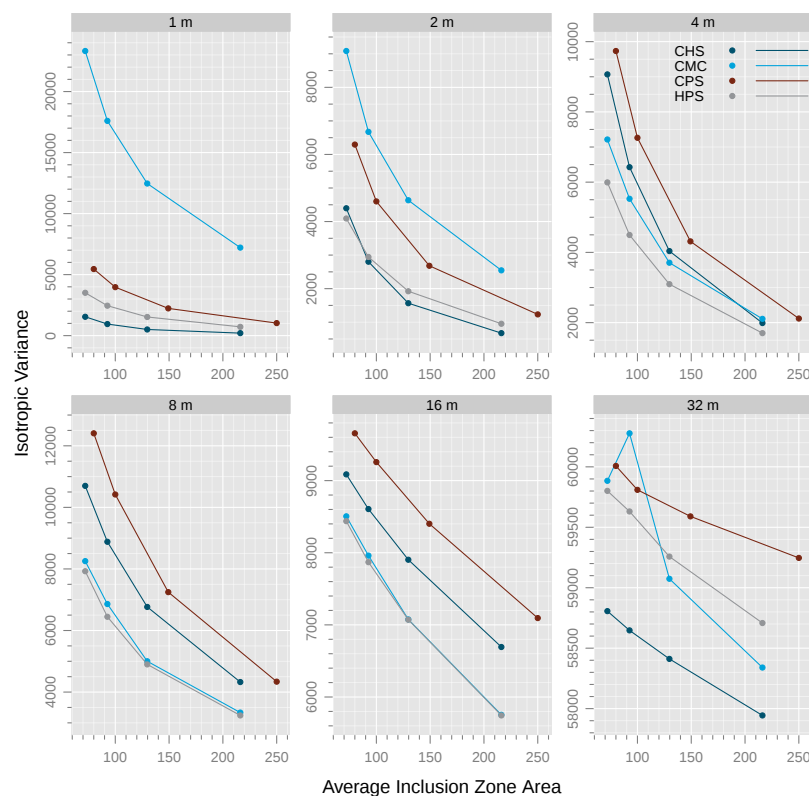


Figure 4. MODWT Smith plots of the simulations in Table 1 where the individual panels show the variance relations by wavelet decomposition distance $\tau'_j, j = 1, \dots, 6$. All panels include wavelet variances, while only the coarsest scale at 32 m contains the scaling (smooth) variance component as well.

Recall that for these simulations, HPS is the best sampling method for volume in Figure 2, which accords with theory. However, when we decompose the sampling surface at different levels, this is no longer consistently the case (Figure 4). For example, HPS has the lowest variance only at the middle ($\tau'_j = 4, 8, 16$ m) scales. At the other scales ($\tau'_j = 1, 2, 32$ m), CHS has uniformly the lowest variance. As noted earlier, for $\tau'_j = 1$ m, the wavelet (h_1) and scaling filters (g_1), as well as their associated two-dimensional anisotropic matrix versions in (12) show that that the decomposition is comparing pairs of sample points. Thus, the associated wavelet variances for each simulation are for this scale

of differences in the overall sampling surfaces. Regarding the simple two-tree examples in Figure S2, note that differences in adjoining cells are either zero or large (at the boundary of inclusion zones) under HPS because of the ‘stair-step’ manner in which the individual surfaces accumulate for each tree. Within an inclusion zone, moving from one point to another causes no change in the estimate because of the flat surface, and thus, no change in the variance at this scale. However, large jumps in variance accrue as the filter straddles adjoining points inside and outside (or vice versa) the inclusion zone, reflecting the well-known importance of checking borderline trees in practice. In contrast, under CHS, the differences are small between adjoining sample points and vary little throughout the inclusion zones depending on the taper of the associated tree stems. The gradual change within individual tree inclusion zones accumulates to associated gradual changes in the overlap zones. This results in lower variance at smaller scales for CHS compared to HPS and the other methods. While one can reason this out from the way we know that sampling works both theoretically and in practice under these two methods for one or two trees, it would be difficult to extend this thinking to a forest with many trees and overlapping inclusion zones. For example, regarding the full simulation surfaces in Figure 1, it would be difficult to justify extending such interpretations to the entire tract without the benefit of the analytic wavelet analysis. In summary, in accordance with theory, moving a point Δxy under HPS can cause a large change in variance near the inclusion zone boundaries, which is reflected in the higher variance for HPS than CHS at this lowest scale, where the same movement under CHS causes a smaller change to the variance. It should be noted that the variance at this finest scale is on the order of a tenth of the overall sampling surface variance for HPS, CHS and CPS (Figure 2), while it comprises a more substantial amount of the overall variance for the spatially random CMC method as would be expected.

CPS consistently has the highest variance of the methods, except at the two smallest scales. Again, this would be expected from theory. The high variance for CMC follows by application of the arguments given above for HPS and CHS. Since CMC is a random, spatially unstructured method, adjacent points can have large differences, especially at the one- and two-meter scales. Figure 4 seems to indicate that CMC can be close to CHS and HPS, especially at the middle scales. Note in the CMC results at the coarsest (32 m) scale that the variance actually increases for $\mathcal{F} = 7 \text{ m}^2 \text{ ha}^{-1}$. This anomaly suggests that the CMC process at this scale violates Smith’s empirical law, though only by a small amount (note the scale for variance on this panel), which is not to be unexpected from a spatially unstructured random process.

The variances presented in Figure 4 are in terms of the average energy, $\bar{\mathcal{E}}$. This includes the empirical wavelet variance (Supplementary Material, Section S.4.2) over all levels, but at the coarsest level, J , also includes the smooth component of the variance. Thus, these add to the sample variance less the squared mean over all levels for each method. A complementary plot that uses only the wavelet variances is presented in Figure S4, which illustrates the relationship between wavelet variance and distance. Note that these figures show clear maxima for HPS, CHS and CPS, and minima for CMC. Geilhufe et al. [63] provides more interpretation for these plots, including the link between the maxima and a characteristic scale [74] to the images.

2.5.4. Distance-Based Taylor Plots

Just as the Smith plots can be applied to the MODWT wavelet analysis to facilitate the interpretation of the sampling variance relationships by scale, the Taylor diagrams can be adapted in a similar manner for a different summary perspective of the results. Figure S5 presents the complete set of Taylor diagrams for the MODWT analysis corresponding to Figure 4. These diagrams have the same interpretation as that for the overall sampling surface variance in Figure 3, but for the distance-based components of the variance relationships. The following discussion will concentrate on distances $\tau'_j = 1, 8 \text{ m}$ as presented in Figure 5 as an illustration of the interpretation and potential use of these diagrams for wavelet decomposition of sampling surfaces.

The analysis of the MODWT Smith plot determined that CHS was the most efficient method on a 1-m scale; it is also possible to determine from that plot that CHSbaf3 has the lowest variance at this scale. This simulation result therefore becomes the standard method to compare against at 1 m. Perhaps the most striking result at this distance is the lack of correlation between the standard and the other sampling methods. The closest in terms of pattern correlation (and difference standard deviation) is CHSbaf5 with $\mathcal{R} = 0.65$. Correlations decrease and difference variances increase in a somewhat predictable manner for the other methods. Note that those methods with similar inclusion zone coverage area and those most closely related appear to be the closest to the standard in terms of $\mathcal{R}$ and $\sqrt{\text{Var}(x - y)}$, with preference to other CHS variants first, followed by HPS and CPS. Furthermore, similar to the Smith plots, the CMC sampling methods cannot be considered to share attributes similar to CHS for reasons already mentioned. Even with the same overall inclusion zone area, CMCbaf3 is a poor substitute compared to the best method at the 1-m scale. Again, the difference lies in the spatially unstructured surface of CMC methods versus the predictably spatially structured surface under CHS.

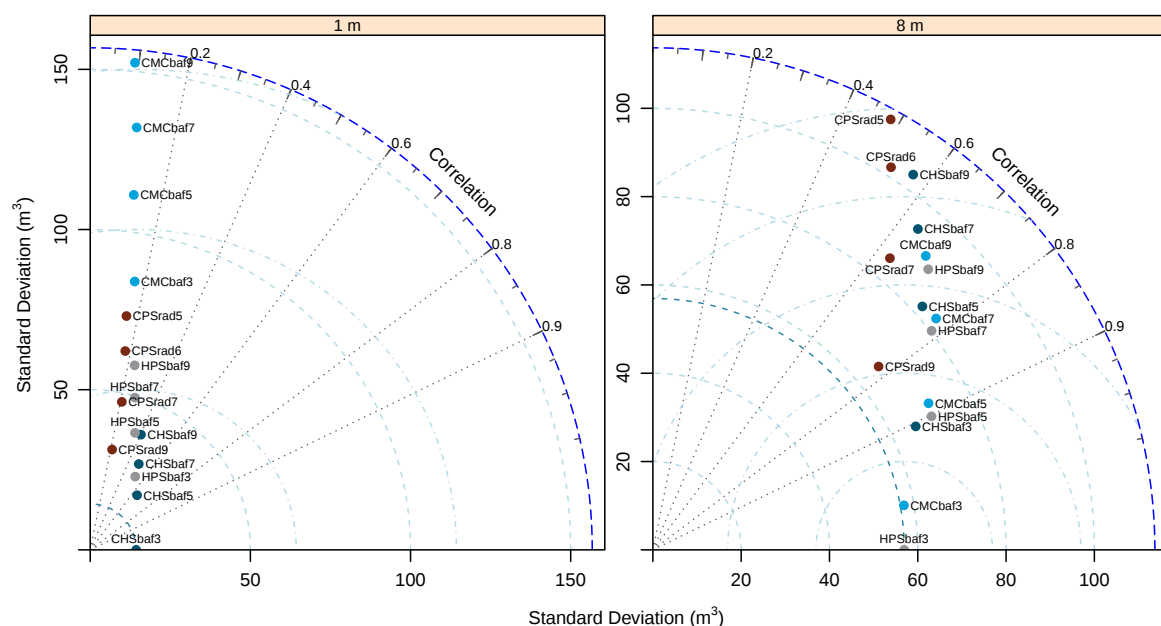


Figure 5. MODWT Taylor diagrams for the simulations in Table 1 where the individual panels represent relationship (10) for wavelet decomposition distances τ_j' , $j = 1, 4$. These panels include wavelet variances only (in terms of their standard deviations).

Note particularly in Figure S5 that each distance decomposition scale for the Taylor diagrams has a different reference (standard) sampling method, which can be discerned from the Smith plots as in the 1-m case. At the 8-m scale, HPSbaf3 becomes the reference method (Figure 5). Perhaps the most salient relationship at this scale is that HPSbaf3 and CMCbaf3 are highly correlated, with pattern correlation $\mathcal{R} = 0.98$ and similar standard deviations. Recall that at a scale of $\tau_j' = 8$ m, the wavelet filters are contrasting the changes in area corresponding to 128 m^2 (Table S2) out of a block twice that size. Therefore, the average difference between the blocks tends to be similar for the two sampling methods as the filters traverse the image. This point will be revisited in the Discussion. The groups of CHSbaf3, HPSbaf5, and CMCbaf5 are also quite highly correlated with HPSbaf3 with $\mathcal{R} \approx 0.9$. The standard deviation for CPSrad9 is close to that for the reference method, but the correlation is less than the aforementioned group, yielding a larger overall difference standard deviation, due at least partially perhaps to the difference in selection probability mechanism between it and HPSbaf3 and the mismatch between inclusion zone coverage.

In general, the Taylor diagrams in Figure S5 allow us to see the progression from small-through large-scale variability while accounting for pattern correlation and difference standard deviation. The smallest scales, 1 and 2 m, show similar patterns that are interpretable with regard to our understanding of the sampling methods themselves and their associated sampling surfaces. For example, CHSbaf3 has the lowest variance for both 1-m and 2-m scales. This makes sense because, as noted on the Smith plot, if we were to move a sample point one meter in any direction, the estimate would change little. However, this same mechanism works against the other sampling methods with different sampling effort, since there can be large jumps between estimates at overlap boundaries within a given method and, thus, low pattern correlation in comparison to CHSbaf3; CMC is the outlier, with essentially no spatial structure. The correlations increase and the difference standard deviations decrease at the 2-m scale. This could be due to the fact that though the scale has increased by only one meter, the area difference comparison is 8 m^2 out of a total block of 16 m^2 ; this evidently results in estimates encompassed by this larger area of sample points being less correlated for CHSbaf3 (note the increase (almost double) in the standard deviation for CHSbaf3 between the 1- and 2-m diagrams). At the same time, the other sampling methods are becoming more correlated with the reference method at this scale. At 4 m, the results are very similar to those of 8 m discussed above, though the pattern correlations with the reference, HPSbaf3, are less and the difference standard deviations larger. The two largest scales point towards a trend of all methods converging as distance increases between sample points.

3. Discussion

The intent of this paper was to present a few tools that could be used in future studies targeted at comparing different areal sampling methods. The Smith plot, as noted in that section, has been used a number of times previously in various forms to help guide the selection of the ‘best’ sampling method. Here, we have extended its use by straightforward application to the scales in a MODWT wavelet analysis. Taylor diagrams do not seem to have been used for either data-model comparison (their original intent) or for sampling design comparisons in forestry (though, see Verbeeck et al. [75] for an eddy covariance application); therefore, this latter application, as well as the extension to wavelet decomposition levels appears to be new. Both Smith plots and Taylor diagrams can be applied to results from either field studies or simulations; they do not require raster data for application. However, wavelet analysis in the form of MODWT does require raster (image) data, and it therefore follows that the Smith and Taylor applications to wavelet results do as well. The wavelet analysis was able to tease out some scale-based observations in an analytical manner that do correspond to our intuition about the sampling methods studied. In this sense, this methodology may be helpful in assessing sampling methods that might be challenging to understand at different scales without it. Perhaps though, a bit more background on the application of wavelets to the sampling surfaces presented here would be helpful in understanding these tools further.

A small, but reasonable overview of information about wavelets has been presented, along with some examples applied to sampling surfaces. Wavelets are a large and evolving subject, and here, we have used one of the most basic decompositions, the MODWT, to illustrate some of the information that can be gleaned from an analysis. Much more is possible. Perhaps though, a little more in the way of illustration would be helpful with regard to what wavelets are doing at different levels (j) and corresponding scales (τ'_j). Recall again that at level $j = 1$, there are $L_1 = 2$ wavelet filter coefficients and also scaling filter coefficients. The wavelet scale is thus $\tau'_1 = 1 \text{ m}$ because there is one meter between each of the sample points when the difference matrices given in (12) are applied to each sample point. The total area covered by the wavelet filter is 4 m^2 , while the area differenced is 2 m^2 (Table S2). The scaling filter is simpler: it is an average of the cells in the 4-m^2 coverage area and thus has scale $2\tau'_1$. The results from the decomposition at any level are new anisotropic images in terms of ‘raw’ wavelet coefficients and a smooth image of scaling coefficients (these should not be confused with the wavelet and scaling filter coefficients). The sum of the anisotropic wavelet images yields

an isotropic image. The smooth image is only retained at level J . The mean square of these images on a pixel-by-pixel basis yields the average energy image, and the covariance is similarly computed between two different images via the relevant formulae given previously.

The above is a simplified review of the process for a given level. Now, as the level of the filters increases by one, so too does the scale increase by the dyadic sequence previously given. The filters themselves also expand dyadically (termed dilation), and similarly, the matrices in (12) expand as the square of the number of filter coefficients, L_j . The area covered by the filters equals $L_j^2 \Delta xy^2$, and the area difference contrasted in this coverage is half this amount. Therefore, as the level or scale of the decomposition increases, the area encompassed in the contrast of the wavelet filter (or averaged in the scaling filter) matrices for each pixel in the image increases as the filter is applied to each pixel in the scene. Boundary pixels are handled by periodization, and ‘wrap-around’ (or the toroidal effect) can be discerned at the edges of the image as the filter scale expands.

The rather concise presentation given of the mathematics behind the wavelet transform for the MODWT application (see also the references given for more details), coupled with the above explanation should provide a basic idea of what is happening in the filter as scale increases. Essentially, at $j = 1$, the filter picks out the finest details available based on the resolution of the image. As j increases, the dilation of the filter coefficients ensures that larger and larger areas are considered when constructing the new images at that scale. These images then are contrasts and averages that show the larger scale (coarser) trends, if any, that are in the original image.

Now, recall that the Taylor diagram in Figure 5 presented the summary of results at $\tau'_1 = 1$ and $\tau'_4 = 8$ m. The former picked out the fine-scale detail, while the latter contrasted areas of 128 m² with the wavelet filter (averaged 256 m² with the scaling filter). At the coarser scale of 8 m, HPSbaf3 and CMCbaf3 were very highly correlated. This results because the size of the blocks contrasted in the wavelet filter is picking out information by the filter matrices in the larger blocks just mentioned. The convergence of HPSbaf3 and CMCbaf3 actually begins at 4 m as noted previously, but becomes stronger as the scale increases. The result is that the two methods look much alike at these coarser scales because there are enough random (spatially unstructured) estimates within the filter blocks at this scale for CMCbaf3 to make the contrast essentially equivalent to that of HPSbaf3. What we are seeing is the CMC estimates averaging to the flat ‘stair-step’ HPS surface in these contrast blocks (since the equivalent BAFs produce exactly the same inclusion zones). Oddly, the spatially structured CHS is less correlated at this scale, due, one conjectures, to the spatially structured arrangement of the estimates within the inclusion zone, potentially restricting the range of estimates available.

Figure 6 presents the raw wavelet coefficient images for HPSbaf3 and CMCbaf3 at scale $\tau'_4 = 8$ m. Visually comparing the two, it is quite evident that the surfaces are indeed very similar and highly correlated. It is interesting to note that this phenomenon occurs when the scale of the filter covers an area (256 m²) that is approximately that of the average inclusion zone area ($a = 216$ m², Table 1). Since the variance images are simply the squares of those in Figure 6 on a pixel basis as noted above, the variance results are very similar as well.

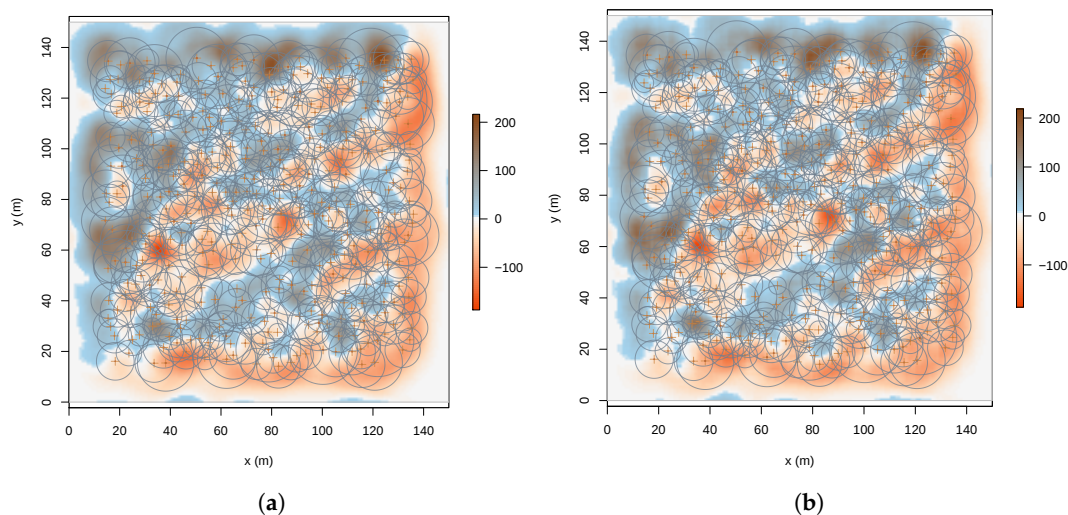


Figure 6. Raw isotropic MODWTwavelet coefficients for scale $\tau'_4 = 8$ m showing the close agreement between HPSbaf3 and CMCbaf3 at this decomposition level. (a) HPS; (b) CMC.

The covariance at 8 m between the results of the images in Figure 6 is presented in Figure 7. Note that the covariance is very similar to the individual variance images because of the high correlation between the methods at this scale. Figure 7 also shows the covariance between HPSbaf3 and CHSbaf3. Because the correlation between these two methods is also quite high at this level, the two figures look quite similar. Notable, however, are several areas where the covariance is negative in the image of Figure 7b, whereas all of the covariances are positive in Figure 7a. The correlations presented in the Taylor diagrams of Figure 5 are computed from the image data in Figure 7 (and similarly for other sampling method pairs) using the methods outlined earlier. Figures 6 and 7 should reinforce that it is the pattern covariance that is calculated on a pixel basis that yields the correlation in the Taylor diagram for each level of the decomposition.

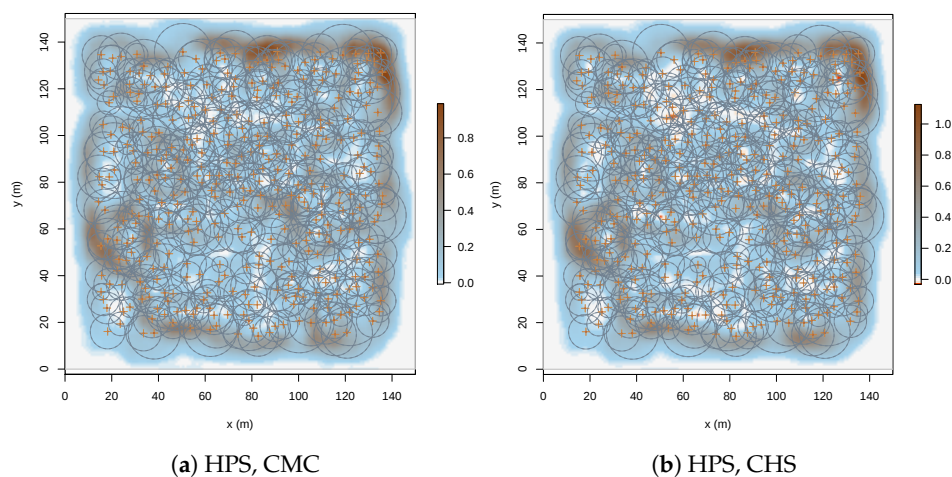


Figure 7. MODWT covariance images for scale $\tau'_4 = 8$ m showing $\text{Cov}(X, Y)$ with (a) $X = \text{HPSbaf3}$ and $Y = \text{CMCbaf3}$; and (b) $X = \text{HPSbaf3}$ and $Y = \text{CHSbaf3}$ at this decomposition level.

4. Summary and Conclusions

The tools discussed here for comparing sampling methods have been illustrated based on sampling surfaces for volume. However, any attribute that can be estimated by the sampling methods

can be used to construct a sampling surface. Application of these tools to the surfaces for different attributes will provide different, possibly conflicting, results with those for volume. This is due to the very nature of areal sampling methods with PPS dominating; optimization is normally for one attribute alone correlated to the dimension used in determining the individual tree's selection probability. In addition, the Taylor diagram presents summaries against the reference method, including correlation. However, it may be useful to also compute correlations among other methods if desired. In the Taylor diagram of Figure 3, it is observed that as the inclusion area for trees of the same size shrinks for sampling methods compared to the reference method, the correlation with the reference tends also to diminish. This is natural as fewer pixels in the methods with smaller average plot radii are covered by inclusion and associated overlap zones and, therefore, have less pattern covariance with the reference, resulting in lower correlation. The Smith plot for the same surfaces, Figure 2, shows a similar relation for the variance. The beauty of using these two displays for comparing sampling methods with different sampling effort is that they are in fact useful in such comparisons, obviating the need to find comparable average inclusion areas to compare competing methods.

Wavelet methods may prove useful for the analysis of variance in sampling surface comparisons. One place where this could be more illuminating is where mapped populations are used in place of the synthetic population used here. Trees within mapped natural populations may be influenced by gradients in edaphic and other factors such that the anisotropic images that have been ignored here (because no such factors exist in the synthetic population) may prove much more useful, uniting scale and directional trends in the comparison of sampling methods (see, e.g., [63]). Regardless, based on the small case study presented here, wavelets can add a new dimension to the comparison of sampling methods by decomposition of the sampling surface variance.

Finally, it should be noted that, just as the results will change when analyzing other stand attributes with the tools described here on the population used, so too will the results change when using a different population. The simulated population used here was chosen to roughly mimic a natural fully-stocked northern hardwoods stand and was chosen for no other reason than to illustrate the efficacy of these methods for more detailed variance comparisons. Having explored the potential usefulness of these methods, they could be employed to contrast populations with differing age structures, species composition, stocking or silvicultural treatments, to name but a few possibilities, in future studies.

Supplementary Materials: The following are available online at www.mdpi.com/1999-4907/8/10/393/s1.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gove, J.H. sampSurf: Sampling Surface Simulations for Areal Sampling Designs in R. 2017. Available online: <http://sampsurf.r-forge.r-project.org/> (accessed on 12 October 2017).
2. Barrett, J.P.; Goldsmith, L. When is n sufficiently large? *Am. Stat.* **1976**, *30*, 67–70.
3. Barrett, J.P.; Nutt, M.E. *Survey Sampling in the Environmental Sciences: A Computer Approach*; COMPRESS, Inc.: Wenworth, NH, USA, 1979; p. 319.
4. Smith, H.F. An Empirical Law Describing Heterogeneity in the Yields of Agricultural Crops. *J. Agric. Sci.* **1938**, *28*, 1–23.
5. Ducey, M.J.; Williams, M.S.; Gove, J.H.; Roberge, S.; Kenning, R.S. Distance limited perpendicular distance sampling for coarse woody material: Theory and field results. *Forestry* **2013**, *86*, 119–128.
6. Rubinstein, R.Y.; Kroese, D.P. *Simulation and the Monte Carlo method*, 2nd ed.; John Wiley and Sons: Hoboken, NJ, USA, 2008; p. 345.
7. Zeide, B. Plot size optimization. *For. Sci.* **1980**, *26*, 251–257.
8. Gambill, C.W.; Wiant, H.V., Jr.; Yandle, D.O. Optimum plot size and BAF. *For. Sci.* **1985**, *31*, 587–594.
9. Lynch, T.B. Optimal plot size or point sample factor for a fixed total cost using the Fairfield Smith relation of plot size to variance. *Forestry* **2017**, *90*, 211–218.

10. Lynch, T.B. Optimal Sample Size and Plot Size or Point Sampling Factor Based on Cost-Plus-Loss using the Fairfield Smith Relationship for Plot Size. *Forestry* **2017**, *90*, 697–709.
11. Robertson, W.M. The line-plot system, its use and application. *J. For.* **1927**, *25*, 157–163.
12. Bitterlich, W. Die Winkelzählprobe. *Allg. For. Holzwirtschaftliche Ztg.* **1948**, *59*, 4–5.
13. Grosenbaugh, L.R. Point-sampling and line-sampling: Probability theory, geometric implications, synthesis. In Occasional Paper 160. 1958. Available online: <https://www.fs.usda.gov/treearch/pubs/2458> (accessed on 5 August 2017).
14. Kitamura, M. On an estimate of the volume of trees in a stand by the sum of critical heights. *Kai Nichi Rin Ko* **1962**, *73*, 64–67.
15. Warren, W.G.; Olsen, P.F. A line intersect method for assessing logging waste. *For. Sci.* **1964**, *3*, 267–276.
16. Cottam, G.; Curtis, J.T. The Use of Distance Measures in Phytosociological Sampling. *Ecology* **1956**, *37*, 451–460.
17. McIntyre, G.M. Estimation of Plant Density Using Line Transects. *J. Ecol.* **1953**, *41*, 319–330.
18. Gregoire, T.G.; Valentine, H.T. *Sampling Strategies for Natural Resources and the Environment*; Applied Environmental Statistics, Chapman & Hall/CRC: New York, NY, USA, 2008.
19. Kershaw, J.A.; Ducey, M.J.; Beers, T.; Husch, B. *Forest Mensuration*, 5th ed.; Wiley-Blackwell: Hoboken, NJ, USA, 2016.
20. Palley, M.N.; Horwitz, L.G. Properties of some random and systematic point sampling estimators. *For. Sci.* **1961**, *7*, 52–65.
21. Leak, W.B.; Yamasaki, M.; Holleran, R. *Silvicultural Guide for Northern Hardwoods in the Northeast*; General Technical Report NRS-132; USDA Forest Service, Northern Research Station: Newtown Square, PA, USA, 2014.
22. Gove, J.H. The sampSurf Package: Sampling Surface Simulation. 2017. Available online: <http://sampsurf.r-forge.r-project.org/> (accessed on 1 August 2017).
23. R Core Team. R: A Language and Environment for Statistical Computing. 2017. Available online: <https://www.R-project.org> (accessed on 1 August 2017).
24. Williams, M.S. New approach to areal sampling in ecological surveys. *For. Ecol. Manag.* **2001**, *154*, 11–22.
25. Williams, M.S. Nonuniform random sampling: An alternative method of variance reduction for forest surveys. *Can. J. For. Res.* **2001**, *31*, 2080–2088.
26. Masuyama, M. A rapid method for estimating basal area in a timber survey—An application of integral geometry to areal sampling problems. *Sankhyā* **1953**, *12*, 291–302.
27. Van Deusen, P. Critical height versus importance sampling for log volume: Does critical height prevail? *For. Sci.* **1990**, *36*, 930–938.
28. Gove, J.H. The “Stem” Class. 2017. Available online: <http://sampsurf.r-forge.r-project.org/> (accessed on 1 August 2017).
29. Bailey, R.L.; Dell, T.R. Quantifying diameter distributions with the Weibull function. *For. Sci.* **1973**, *19*, 97–104.
30. Fast, A.J.; Ducey, M.J. Height-Diameter Equations for Select New Hampshire Tree Species. *North. J. Appl. For.* **2011**, *28*, 157–160.
31. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S*, 4th ed.; Springer: New York, NY, USA, 2002; ISBN 0-387-95457-0.
32. Thompson, S.K. *Sampling*; John Wiley & Sons: New York, NY, USA, 1992.
33. Gove, J.H. Monte Carlo Sampling Methods in sampSurf. 2013. Available online: <http://sampsurf.r-forge.r-project.org/> (accessed on 1 August 2017).
34. Iles, K. Some techniques to generalize the use of variable plot and line intersect sampling. In *Forest Resource Inventories Workshop Proceedings, Volume 1*; Frayer, W.E., Ed.; Colorado State University: Fort Collins, CO, USA, 1979; pp. 270–278.
35. McTague, J.P.; Bailey, R.L. Critical height sampling for stand volume estimation. *For. Sci.* **1985**, *31*, 899–912.
36. Lynch, T.B.; Gove, J.H. An antithetic variate to facilitate upper-stem height measurements for critical height sampling with importance sampling. *Can. J. For. Res.* **2013**, *43*, 1151–1161.
37. Lynch, T.B.; Chang, S.T.; Chandler, J.P. Estimation of individual tree volume by importance sampling and antithetic variates from the cylindrical shells integral. *Can. J. For. Res.* **1992**, *22*, 326–335.

38. Gove, J.H.; Ducey, M.J.; Valentine, H.T.; Williams, M.S. A comprehensive comparison of perpendicular distance sampling methods for sampling downed coarse woody debris. *Forestry* **2013**, *86*, 129–143.
39. Williams, M.S.; Gove, J.H. Perpendicular distance sampling: An alternative method for sampling downed coarse woody debris. *Can. J. For. Res.* **2003**, *33*, 1564–1579.
40. Gove, J.H.; Ringvall, A.; Ståhl, G.; Ducey, M.J. Point relascope sampling of downed coarse woody debris. *Can. J. For. Res.* **1999**, *29*, 1718–1726.
41. Gove, J.H.; Van Deusen, P.C. On fixed-area plot sampling for downed coarse woody debris. *Forestry* **2011**, *84*, 109–117.
42. Deming, W.E. *Some Theory of Sampling*; John Wiley: Hoboken, NJ, USA, 1950.
43. Mahalanobis, P.C. A Sample Survey of the Acreage under Jute in Bengal. *Sankhyā* **1940**, *4*, 511–530.
44. Mahalanobis, P.C. On Large-scale sample surveys. *Philosophical Trans. R. Soc.* **1944**, B231, 329–451.
45. Koch, E.J.; Rigney, J.A. A Method of Estimating Optimum Plot Size from Experimental Data. *Agron. J.* **1951**, *43*, 17–21.
46. Hatheway, W.H.; Williams, E.J. Efficient Estimation of the Relationship between Plot Size and the Variability of Crop Yields. *Biometrics* **1958**, *14*, 207–222.
47. Swallow, W.H.; Wehner, T.C. Optimum Plot Size Determination and its Application to Cucumber Yield Trials. *Euphytica* **1986**, *35*, 421–432.
48. Whittle, P. On the Variation of Yield Variance with Plot Size. *Biometrika* **1956**, *43*, 337–343.
49. Pearce, S.C. An examination of Fairfield Smith's law of environmental variation. *J. Agric. Sci.* **1976**, *87*, 21–24.
50. Freese, F. A Relationship of Plot Size to Variability. *J. For.* **1961**, *59*, 679.
51. O'Regan, W.G.; Palley, M.N. A computer technique for the study of forest sampling methods. *For. Sci.* **1965**, *11*, 99–114.
52. Arvanitis, L.G.; O'Regan, W.G. Computer simulation and economic efficiency in forest sampling. *Hilgardia* **1967**, *38*, 133–164.
53. Wensel, L.C.; John, H.H. A Statistical Procedure for Combining Different Types of Sampling Units in a Forest Inventory. *For. Sci.* **1969**, *15*, 307–317.
54. Reich, R.M.; Arvanitis, L.G. Sampling unit, spatial distribution of trees, and precision. *North. J. Appl. For.* **1992**, *9*, 3–6.
55. Yang, T.R.; Hsu, Y.H.; Kershaw, J.A.; McGarrigle, E.; Kilham, D. Big BAF sampling in mixed species forest structures of northeastern North America: Influence of count and measure BAF under cost constraints. *Forestry* **2017**.
56. Marshall, D.D.; Iles, K.; Bell, J.F. Using a large-angle gauge to select trees for measurement in variable plot sampling. *Can. J. For. Res.* **2004**, *34*, 840–845.
57. Cohen, J.E. Mathematical Population Biologist. In *Bulletin of the American Academy of Arts & Sciences*; American Academy of Arts & Sciences: Cambridge, MA, USA, 2015; pp. 66–67.
58. Taylor, K.E. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* **2001**, *106*, 7183–7192.
59. Gleckler, P.J.; Taylor, K.E.; Doutriaux, C. Performance metrics for climate models. *J. Geophys. Res.* **2008**, *113*, 1–20.
60. Pincus, R.; Batstone, C.; Hofmann, R.J.P.; Glecker, K.E.T.P.J. Evaluating the present-day simulation of clouds, precipitation, and radiation in climate models. *J. Geophys. Res.* **2008**, *113*, 1–20.
61. Percival, D.B.; Walden, A.T. *Wavelet Methods for Time Series Analysis*; Cambridge University Press: Cambridge, UK, 2000.
62. Mondal, D.; Percival, D.B. Wavelet variance analysis for random fields on a regular lattice. *IEEE Trans. Image Proc.* **2012**, *21*, 537–549.
63. Geilhufe, M.; Percival, D.B.; Stern, H.L. Two-dimensional wavelet variance estimation with application to sea ice SAR images. *Comput. Geosci.* **2013**, *54*, 351–360.
64. Lark, R.M.; Webster, R. Analysing soil variation in two dimensions with the discrete wavelet transform. *Eur. J. Soil Sci.* **2004**, *55*, 777–797.
65. Ma, Z.; Zhang, L. Modeling Bird Species Richness at Multiple Spatial Scales Using Two-Dimensional Wavelet Analysis. *For. Sci.* **2015**, *61*, 1–16.
66. Walker, J.S. *A Primer on Wavelets and Their Scientific Applications*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2008.

67. Gove, J.H. The ssWavelets Package: Wavelet Functionality for Package sampSurf. 2017. Available online: <http://sswavelets.r-forge.r-project.org/> (accessed on 1 August 2017).
68. Whitcher, B. Waveslim: Basic Wavelet Routines for One-, Two- and Three-Dimensional Signal Processing. 2015. Available online: <http://lib.stat.cmu.edu/R/CRAN/web/packages/waveslim/index.html> (accessed on 1 August 2017).
69. Addison, P.S. *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance*; Taylor & Francis: Abingdon, UK, 2002.
70. Ripley, B.D. *Spatial Statistics*; John Wiley and Sons: Hoboken, NJ, USA, 1981.
71. Gove, J.H. A User's Guide to the ssWavelets Package. 2017. Available online: <http://sswavelets.r-forge.r-project.org/> (accessed on 1 August 2017).
72. Cochran, W.G. *Sampling Techniques*; John Wiley and Sons: Hoboken, NJ, USA, 1977.
73. Percival, D.B.; Mondal, D. A wavelet variance primer. *Handb. Stat.* **2012**, *30*, 623–657.
74. Keim, M.J.; Percival, D.B. Assessing characteristic scales using wavelets. *Appl. Stat.* **2015**, *64*, 377–393.
75. Verbeeck, H.; Peylin, P.; Bacour, C.; Bonal, D.; Steppe, K.; Ciais, P. Seasonal patterns of CO₂ fluxes in Amazon forests: Fusion of eddy covariance data and the ORCHIDEE model. *J. Geophys. Res.* **2011**, *116*, 1–19. doi:10.1029/2010JG001544.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Supplementary Material

Some Refinements on the Comparison of Areal Sampling Methods via Simulation

J. H. Gove^a

^aUSDA Forest Service, Northern Research Station, 271 Mast Road, Durham, NH 03824 USA

Contents

S.1	Introduction	1	S.2.2	Raster sampling surfaces	2
S.2	Sampling Surfaces	1	S.3	Taylor Diagrams	3
S.2.1	A simple example	1	S.4	Wavelets	3
			S.4.1	Wavelet levels, scale, etc.	3
			S.4.2	Wavelet Variance Plot	5
			S.4.3	Wavelet Taylor Diagrams	6

S.1 Introduction

The material in the following sections presents more details on certain results given in the main text for those that are interested. All results both here and in the main manuscript were derived using the **R** (R Core Team, 2017) software system. The two main **R** packages used in the analysis are **sampSurf** (Gove, 2017a) and **ssWavelets** (Gove, 2017b) as noted in the main manuscript. Other related information is available from the author upon request.

S.2 Sampling Surfaces

S.2.1 A simple example

Figure S.1 presents a simple set of sampling surfaces that are much less complex and busy than those in Figure 1 in the main text where a large number of overlapping inclusion zones obscure the finer details of surfaces with the respective methods. The figure illustrates the differences in the surfaces when only two trees are present. Note that the same two trees are used in each subfigure, the only change is the sampling method. HPS and CPS have flat or constant sampling surfaces within an individual tree's inclusion zone. CHS and CMC have surfaces that vary from sample point to sample point with a given tree's inclusion zone.

Similarly, Figure S.2 presents a corresponding set of figures to those in Figure S.1, but in perspective view. The two figures may be compared for a better idea of sampling surface form under these sampling methods.

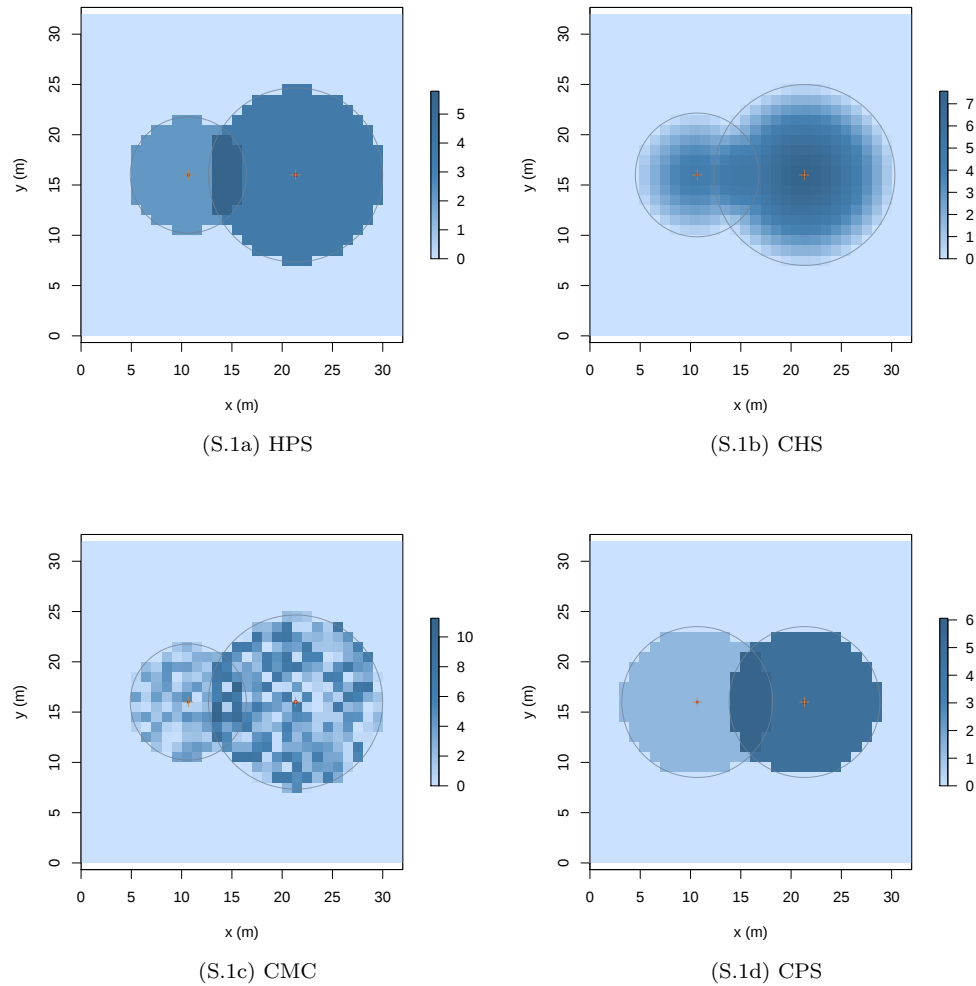


Figure S.1: Four simple sampling surfaces using two trees with DBH= 20 (left) and 30 cm (right) and respective total heights of 15 and 25 m, illustrating the four sampling methods: (a) HPS, (b) CHS, (c) CMC all with $\mathcal{F} = 3 \text{ m}^2 \text{ ha}^{-1}$, and (d) CPS with $R = 7.5 \text{ m}$. Tree positions are denoted by '+', while the circles correspond each individual tree's inclusion zone.

S.2.2 Raster sampling surfaces

The sampling surface simulations shown in Figure S.3 correspond to the perspective view in Figure 1. The benefit of the raster view is that it allows one to see the individual inclusion zones, and it more clearly demonstrates the scale variation between the images. However, it is difficult to get a clear understanding of the shape of the overall surfaces without viewing individual larger versions of the subfigures.

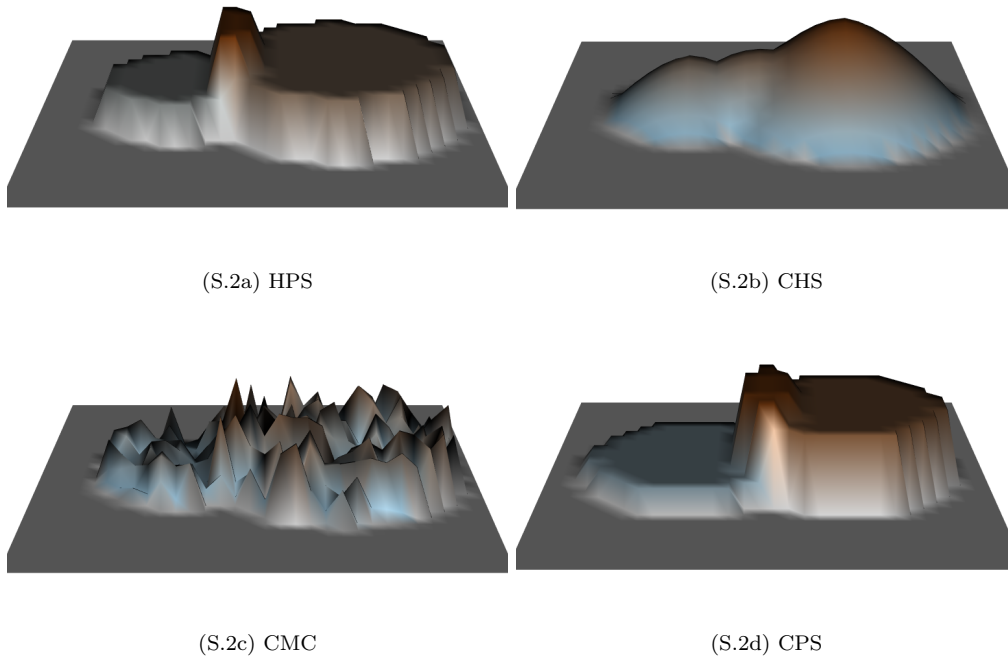


Figure S.2: Perspective view of the sampling surfaces in Figure S.1 using two trees with DBH= 20 (left) and 30 cm (right) and respective total heights of 15 and 25 m, illustrating the four sampling methods: (a) HPS, (b) CHS, (c) CMC all with $\mathcal{F} = 3 \text{ m}^2 \text{ ha}^{-1}$, and (d) CPS with $R = 7.5 \text{ m}$.

S.3 Taylor Diagrams

The information in Table S.1 was computed directly from the `sampSurf` simulations, whose main results appear in Table 1 and may be cross-referenced for convenience. The table presents the variance differences, covariances and correlations required in (5) and (6) for presentation in the Taylor diagram in Figure 3

S.4 Wavelets

The following sections provide material in support of the wavelet analysis component of the manuscript.

S.4.1 Wavelet levels, scale, etc.

Table S.2 presents the levels, j , physical scale, τ'_j , and the number of Haar wavelet and scaling filter coefficients, L_j , in the MODWT analysis. In addition, the total area covered by the wavelet and

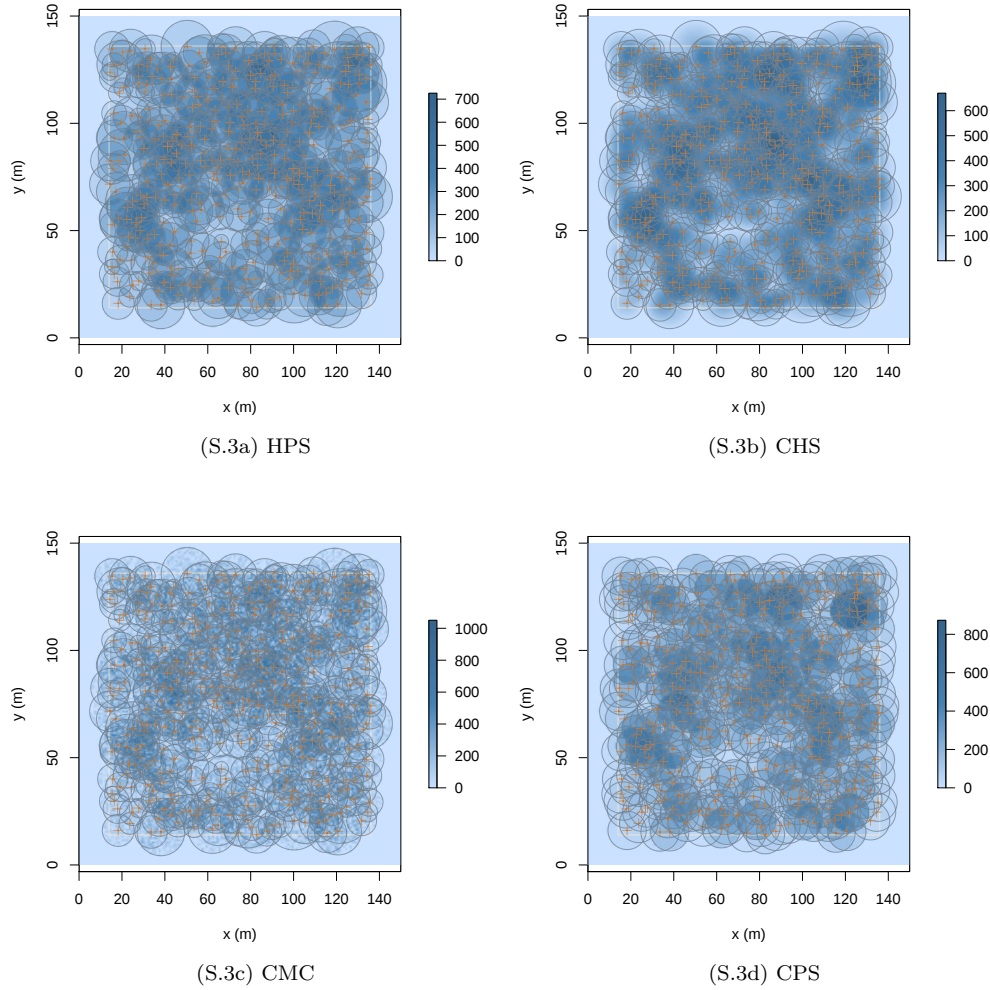


Figure S.3: Raster view of the sampling surfaces from the simulations in Table 1 illustrating the four sampling methods: (a) HPS, (b) CHS, (c) CMC all with $\mathcal{F} = 3 \text{ m}^2 \text{ ha}^{-1}$, and (d) CPS with $R = 8.92 \text{ m}$. Tree positions are denoted by '+', while the circles correspond each individual tree's inclusion zone.

scaling filters, and the area difference—or change in area—associated with the wavelet filters are also presented. Note that since the resolution of the images are $\Delta xy = 1 \text{ m}$ in the simulations, one can think of the various area quantities in either m^2 or pixels.

Table S.1: Standard deviation difference, covariance and correlation results from the **sampSurf** simulations required for the Taylor diagram in Figure 3 with HPSbaf3 as the reference sampling method.

Sampling Method	$\mathcal{F}$ or R	Method ID (m ³)	St. Dev. (m ³)	$\sqrt{\text{Var}(x - y)}$ (m ³)	$\text{Cov}(x, y)$ (m ³) ²	$\text{Corr}(x, y)$
HPS	3.0	HPSbaf3	157.76	0.00	24888.11	1.00
HPS	5.0	HPSbaf5	177.95	88.36	24373.73	0.87
HPS	7.0	HPSbaf7	194.20	120.04	24095.50	0.79
HPS	9.0	HPSbaf9	208.79	143.76	23906.56	0.73
CHS	3.0	CHSbaf3	163.29	56.45	24183.30	0.94
CHS	5.0	CHSbaf5	184.48	106.44	23794.95	0.82
CHS	7.0	CHSbaf7	202.88	137.31	23598.14	0.74
CHS	9.0	CHSbaf9	220.07	162.26	23494.55	0.68
CMC	3.0	CMCbaf3	182.96	92.89	24866.66	0.86
CMC	5.0	CMCbaf5	214.11	149.53	24186.03	0.72
CMC	7.0	CMCbaf7	241.45	185.99	24297.19	0.64
CMC	9.0	CMCbaf9	264.82	217.51	23852.90	0.57
CPS	8.92	CPSrad9	169.97	87.28	23080.03	0.86
CPS	6.89	CPSrad7	196.04	127.44	23539.40	0.76
CPS	5.64	CPSrad6	222.11	164.65	23555.71	0.67
CPS	5.05	CPSrad5	239.80	188.00	23523.68	0.62

Note: the second column lists the BAF (m² ha⁻¹) for methods based on HPS, and plot radius (m), for CPS. The third column lists identifiers used in the main text for each sampling protocol (sampling method plus design parameters).

Table S.2: The levels with associated scale and area dimensions for MODWT analysis.

Level j	Scale τ'_j (m)	L_j	Coverage Area (m ²)	Area Change (m ²)
1	1	2	4	2
2	2	4	16	8
3	4	8	64	32
4	8	16	256	128
5	16	32	1024	512
6	32	64	4096	2048

Note: The coverage area equals $L_j^2 \Delta x y^2$.

S.4.2 Wavelet Variance Plot

The wavelet variance includes those components of the MODWT decomposition that include the wavelet filter; that is, all variances except the scaling-scaling, or smooth variances. The latter can dominate the former in terms of magnitude (depending on the level, J , to which the filters are applied) if included in a summary. Note that (12) is in the form $E[X^2] - E[X]^2$, and implies that

(e.g., Mondal and Percival, 2012, first equation, p. 543)

$$\hat{\sigma}_X^2 = \sum_{j=1}^J \hat{\sigma}_{DVH_j}^2 + \hat{\sigma}_{Z_J}^2$$

where, analogous to the above,

$$\hat{\sigma}_{DVH_j}^2 = \frac{1}{NM} (\|\mathbf{D}_j\|^2 + \|\mathbf{V}_j\|^2 + \|\mathbf{H}_j\|^2) \quad (1)$$

$$\hat{\sigma}_{DVH}^2 = \sum_{j=1}^J \hat{\sigma}_{DVH_j}^2 \quad (2)$$

$$\hat{\sigma}_{Z_J}^2 = \frac{1}{NM} \|\mathbf{Z}_J\|^2 - \bar{X}^2 \quad (3)$$

where $\hat{\sigma}_{DVH_j}^2$ is the j th level *empirical wavelet variance* with total $\hat{\sigma}_{DVH}^2$, and $\hat{\sigma}_{Z_J}^2$ is the *empirical scaling variance*.

Figure S.4 displays the overall isotropic empirical wavelet variances, $\hat{\sigma}_{DVH_j}^2$, plotted against distance, τ_j , for the four sets of simulations in Table 1. Similar plots can be generated for anisotropic components as well (e.g., Geilhufe et al., 2013). Critical height sampling and HPS both show a similar overall trend with variance increasing as the BAF increases. Notably, depending on the BAF, variance peaks at approximately 16 m or greater for HPS, while it peaks at a shorter distance (~ 8 m and larger) under CHS. The variance for CPS more clearly peaks at approximately 8–16 m, and the variances uniformly increase with decreasing plot size. The results for CMC are almost inverted from the others. It is conjectured that this result is due to the completely spatially random nature of the sampling surface: whereas each of the other surfaces have some degree of spatial structure, CMC does not (Figure S.2). An interesting observation is that, even with the unstructured surface, the variances are consistent across each of the BAF's, yielding a minimum variance at approximately 4 m. Each of the sampling methods tends to converge to what appears to be a stable variance (resembling a variogram sill at approximately 6,800 to 8,000) for all BAF's (plot sizes) as the scale (distance) increases.

S.4.3 Wavelet Taylor Diagrams

The full set of Talyor diagrams for the MODWT wavelet decomposition of the simulated sampling surfaces is given in Figure S.5.

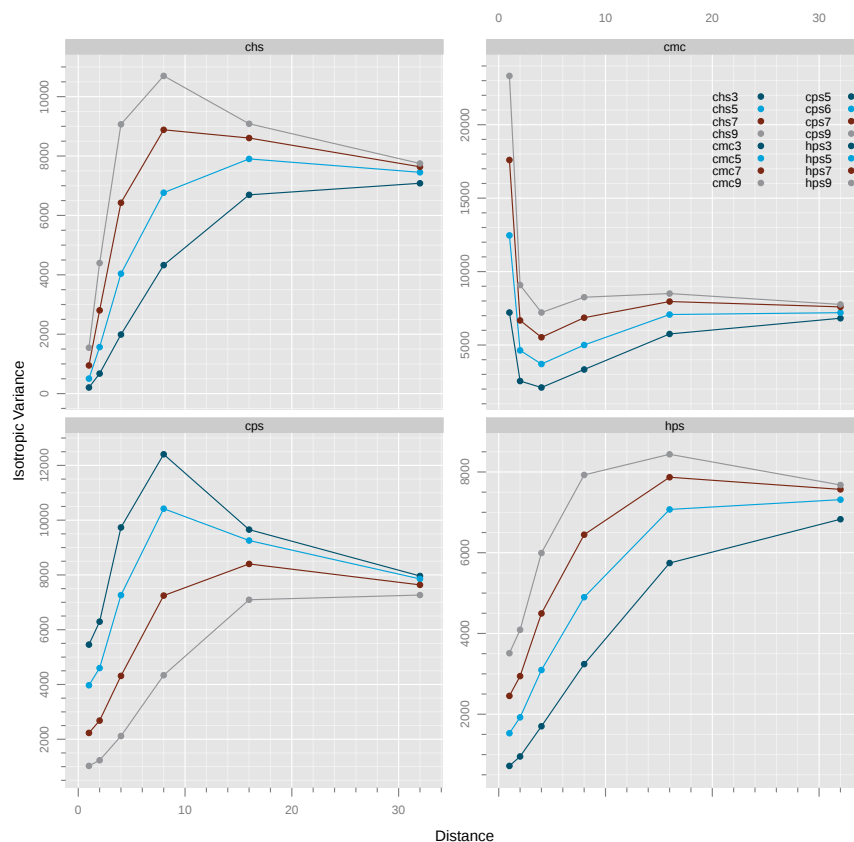


Figure S.4: Empirical wavelet variance plot of the simulations in Table 1 by distance.

References

- M. Geilhufe, D. B. Percival, and H. L. Stern. Two-dimensional wavelet variance estimation with application to sea ice SAR images. *Computers and Geosciences*, 54:351–360, 2013. 6
- J. H. Gove. The sampSurf package: Sampling surface simulation, 2017a. URL <http://sampsurf.r-forge.r-project.org/>. 1
- J. H. Gove. The ssWavelets package: Wavelet functionality for package sampSurf, 2017b. URL <http://sswavelets.r-forge.r-project.org/>. 1
- D. Mondal and D. B. Percival. Wavelet variance analysis for random fields on a regular lattice. *IEEE Transactions on Image Processing*, 21(2):537–549, 2012. 6
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL <https://www.R-project.org>. 1

