

Generalized and synthetic regression estimators for randomized branch sampling

David L. R. Affleck^{1*} and Timothy G. Gregoire²

¹College of Forestry & Conservation, University of Montana, Missoula, MT 59812, USA

²Yale School of Forestry & Environmental Studies, New Haven, CT 06511, USA

*Corresponding author. E-mail: david.affleck@umontana.edu

Received 14 January 2015

In felled-tree studies, ratio and regression estimators are commonly used to convert more readily measured branch characteristics to dry crown mass estimates. In some cases, data from multiple trees are pooled to form these estimates. This research evaluates the utility of both tactics in the estimation of crown biomass following randomized branch sampling (RBS). Synthetic generalized regression (GREG) estimators are developed, and their properties examined against standard estimators. It is shown that synthetic GREG estimators with zero or low design-bias can be obtained, and that the variance of a design-unbiased class of GREG estimators can be unbiasedly estimated. Simulated sampling from 20 censused crowns of two Rocky Mountain species indicated that improvements in accuracy can be obtained through GREG estimation following RBS. Simulations also showed that synthetic GREG estimators that pool data from multiple trees can stabilize coefficients of multivariate regression models to provide improved accuracy over direct GREG estimators. However, for the univariate regression models that proved most adept for the censused crowns, direct GREG estimation provided the lowest average root mean squared error (RMSE) for RBS. Simulations also showed that model-based branch aggregation estimators have generally low RMSE but can be heavily design-biased. For the crown forms studied, use of branch auxiliary information at both the design and estimation stages through RBS and GREG estimation appears to be more efficient than using the information only at the estimation stage following simple or stratified random sampling.

Introduction

Forest biomass and carbon stocks are increasingly relevant both to broad environmental policy and to specific resource management decisions. Forest carbon is now commonly monitored to meet international conventions and in many regions is, or recently has been, actively traded in regulated markets. At the forest management level, tree biomass stocks are now inventoried to inform diverse planning considerations involving not only wood fibre yield, but also bio-energy potential and wildland fuel treatment needs. Accordingly, there has been renewed interest in the development of tree biomass models and in the collection of tree biomass data (e.g. Zeng and Tang, 2012; Jiménez *et al.*, 2013; Ritchie *et al.*, 2013). This research is concerned with the latter, and in particular with the properties of alternative regression-based estimators of crown mass for use with randomized branch sampling (RBS) in felled-tree studies. Focus centres primarily on sampling strategies suitable for conifer species with excurrent branching habits.

Measuring the oven-dry biomass of an individual conifer crown is inevitably a destructive and time-consuming undertaking. It is

also necessarily a sample-based undertaking: excepting very small trees, measuring the oven-dry mass of an entire crown is infeasible. As a result, the majority of felled-tree studies have pursued complete measurement only of auxiliary crown characteristics related to dry mass, and have then drawn branch samples for dry mass determination. In particular, many studies have censused branch green weights (e.g. Zeng and Tang, 2012; Jiménez *et al.*, 2013), whereas others have censused branch diameters and attachment heights (or possibly branch lengths; Schneider *et al.*, 2008; Ritchie *et al.*, 2013). Branch samples have subsequently been drawn to estimate, respectively, mean branch moisture content or a set of coefficients relating mean branch dry mass to various branch dimensions (diameter, height and length). These widely used strategies make use of auxiliary crown information at the estimation stage, applying ratio or regression estimators. Yet, at the sample design stage, they frequently utilize only qualitative auxiliary information in order to establish strata within crowns (e.g. each crown is split into segments in recognition of broad trends in branch size or moisture content). Indeed, accounts of sample designs in the past studies that have estimated crown

(or crown component) biomass from moisture content ratios or dry mass regression relationships have typically focused much more on crown stratification rules than on within-strata branch selection protocols – and the latter appear generally to take the form of simple random sampling (SRS) or convenience sampling. The study by [Monserud and Marshall \(1999\)](#), in which branch sampling with antithetic variates is applied, provides a notable exception.

Exceptions to the two-part methodologies described above are strategies based on RBS. With RBS the cycle-free branching pattern of a tree is exploited to obtain probability samples without first (and perhaps never) counting or measuring the full set of branches (see, e.g. [Valentine *et al.*, 1984](#); [Williams, 1989](#)). Another important element of RBS is the ability to incorporate easily obtained measures of branch size (e.g. branch basal area) into branch selection probabilities. Doing so allows the precision of crown biomass component estimators to be determined not by large inter-branch variations in mass but rather by relatively small variations in branch biomass-to-size ratios. This might be achieved even more effectively by probability proportional-to-size (PPS) sampling, though in applications PPS designs require advance knowledge of the sizes and positions of every branch. Previous studies have examined the gains in precision that might be obtained by RBS and PPS designs by incorporating auxiliary information into branch selection probabilities ([Temesgen *et al.*, 2011](#); [Schlecht and Affleck, 2014](#)). With both designs, it is also possible to utilize the same or additional auxiliary information at the estimation stage by applying generalized regression (GREG) estimators ([Gregoire and Valentine, 2008](#), pp. 180–192). Yet, the extent to which doing so might further improve the precision of crown biomass estimates remains an open question.

In addition to differences in branch selection protocols, past felled-tree studies also differ in how they have made use of branch dry mass data obtained from multiple trees. In studies where the green mass of the whole crown is first obtained and branch samples are then used to estimate moisture content, sample branch data are generally utilized only in the estimators of biomass for the crown from which they were collected. That is, moisture contents of branches drawn from other trees are rarely used to convert the green mass of a subject tree's crown (or its components) to dry mass. Likewise, studies that have drawn probability samples of branches using RBS and then utilized design weights for estimation have treated each sampled tree crown independently. Formally, these strategies have utilized direct estimators (*sensu* [Rao, 2003](#), Chapter 1) with the requisite dry mass data coming only from the tree in question. In contrast, where linear dimensions such as branch diameter and height are censused and branch samples are then used to develop ratio/regression relationships, dry mass sample information from multiple trees (of the same species) is often pooled (e.g. [Loomis *et al.*, 1966](#); [Bartelink, 1996](#); [Schneider *et al.*, 2008](#); [Ritchie *et al.*, 2013](#)). Common, pooled-sample regression coefficients are then applied in biomass estimations across sets of destructively sampled trees. That is, indirect synthetic estimators ([Rao, 2003](#), §4.2) are employed, with each tree's biomass estimator drawing on branch dry mass data collected from other trees.

Despite these methodological differences, per-tree branch samples in felled-tree studies are commonly quite small (i.e. 5–10 branches per tree). This can be attributed to (i) the

imperative for sampling a large number of trees given underlying interests in among-tree variation, (ii) relatively small among-branch variation in dry mass given other branch dimensions (especially green mass, but also often diameter) and (iii) the generally burdensome nature of selecting, processing and oven-drying branch materials. Whatever the reasons, the prevalence of small sample sizes has repercussions for the properties of commonly used ratio and regression estimators. In particular, although it is established that these estimators offer improved accuracy as sample sizes increase ([Särndal *et al.*, 1992](#), Chapter 6), with small samples both bias and variability may be appreciable. Pooling branch sample information across trees where potential commonalities exist (e.g. in species or in spatio-temporal location of sampling) then becomes appealing as it offers a means of improving the precision of ratio/regression coefficients. Yet, it remains unclear under what conditions this would improve the sampling properties of crown biomass estimates at the tree level. [Estavio and Särndal \(2004\)](#) show that the use of data from beyond the domain of interest (in the present case, a tree) cannot improve the accuracy of GREG or calibration estimators when the domain sample size is large. Broadly, the application of pooled-sample regression coefficients presumes that at some level the regression relationships are homogeneous across trees. This homogeneity may itself be difficult to establish when per-tree branch samples are small. Moreover, depending on the form of the estimator used, failure of this assumption could induce substantial bias or imprecision in individual crown mass estimates. Bias, in particular, could have far-reaching implications if the latter estimates are themselves subsequently used as the raw material for development of biomass equations to be applied over large areas.

This research explores the bias and accuracy of alternative regression estimators of crown component biomass for RBS and other branch-based crown sampling designs. The first objective is to describe the source and potential magnitude of the design-bias of tree-level GREG estimators of crown mass, as well as how this bias can be mitigated or even eliminated by drawing on branch data collected from other trees. The second objective is to evaluate the potential gains in accuracy that might be achieved by these estimators, and how these gains are impacted by the sampling design. The third objective is to contrast the design-based strategies against standard synthetic regression estimators that also draw on branch relationships fitted across multiple trees. The next sections define the estimators and develop analytical results related to their properties under RBS and alternative branch sampling designs. Subsequently, results from a set of simulated sampling experiments aimed at estimating the known crown mass of 20 individual Douglas-fir (*Pseudotsuga menziesii* var. *glauca*) and western larch (*Larix occidentalis*) trees are described. Synthesis and recommendations are offered in the discussion.

Randomized branch sampling

[Gregoire and Valentine \(2008, Chapter 13\)](#) provide a comprehensive treatment of RBS. Briefly, the design is grounded in the natural divisibility of a tree's anatomical structure into discrete segments of branch and stem material (see Figure 1a). These segments are connected by nodes that may be true branching

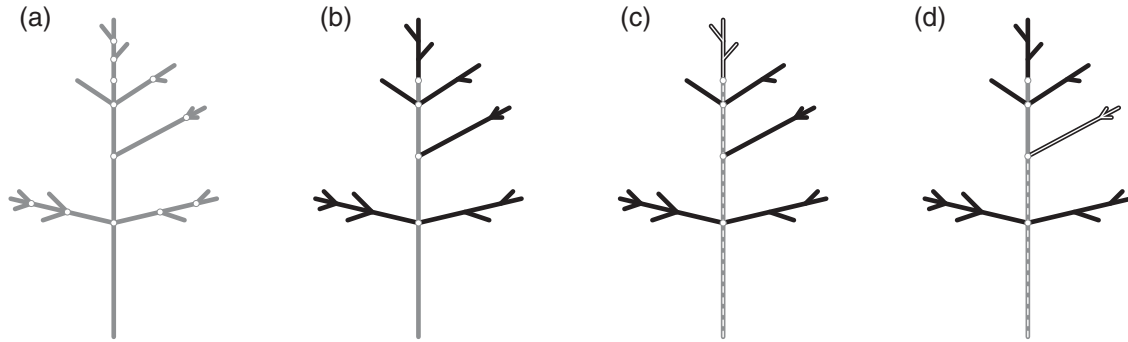


Figure 1 Tree with excurrent crown form (a) divided into discrete segments (solid lines) by nodes (open dots) at branching points and/or diameter thresholds; (b) with RBS terminal segments (black lines) defined as first-order branches plus the tree tip; (c) and (d) with individual terminal segments (white inlay) selected by distinct RBS paths (white dashed lines).

furcations or may be defined at operationally meaningful limits (e.g. the point at which a branch's diameter first falls below a prescribed size). The base of a tree defines the start of a singular basal segment; this is connected by sequences of medial segments to the tips of the branches and stem that collectively define the end points of a set of terminal segments. Typically, the number and size of the terminal segments on a tree are unknown, but it is axiomatic that the tip of each terminal segment can be reached by a unique path through the anterior nodes and segments of the tree.

For simplicity, in the sequel the terminal segments are defined as the first-order branches emanating from the main stem along with the tree tip above a specified diameter threshold (Figure 1b). A conventional definition for the main stem is the path from the base of the tree, passing into the largest diameter limb emerging at every successive branching node and terminating at some minimum diameter (e.g. the gray segments in Figure 1b). Without limiting generality, this definition is adopted here. In turn, the crown is defined as the complement of the main stem and thus consists of the collection of N terminal segments, with all connected higher-order branches being subsumed with their first-order parents into these terminal segments. The mass of a particular crown component on an individual tree can then be written as

$$\tau_y = \sum_{\mathcal{C}} y_k$$

where y_k is the dry mass component of interest for the k th terminal segment and the sum extends across the full set of these segments $\mathcal{C}$. For example, y might denote the dry mass of a whole first-order branch, of its foliage, or of the portion of its branchwood below some diameter threshold (e.g. of 1-h fuels defined by diameters ≤ 0.625 cm; Brown, 1978).

Estimation of τ_y proceeds from measurements on terminal segments selected by RBS. As described in Gregoire and Valentine (2008, §13.2), RBS is initiated at the base of the tree and conditional selection probabilities are assigned to each segment encountered at every successive node. By this process, a path from the basal segment to a single terminal segment is ultimately selected (Figure 1c,d). The (unconditional) selection probability p_k of this terminal segment can be determined recursively from the

conditional selection probabilities of all medial segments along the path leading to its selection. Replicating this procedure n times on a given tree, one obtains a sample s of terminal segments. Estimation of crown biomass on the tree can then proceed via the unbiased Hansen–Hurwitz (HH) estimator

$$\hat{\tau}_{yp} = \frac{1}{n} \sum_s \frac{y_k}{p_k} \quad (1)$$

In practice, RBS is often taken further and used to select a path that extends into a first-order branch. Thus, one might continue to assign conditional selection probabilities at higher-order branching nodes until, say, a single segment with basal diameter below 0.625 cm is chosen. Yet, doing so can be viewed as simply applying RBS to sub-sample and unbiasedly estimate the component biomass of the parent first-order branch. As a result, to ensure sufficient generality to capture crown biomass sampling applications while simplifying notation, we do not consider such further levels of sampling.

From the above, it is evident that RBS is a form of with-replacement, unequal probability sampling. The variance of $\hat{\tau}_{yp}$ is therefore given by

$$\text{var}[\hat{\tau}_{yp}] = \frac{1}{n} \sum_{\mathcal{C}} p_k \left(\frac{y_k}{p_k} - \tau_y \right)^2 \quad (2)$$

This variance formula also applies to the HH estimator in the cases of (i) with-replacement SRS of terminal segments with equal selection probabilities $p_k = 1/N$, and (ii) PPS sampling with arbitrary p_k ($\sum_{\mathcal{C}} p_k = 1$). In contrast to RBS, both these designs require a priori information on the number or sizes of terminal segments in order to affix valid p_k and undertake sampling. In the case of RBS, only partial information on the number and size of terminal and medial segments must be obtained (i.e. only for segments encountered along the selected paths). Strong allometric relationships between branch component biomass and relatively easy-to-measure dimensions like branch basal area and branch length suggest reducing the variance (2) by ensuring that the p_k are functions of one or both of these dimensions (see Gregoire and Valentine, 2008, Chapter 13). However, it should be noted that RBS is not a PPS design and as a rule cannot ensure that p_k is

strictly proportional to the basal area or any other characteristic of a terminal segment.

The general difficulty of identifying, counting and/or measuring all terminal segments in a tree favours in many applications the use of RBS in combination with the HH estimator. Yet, in the case of many conifers or other species under strong apical dominance, excurrent crown forms facilitate the counting and measurement of all the first-order branches that constitute the tree's crown – either in the course of sampling or subsequently. Schlecht and Affleck (2014) found that the expected height of the uppermost branch selected by RBS with $n \geq 5$ was typically above the crown midpoint in the Douglas-fir (*Pseudotsuga menziesii*) and western larch (*Larix occidentalis*) trees they examined. More generally, it can be anticipated that a substantial number of first-order branches will be counted and measured for diameter or other auxiliary characteristics in the course of RBS. As a result, once the sample of branches is selected by RBS only the acropetally higher first-order branches will need to be measured to complete a census of these auxiliary variables. Thus, denoting the auxiliary variables measured on terminal segment k by the q -vector $\mathbf{x}_k$, in some applications it may be feasible to obtain the conforming vector of tree totals

$$\boldsymbol{\tau}_x = \sum_c \mathbf{x}_k$$

For example, $\boldsymbol{\tau}_x$ may include the total number of terminal segments (N), their aggregate basal area, their total length or the sum of their attachment heights (or crown depths). These can be contrasted with corresponding estimates available from RBS

$$\hat{\tau}_{xp} = \frac{1}{n} \sum_s \frac{\mathbf{x}_k}{p_k}$$

The effort required for the auxiliary variable census may prove worthwhile if elements of $\hat{\tau}_{xp}$ are correlated with $\hat{\tau}_{yp}$.

Regression estimators

Within the design-based approach to inference, the GREG estimator provides a flexible means of drawing on auxiliary information following the application of RBS or other unequal probability designs. One form of the GREG estimator is

$$\hat{\tau}_{yr} = \hat{\tau}_{yp} - (\hat{\tau}_{xp} - \boldsymbol{\tau}_x)^T \hat{\mathbf{B}} \quad (3)$$

where $\hat{\mathbf{B}}$ is a q -vector of regression coefficients estimated from the $(y_k, \mathbf{x}_k)$ data in the sample s . For example, $\hat{\mathbf{B}}$ can be estimated as

$$\hat{\mathbf{B}} = \left(\sum_s \frac{\mathbf{x}_k^T \mathbf{x}_k}{\sigma_k^2 p_k} \right)^{-1} \left(\sum_s \frac{\mathbf{x}_k^T y_k}{\sigma_k^2 p_k} \right) \quad (4)$$

where the σ_k^2 are a set of observation weights. Both the composition of the auxiliary vector and the weights are typically suggested by a working model for the distribution of y_k given $\mathbf{x}_k$.

Specifically, estimator (4) is motivated by an approximation to the conditional distribution of branch mass of the form

$$\mathcal{E}(y_k | \mathbf{x}_k) = \mathbf{x}_k^T \mathbf{B} \quad (5a)$$

$$\mathcal{V}(y_k | \mathbf{x}_k) = \sigma_k^2 = \sigma^2 g(\mathbf{x}_k) \quad (5b)$$

where $\mathcal{E}(\cdot)$ and $\mathcal{V}(\cdot)$ define the expectation and variance under a working stochastic model, and $g(\cdot)$ is scalar-valued function of branch characteristics. The GREG estimator (3) can be called a model-assisted estimator (sensu Särndal *et al.*, 1992, p. 227) since it is motivated by a stochastic model for the variable of interest but retains a number of key properties (including design consistency) regardless of the validity of that model.

Written in the form (3), it is evident that the GREG estimator provides an additive adjustment to $\hat{\tau}_{yp}$. This random adjustment is obtained from the deviation of the RBS estimates from their known population totals in the auxiliary variable space. In general, the GREG estimator is not unbiased since the expectation of this adjustment term,

$$\begin{aligned} E[(\hat{\tau}_{xp} - \boldsymbol{\tau}_x)^T \hat{\mathbf{B}}] &= E[(\hat{\tau}_{xp} - \boldsymbol{\tau}_x)^T (\hat{\mathbf{B}} - E[\hat{\mathbf{B}}])] \\ &= \sum_{j=1}^q \text{cov}[\hat{\tau}_{xp}, \hat{B}_j] \end{aligned}$$

is non-null except in special cases. The adjustment term also adds variability to (3). Nonetheless, the GREG estimator can offer substantial improvements in accuracy over $\hat{\tau}_{yp}$. When the $\mathbf{x}$ are chosen judiciously and the working model (5a) provides a useful approximation, the two terms on the right-hand side of (3) will be positively correlated. Since

$$\text{var}[\hat{\tau}_{yr}] = \text{var}[\hat{\tau}_{yp}] + \text{var}[(\hat{\tau}_{xp} - \boldsymbol{\tau}_x)^T \hat{\mathbf{B}}] - 2 \text{cov}[\hat{\tau}_{yp}, (\hat{\tau}_{xp} - \boldsymbol{\tau}_x)^T \hat{\mathbf{B}}]$$

the reduction in the overall variance of the GREG estimator attributable to this correlation can more than offset the additional variability associated with the second term.

It should be noted that the GREG estimator defined above is based on the branch selection probabilities p_k . The alternative form of this estimator in which the sample data are weighted by inclusion probabilities is more commonly presented, but the p -weighted form (3) is more readily applicable to RBS data (cf. Särndal *et al.*, 1992, §6.4).

Sampling properties

Analytical results concerning the sampling distribution of $\hat{\tau}_{yr}$ are few owing to the nonlinearities in the second term of (3). There are, however, two cases wherein such results can be obtained. The first is the large sample setting, where a first-order Taylor linearization (see, e.g. Särndal *et al.*, 1992, §6.6) can be used to show that $\hat{\mathbf{B}}$ is approximately design-unbiased for

$$\mathbf{B} = \left(\sum_c \frac{\mathbf{x}_k^T \mathbf{x}_k}{\sigma_k^2} \right)^{-1} \left(\sum_c \frac{\mathbf{x}_k^T y_k}{\sigma_k^2} \right) \quad (6)$$

and the GREG estimator is approximately design-unbiased for τ_y . Also, for large n , the variance of (3) can be approximated as

$$\text{var}[\hat{\tau}_{yr}] \approx \frac{1}{n} \sum_c p_k \left(\frac{e_k}{p_k} - \tau_e \right)^2 \quad (7)$$

where $e_k = y_k - \mathbf{x}_k^T \mathbf{B}$ and $\tau_e = \sum_c e_k$. As such, if there are strong linear correlations between y_k and elements of $\mathbf{x}_k$, the mean squared error of $\hat{\tau}_{yr}$ can be substantially lower than that of $\hat{\tau}_{yp}$. Yet, the approximate variance formula (7) also points to the relevance of the sample design (via the p_k) even in large samples. In particular, the variance will be reduced if the p_k are approximately proportional to the absolute size of the e_k . That is, an efficient design for GREG estimation assigns larger selection probabilities where the absolute deviations of y_k from $\mathbf{x}_k^T \mathbf{B}$ are larger.

Though informative, these results may have little relevance in applied crown mass sampling efforts. Cochran (1977, §6.3) suggests a minimum sample size of 30 as a working rule for the use of asymptotic results for the simple ratio estimator; this is considerably larger than the standard number of branches selected for dry mass determination in most felled-tree studies. For finite samples, more exacting analysis of the GREG estimator shows that its bias depends on the sampling design and on the variance-covariance properties of y_k and $\mathbf{x}_k$ (Lepik, 2007). Likewise, variance expression (7) is obtained assuming convergence of $\hat{\mathbf{B}}$ to $\mathbf{B}$. In small samples, the variance of $\hat{\mathbf{B}}$ may be large relative to the variances of (and correlation between) $\hat{\tau}_y$ and $\hat{\tau}_{xp}$.

In the context of the approximate variance (7), it is also worth noting that branch mass data are often heteroscedastic, with larger variability in dry mass in larger branches (see, e.g. Kershaw and Maguire, 1995; Monserud and Marshall, 1999; Ritchie et al., 2013). On the one hand, this suggests increased variability in $\hat{\mathbf{B}}$ under size-dependent sampling designs such as RBS or PPS relative to SRS. The former favour the selection of larger branches, which carry less information concerning the association between y_k and $\mathbf{x}_k$. On the other hand, positive correlations between branch selection probabilities p_k and absolute regression residuals $|e_k|$ will reduce the approximate variance (7). Under heteroscedasticity, the largest absolute deviations of y_k from $\mathbf{x}_k^T \mathbf{B}$ are likely to be found among the same larger branches favoured by PPS and RBS (relative to SRS). Still, this cannot guarantee that the correlation between p_k and $|e_k|$ will be positive under these designs. In contrast, under SRS there is at least an assurance that the $p_k = 1/N$ cannot be negatively correlated with $|e_k|$.

The second setting in which analytical results can be obtained for (3) is that where values for the regression coefficients are obtained independently of the branch sample s . If $\hat{\mathbf{B}}$ is independent of $\hat{\tau}_{yp}$ and $\hat{\tau}_{xp}$, then the GREG estimator is unbiased for any sample size. It also has conditional variance

$$\text{var}[\hat{\tau}_{yr} | \hat{\mathbf{B}}] = \frac{1}{n} \sum_c p_k \left(\frac{\hat{e}_k}{p_k} - \tau_{\hat{e}} \right)^2 \quad (8)$$

where $\hat{e}_k = y_k - \mathbf{x}_k^T \hat{\mathbf{B}}$ and $\tau_{\hat{e}} = \sum_c \hat{e}_k$. When $\hat{\mathbf{B}}$ is assigned a value without appeal to sample data, the GREG estimator (3) is also

known as the difference estimator (Särndal et al., 1992, §6.3). For a given design, the optimal value to assign to $\hat{\mathbf{B}}$ is

$$\hat{\mathbf{B}}_{\text{opt}} = \text{var}[\hat{\tau}_{xp}]^{-1} \text{cov}[\hat{\tau}_{xp}, \hat{\tau}_{yp}]$$

With $\hat{\mathbf{B}}_{\text{opt}}$ substituted for $\hat{\mathbf{B}}$ in (3), the GREG estimator is at least as efficient as the HH estimator (1) for any sample size. The difficulty, of course, is obtaining the value $\hat{\mathbf{B}}_{\text{opt}}$, or a reasonable approximation thereof, without relying on the branch information in s .

One approach to approximating $\hat{\mathbf{B}}_{\text{opt}}$ is to draw on information obtained from other felled trees sampled independently of the tree in question. That is, if branch data from multiple similar trees (e.g. of the same species, size class or geographic origin) are available, then it may be possible to borrow strength from those data in order to more accurately estimate the crown mass of every individual felled tree. Denote by $s-$ the sample of branch y and $\mathbf{x}$ data from trees other than the tree in question, and let $\hat{\mathbf{B}}_-$ be an estimate of the regression coefficients obtained from this sample:

$$\hat{\mathbf{B}}_- = \left(\sum_{s-} \frac{\mathbf{x}_k^T \mathbf{x}_k}{\sigma_k^2 p_k} \right)^{-1} \left(\sum_{s-} \frac{\mathbf{x}_k^T y_k}{\sigma_k^2 p_k} \right)$$

where the weights are assigned a structure under a working model for branch mass that applies across trees. This allows for the synthetic regression estimator

$$\hat{\tau}_{yr-} = \hat{\tau}_{yp} - (\hat{\tau}_{xp} - \tau_{\mathbf{x}})^T \hat{\mathbf{B}}_- \quad (9)$$

Since $\hat{\mathbf{B}}_-$ is independent of $\hat{\tau}_{xp}$ for a given focal tree, estimator (9) is unbiased for that tree's total:

$$\text{E}[\hat{\tau}_{yr}] = \text{E}[\hat{\tau}_{yp}] - \text{E}[(\hat{\tau}_{xp} - \tau_{\mathbf{x}})^T] \times \text{E}[\hat{\mathbf{B}}_-] = \tau_y - \mathbf{0}^T \times \text{E}[\hat{\mathbf{B}}_-] = \tau_y$$

where $\mathbf{0}$ is a $q \times 1$ null vector. This result applies regardless of how much the focal tree might deviate in form or branch allometry from the other trees contributing to $s-$. Also, while per-tree sample sizes may be small, the aggregated sample $s-$ from multiple trees could be large, and thus $\hat{\mathbf{B}}_-$ conditional on the felled trees in hand could be quite precise. Furthermore, if the nature of the relationship between y and $\mathbf{x}$ is consistent across trees, the precision of (9) could be superior to $\hat{\tau}_{yp}$ even for small s . On the other hand, for any individual tree $\hat{\mathbf{B}}_-$ could deviate substantially from $\hat{\mathbf{B}}_{\text{opt}}$ leading to poor correspondence between y and $\mathbf{x}^T \hat{\mathbf{B}}_-$. In such cases, the variance of $\hat{\tau}_{yr-}$ would increase, possibly above that of $\hat{\tau}_{yp}$.

With no information from s being used, it is apparent that there is real potential for $\hat{\mathbf{B}}_-$ to be a poor estimator of the regression relationships for the tree of interest. Thus, it is also reasonable to consider estimating the regression coefficients using the combined samples s and $s-$. Denoting this combined sample from all felled trees as $s+ = \{s, s-\}$, the regression coefficients could be

estimated as

$$\hat{\mathbf{B}}_+ = \left(\sum_{s+} \frac{\mathbf{x}_k^T \mathbf{x}_k}{\sigma_k^2 p_k} \right)^{-1} \left(\sum_{s+} \frac{\mathbf{x}_k^T y_k}{\sigma_k^2 p_k} \right)$$

leading to the alternative synthetic estimator of total crown mass

$$\hat{\tau}_{yr+} = \hat{\tau}_{yp} - (\hat{\tau}_{xp} - \tau_x)^T \hat{\mathbf{B}}_+ \quad (10)$$

Rao (2003) points out that this is also known as the ‘survey regression’ estimator. Inasmuch as $\hat{\mathbf{B}}_+$ is not independent of $\hat{\tau}_{xp}$, the latter will generally be biased for τ_y . Yet, the correlation between $\hat{\mathbf{B}}_+$ and $\hat{\tau}_{xp}$, and thus the bias of $\hat{\tau}_{yr+}$, can be expected to be small and decreasing as the number of felled trees contributing to $s+$ increases (Rao, 2003, p. 20). Under the continuing assumption that per-tree samples are small, $\hat{\mathbf{B}}_+$ should have comparable precision to $\hat{\mathbf{B}}_-$. Also, because $\hat{\mathbf{B}}_+$ is informed by data from the tree in question, it can be expected to more closely approach $\hat{\mathbf{B}}_{\text{opt}}$ for a given tree, leading to higher precision in $\hat{\tau}_{yr+}$ relative to $\hat{\tau}_{yr-}$. Thus, the use of $s+$ or $s-$ represents a bias-variance tradeoff, though it is unclear how the corresponding GREG estimators of τ_y will perform in terms of accuracy (MSE) relative to $\hat{\tau}_{yp}$ or $\hat{\tau}_{yr}$.

For estimating precision, the standard list sampling variance estimator is readily adapted for use with $\hat{\tau}_{yr-}$. With $\hat{e}_k = y_k - \mathbf{x}_k^T \hat{\mathbf{B}}_-$ and letting $\hat{\tau}_{\hat{e}} = (1/n) \sum_s (\hat{e}_k / p_k)$, the variance of (9) given a selected set of felled trees can be unbiasedly estimated as

$$v(\hat{\tau}_{yr-}) = \frac{1}{n(n-1)} \sum_s \left(\frac{\hat{e}_k}{p_k} - \hat{\tau}_{\hat{e}} \right)^2 \quad (11)$$

Unbiasedness again follows from the independence of s and $\hat{\mathbf{B}}_-$:

$$E[v(\hat{\tau}_{yr-})] = E_{s-} E_s[v(\hat{\tau}_{yr-}) | \hat{\mathbf{B}}_-] = E_{s-}[\text{var}(\hat{\tau}_{yr-} | \hat{\mathbf{B}}_-)] = \text{var}(\hat{\tau}_{yr-})$$

Analogous statistics could be adopted for estimation of $\text{var}[\hat{\tau}_{yr+}]$ (with $\hat{e}_k = y_k - \mathbf{x}_k^T \hat{\mathbf{B}}_+$) and $\text{var}[\hat{\tau}_{yr}]$ ($\hat{e}_k = y_k - \mathbf{x}_k^T \hat{\mathbf{B}}$), but the dependence of both $\hat{\mathbf{B}}_+$ and $\hat{\mathbf{B}}$ on s breaks the second equality in the preceding argument. This will lead to biased variance estimation, though the bias can be expected to be low for $v(\hat{\tau}_{yr+})$ if s forms only a small proportion of the pooled sample $s+$.

The utility of the synthetic biomass estimators (9) and (10) clearly hinges on the consistency of regression relationships across the set of trees available for data pooling. Interestingly, however, a lack of consistency impacts primarily the variance, not the bias, of these estimators. This stands in contrast to a fully model-based approach, wherein crown biomass might be obtained from an estimator of the form

$$\hat{\tau}_{ym} = \tau_x^T \hat{\mathbf{B}}_+ = \sum_c \mathbf{x}_k^T \hat{\mathbf{B}}_+ \quad (12)$$

with the last expression conveying why this approach is also known as the branch aggregation method (i.e. it is the sum of branch-level predictions). In the small area estimation literature,

(12) is also known as the regression-synthetic estimator (Rao, 2003, p. 46). From a design-based perspective, this estimator is clearly biased for τ_y . Perhaps more importantly, the potential magnitude of the bias for any given tree is large and largely unpredictable, being a function of the overall goodness-of-fit of the working model as well as the deviance of the subject tree from the set of felled trees contributing to $\hat{\mathbf{B}}_+$. That said, this model-based estimator will be much more precise than $\hat{\tau}_{yp}$ or any of the regression estimators defined above, for the simple reason that it is subject only to variation in $\hat{\mathbf{B}}_+$ (itself based on the potentially large sample $s+$). Thus, the model-based estimator $\hat{\tau}_{ym}$ trades a potentially large design-bias for consistently high precision.

Simulation

To provide an empirical evaluation of the small sample performance of the regression estimators defined above, they were applied to simulated branch samples drawn from a set of fully enumerated tree crowns. The primary goal was to identify whether the synthetic GREG estimators $\hat{\tau}_{yr-}$ and $\hat{\tau}_{yr+}$ could provide improved accuracy for small RBS samples (relative to the HH estimator) without incurring substantial bias. Secondly, it was of interest to determine whether the improvements in accuracy, if any, offered by these estimators were similar for RBS, PPS, SRS and stratified SRS (STRS) designs.

Crown parameters

The population data consist of measurements from 10 Douglas-fir and 10 western larch trees felled in western Montana, USA. On each tree, the green mass (y), basal area (a) and attachment height (h) were obtained on every live first-order branch that (i) had a basal diameter of at least 0.5 cm and (ii) was attached below the point where the stem tapered to a 5 cm diameter. The green mass and height of the tree tip above this 5 cm diameter threshold were also measured, and this tip was treated as a single first-order branch. The tree-level sums of these green mass measurements (first-order branches and tip) were then treated as the parameters of interest. Since depth in the crown is a more commonly used measure of position in branch modelling studies, attachment heights were re-scaled to crown depths as

$$d = \frac{H - h}{H - h_{(1)}}$$

where H is the tree height and $h_{(1)}$ is the height of the lowest live branch on the tree. Tree and crown characteristics are summarized in Figure 2 and the dataset is described in more detail in Schlecht and Affleck (2014).

The form and strength of the relationships between branch green mass, basal area and crown depth are shown for one randomly selected tree in Figure 3. This figure also depicts variations in trends across trees. Of the variables measured, the best predictor of branch green mass was branch basal area: this association was consistently strong ($r \geq 0.85$) and linear, though variation in branch mass also increased with basal area. Branch mass also increased with crown depth, yet the strength and form of the relationship varied substantially across trees.

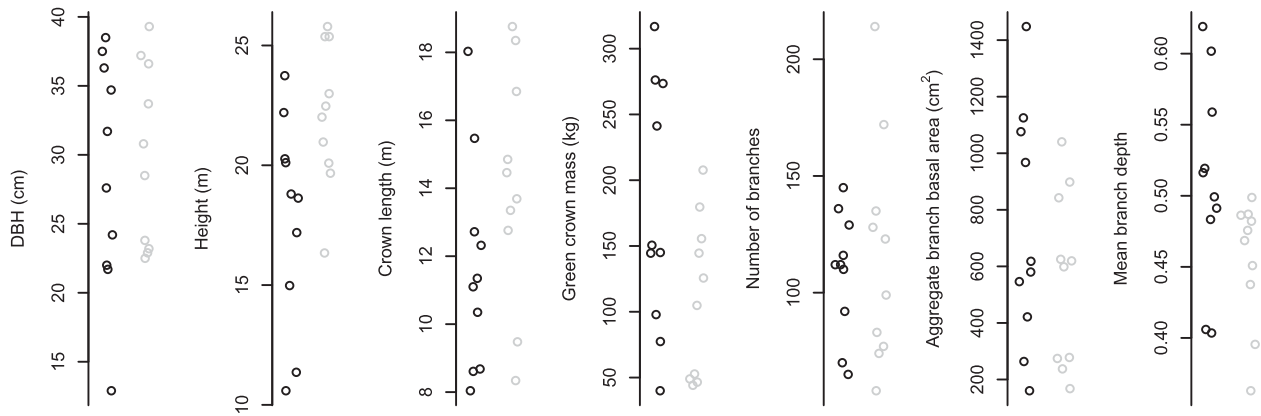


Figure 2 Distributions of tree sizes and crown attributes for the 20 branch populations used in simulated sampling (black points denote Douglas-fir trees; grey points denote western larch).

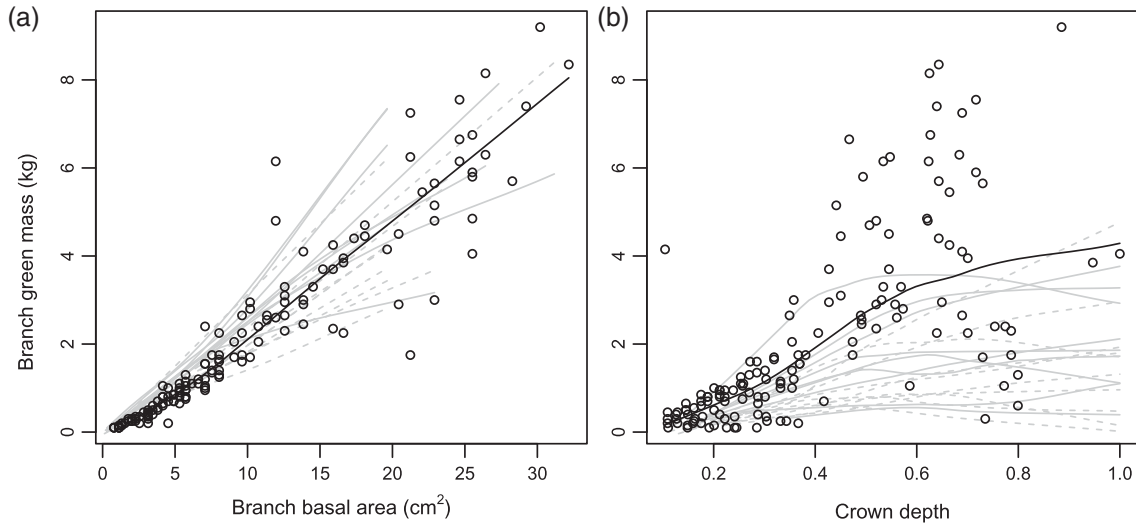


Figure 3 Branch green mass versus (a) branch basal area and (b) crown depth for a single Douglas-fir tree. Black lines denote loess curves for the selected tree; grey lines denote loess curves for the other 19 trees.

Sampling strategies

Branch sampling was simulated independently on each of the 20 crowns. Sampling designs included PPS and both stratified and unstratified SRS. STRS was implemented by dividing each tree crown into five strata of equal length along the bole, and drawing equal numbers of branches from each. For PPS, the branch selection probabilities were set proportional to branch basal area ($p_k \propto a_k$). Similarly, in simulating RBS, all conditional branch selection probabilities were set proportional to segment (branch or stem) basal areas. PPS and RBS designs with branch selection probabilities set proportional to diameter^{8/3} (see Gregoire and Valentine, 2008, p. 414) were also evaluated, but these designs produced slightly more variable crown mass estimates on average as well as qualitatively similar differences among estimators. Under each design, 250 000 samples of 5, 10, 15 and 20 branches were drawn (with replacement) from each

tree. The number of simulated samples was chosen to ensure the Monte Carlo standard error fell below 1 per cent of the mean for every strategy when averaged across all trees.

In addition to the HH estimator, the standard p -weighted GREG estimator $\hat{\tau}_{yr}$ (3), the synthetic GREG estimators $\hat{\tau}_{yr-}$ (9) and $\hat{\tau}_{yr+}$ (10) and the model-based estimator $\hat{\tau}_{ym}$ (12) were evaluated on each simulated sample. Two distinct models of the form (5a) were specified for branch green mass, leading to two distinct forms for each of the regression estimators. The first model was a multivariate linear regression with $q = 3$ and $\mathbf{x}_k = [1 \ a_k \ d_k]^T$; the second was a univariate (ratio or no-intercept) model with $q = 1$ and $x_k = a_k$. The multivariate-form regression estimators necessitated the computation of three coefficients and $\hat{\tau}_{xp} = [\hat{N}_p \ \hat{\tau}_{ap} \ \hat{\tau}_{dp}]^T$, where the elements of $\hat{\tau}_{xp}$ denote HH estimates of the total number of branches, the aggregate branch basal area and the total branch depth, respectively. Under the second

model, the univariate-form estimators required the computation of only a single regression coefficient and $\hat{\tau}_{xp} = \hat{\tau}_{ap}$. In the calculations of $\hat{\mathbf{B}}_-$ and $\hat{\mathbf{B}}_+$ for the synthetic estimators, only data from trees of the same species were pooled ($\hat{\mathbf{B}}_-$ is computed from 9 trees' samples and $\hat{\mathbf{B}}_+$ from 10). Results for regression estimators with observation weights set proportional to branch basal area ($\sigma_k^2 \propto a_k$) are presented. Modifying these weights to accord with models specifying constant branch mass variance ($\sigma_k^2 = \sigma^2$) or increased heteroscedasticity ($\sigma_k^2 \propto a_k^2$) did not appreciably alter the estimators' relative performances for the sample sizes examined. Finally, none of the models differentiated among branches by stratum, and thus all regression coefficients under STRS were obtained by pooling data from all five crown strata within trees.

Altogether, eight regression estimators (4 per model) plus the HH estimator were applied to each sample, and samples were drawn using four designs. The performance of each of these 36 strategies was summarized separately for each tree and sample size by estimating relative bias

$$\text{Bias} = 100 \frac{\sum \hat{\tau}_y - \tau_y}{\tau_y}$$

and root mean squared error (RMSE)

$$\text{RMSE} = 100 \frac{\sqrt{\sum (\hat{\tau}_y - \tau_y)^2}}{\tau_y}$$

where $\hat{\tau}_y$ represents one of the crown biomass estimators described above, τ_y the known total for the tree in question and the summations extend across the set of 250 000 simulations for a given tree, sample size and design. The relative RMSEs of $\hat{\tau}_{yr}$, $\hat{\tau}_{yr-}$

and $\hat{\tau}_{yr+}$ under RBS were also compared with sample estimates furnished by list sampling variance estimators of the form (11).

Results

Bias

As expected, the HH and $\hat{\tau}_{yr-}$ estimators were unbiased under all simulation scenarios. The estimated biases of the other estimators under RBS are summarized in Figures 4 and 5. At $n = 5$, the multivariate-form regression estimators $\hat{\tau}_{yr}$ and $\hat{\tau}_{ym}$ exhibited biases exceeding ± 5 per cent for most trees (Figure 4). The design-bias in $\hat{\tau}_{ym}$ is attributable to lack of fit in the underlying model, either in its form for any given tree or in the homogeneity of coefficients across trees. Moreover, the percent bias of $\hat{\tau}_{ym}$ did not change with sample size, reflecting the fact that this estimator is not design-consistent. In contrast, the bias of the design-consistent $\hat{\tau}_{yr}$ declined with increasing sample size (at a rate of approximately $n^{-1/2}$). Unlike $\hat{\tau}_{ym}$, $\hat{\tau}_{yr}$ contains a bias correction term that effectively adjusts for lack of model fit as the sample size increases. The bias of the synthetic multivariate-form estimator $\hat{\tau}_{yr+}$ also declined with sample size, although its percent bias never exceeded ± 5 per cent under RBS (Figure 4). This reflects weak correlation between each tree's $\hat{\tau}_{xp}$ and the combined-tree $\hat{\mathbf{B}}_+$ when the latter is based on nine trees beyond the tree in question.

The absolute percent biases in $\hat{\tau}_{yr}$ under PPS were similar to those under RBS, but were smaller under SRS and STRS (results not shown). For the latter designs, the percent bias of $\hat{\tau}_{yr}$ under the multivariate model was below ± 5 per cent (SRS) or ± 7 per cent (STRS) for all trees at $n = 5$. Under SRS and STRS, the estimated number of branches is always equal to N ; as noted in Lepik (2007), this eliminates one source of bias in the GREG estimator.

Applying the regression estimators constructed under the simpler univariate model (with branch area as the sole predictor

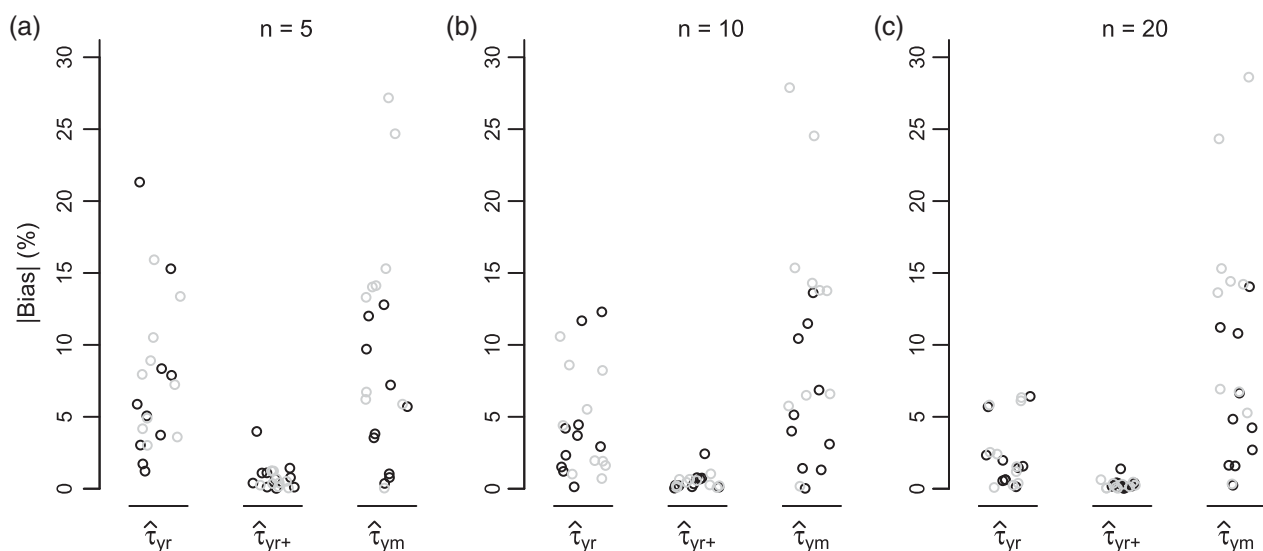


Figure 4 Percent bias of alternative multivariate-form regression estimators of crown mass under RBS by sample size (black points are Douglas-fir trees; grey points are western larch).

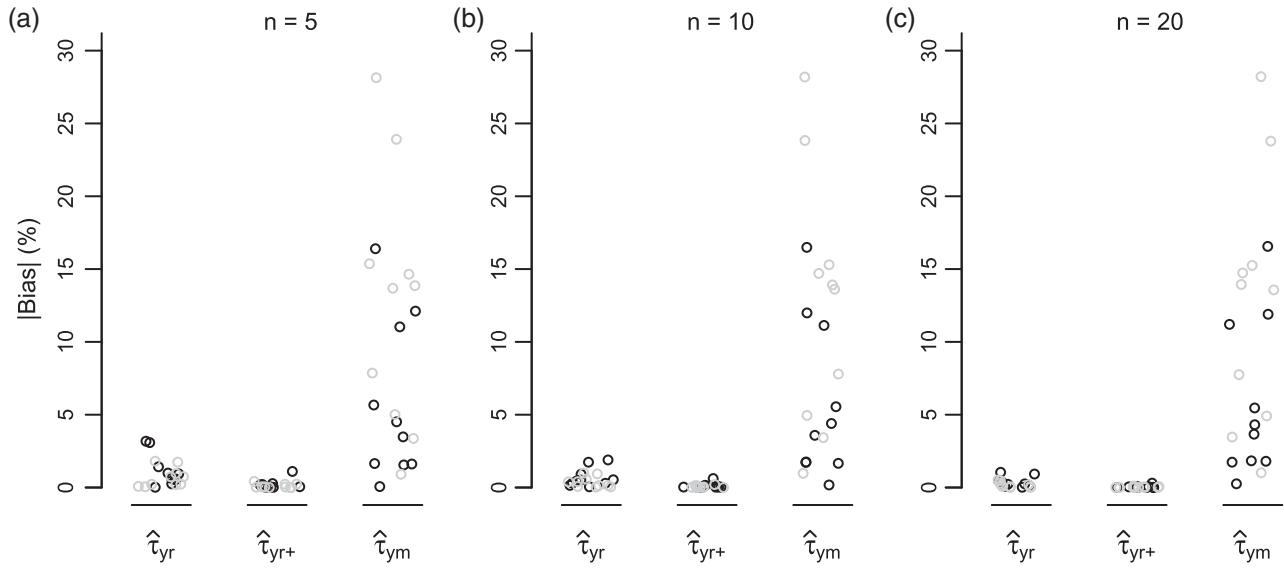


Figure 5 Percent bias of alternative univariate-form estimators of crown mass under RBS by sample size (black points are Douglas-fir trees; grey points are western larch).

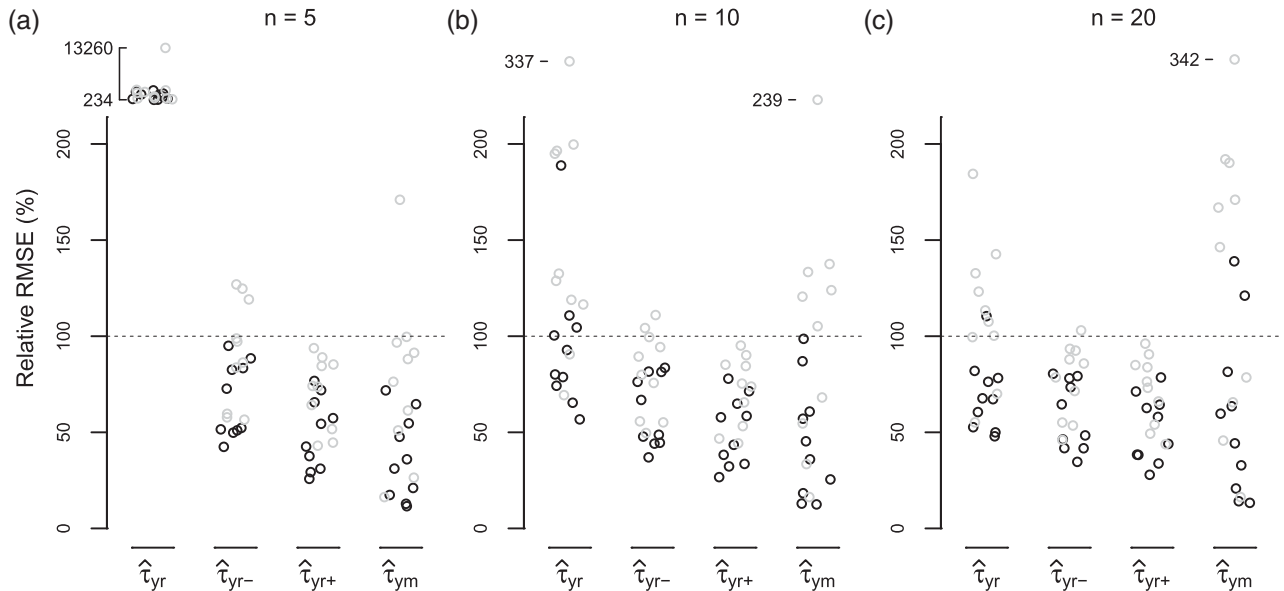


Figure 6 RMSEs of multivariate-form regression estimators of crown mass under RBS relative to the standard errors of HH estimators by sample size (black points are Douglas-fir trees; grey points are western larch). Note that vertical axis scales vary above 220%.

of mass), the biases of $\hat{\tau}_{yr+}$ and especially $\hat{\tau}_{yr}$ with RBS were reduced substantially (Figure 5). The biases of the univariate and multivariate forms of $\hat{\tau}_{ym}$ were essentially the same, and similar across all designs. This suggests that a large portion of the biases stem either from (i) heterogeneity in mean mass-to-basal-area ratios across trees, or (ii) nonlinearities in mass-to-basal-area relationships. Under PPS, the univariate $\hat{\tau}_{yr}$ and $\hat{\tau}_{yr+}$ estimators reduce mathematically to the HH estimator and are design-unbiased.

Accuracy

The accuracy of each multivariate-form regression estimator under RBS is expressed relative to the precision of the HH estimator in Figure 6 and Table 1. In the multivariate form, $\hat{\tau}_{yr}$ was more accurate than the HH estimator at $n = 10$ for only 8 of 20 trees, and for only 13 trees at $n = 20$. At $n = 5$, the RMSE of this estimator greatly exceeded that of the HH estimator for all trees (by a factor of 19, on average). This poor performance at $n = 5$ is attributable to high variability as well as frequently large

bias (Figure 4). In contrast, the variance of the unbiased synthetic estimator $\hat{\tau}_{yr-}$ in its multivariate form exceeded that of the HH estimator at $n = 5$ only for 3 trees, and for 8 trees its standard error was less than two-thirds that of the HH estimator (Figure 6). Further, the biased estimator $\hat{\tau}_{yr+}$ was consistently more accurate than the HH estimator even at $n = 5$, with RMSEs less than half the standard error of the HH estimator for 7 of 20 trees. Except for one tree, the biased model-based estimator $\hat{\tau}_{ym}$ had lower RMSE than the HH estimator at $n = 5$. However, both the bias (Figure 4) and precision of this estimator under RBS changed only slightly

Table 1 RMSE of multivariate regression estimators as a percentage of the standard error of the HH estimator under RBS averaged across trees by sample size, design and number of predictors

n	Estimator	RBS	PPS	SRS	STRS
5	$\hat{\tau}_{yp}$	100.0	54.6	195.3	192.8
	$\hat{\tau}_{yr}$	1912.5	1576.6	300.2	74.8
	$\hat{\tau}_{yr-}$	79.0	64.8	79.1	81.4
	$\hat{\tau}_{yr+}$	59.8	54.3	74.1	74.9
	$\hat{\tau}_{ym}$	57.4	55.5	55.6	54.6
10	$\hat{\tau}_{yp}$	100.0	54.6	195.2	192.5
	$\hat{\tau}_{yr}$	126.9	89.6	68.1	61.2
	$\hat{\tau}_{yr-}$	71.3	60.2	78.1	80.7
	$\hat{\tau}_{yr+}$	61.0	54.5	75.1	77.0
	$\hat{\tau}_{ym}$	74.3	72.7	73.0	72.3
20	$\hat{\tau}_{yp}$	100.0	54.7	195.4	192.8
	$\hat{\tau}_{yr}$	91.1	66.9	65.6	62.0
	$\hat{\tau}_{yr-}$	67.8	57.8	77.8	80.5
	$\hat{\tau}_{yr+}$	61.8	54.6	75.8	78.3
	$\hat{\tau}_{ym}$	100.3	98.6	98.9	98.3

with sample size, and at $n = 20$ its RMSE was lower than that of the HH estimator for only 12 of the 20 trees.

The RMSEs of all four multivariate-form regression estimators relative to one another were similar under RBS and PPS. The synthetic regression estimators under PPS proved more accurate than their counterparts under RBS, and more accurate than the HH estimator under RBS (Table 1). However, applying PPS, the HH estimator was more accurate than $\hat{\tau}_{yr-}$ and nearly equivalent in RMSE to $\hat{\tau}_{yr+}$, that is, across the 20 trees, the HH estimator under α -weighted PPS performed as well as any of the multivariate-form regression estimators. In contrast, under SRS and STRS, the regression estimators showed marked and consistent improvements in accuracy relative to the HH estimator under RBS (Table 1), with the exception of $\hat{\tau}_{yr}$ at $n = 5$ under unstratified SRS. Thus, for the designs where no branch size/position information is used at the design stage, the use of such information by the synthetic estimators $\hat{\tau}_{yr-}$ and $\hat{\tau}_{yr+}$ uniformly improved the precision of crown biomass estimates. Interestingly, under STRS, neither $\hat{\tau}_{yr-}$ nor $\hat{\tau}_{yr+}$ outperformed $\hat{\tau}_{yr}$ at any sample size.

Applying the univariate-form regression estimators, across all sample sizes the standard and synthetic estimators were all at least as accurate under RBS as the HH estimator (Figure 7). Indeed, there was little difference between the accuracies of $\hat{\tau}_{yr}$, $\hat{\tau}_{yr-}$ and $\hat{\tau}_{yr+}$, and all of them had RMSEs that were on average less than 65 per cent of the standard error of the HH estimator at $n = 5$ (Table 2). Surprisingly, the standard estimator $\hat{\tau}_{yr}$ proved more accurate than $\hat{\tau}_{yr-}$ and $\hat{\tau}_{yr+}$ at all sample sizes, with the largest percentage improvement in accuracy at $n = 5$. These estimators all utilize a single regression coefficient expressing a biomass-to-basal-area ratio, and these differ only in the branch data from which they draw: $\hat{B}$ from the subject tree only (s); $\hat{B}_-$ only from other trees of the same species (s_-); $\hat{B}_+$ from all trees of the same species (s_+). Evidently, the divergence of $\hat{B}_-$ and $\hat{B}_+$

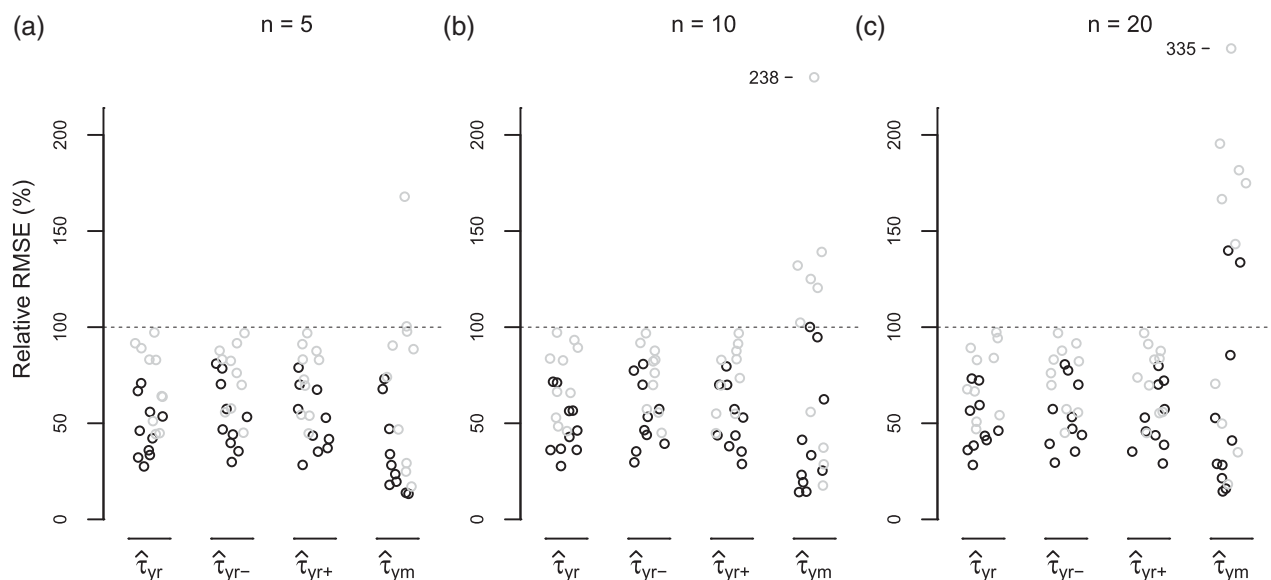


Figure 7 RMSEs of univariate-form regression estimators of crown mass under RBS relative to the standard errors of HH estimators by sample size (black points are Douglas-fir trees; grey points are western larch). Note that vertical axis scales vary above 220%.

Table 2 RMSE of univariate regression estimators as a percentage of the standard error of the HH estimator under RBS averaged across trees by sample size, design and number of predictors

<i>n</i>	Estimator	RBS	PPS	SRS	STRS
5	$\hat{\tau}_{yp}$	100.0	54.6	195.3	192.8
	$\hat{\tau}_{yr}$	58.9	–	65.9	59.9
	$\hat{\tau}_{yr-}$	64.2	–	78.1	81.1
	$\hat{\tau}_{yr+}$	62.5	–	75.2	76.6
	$\hat{\tau}_{ym}$	53.8	51.4	54.2	53.3
10	$\hat{\tau}_{yp}$	100.0	54.6	195.2	192.5
	$\hat{\tau}_{yr}$	60.4	–	67.3	62.3
	$\hat{\tau}_{yr-}$	64.0	–	77.6	80.7
	$\hat{\tau}_{yr+}$	63.0	–	75.8	78.0
	$\hat{\tau}_{ym}$	71.2	68.9	71.3	70.6
20	$\hat{\tau}_{yp}$	100.0	54.7	195.4	192.8
	$\hat{\tau}_{yr}$	61.5	–	68.7	65.7
	$\hat{\tau}_{yr-}$	64.0	–	77.5	80.6
	$\hat{\tau}_{yr+}$	63.4	–	76.3	78.9
	$\hat{\tau}_{ym}$	96.6	94.6	96.6	96.0

from the optimal ratio for the tree in question more than offsets their improved precision relative to $\hat{B}$. Also, the degraded accuracy of these estimators in multivariate form (Figure 6) must stem from poor precision or among-tree heterogeneity in the intercept and crown depth regression coefficients.

As in the other scenarios examined, the high precision of the univariate model-based estimator $\hat{\tau}_{ym}$ under RBS led to low RMSEs (Figure 7) for those trees for which the estimator had small bias. Conversely, for other trees, large design-bias in this estimator led to large RMSEs. Moreover, the RMSEs of $\hat{\tau}_{ym}$ changed very little with n under RBS, such that the accuracy of this estimator declined relative to the precision of the HH estimator as sample size increased (Table 2). Similar results obtained for the univariate-form estimator $\hat{\tau}_{ym}$ under PPS, SRS and STRS.

Under PPS, the univariate-form estimators $\hat{\tau}_{yr}$, $\hat{\tau}_{yr-}$ and $\hat{\tau}_{yr+}$ are identical to each other and to $\hat{\tau}_{yp}$. As shown in Table 2, this strategy provided the highest accuracy for all sample sizes except at $n = 5$ where the model-based estimator offered lower mean RMSE. Drawing branches by SRS or STRS, all of the univariate-form regression estimators but $\hat{\tau}_{ym}$ were more accurate than the simple or stratified expansion estimators for all trees ($n \geq 5$). They also offered improvements in accuracy over the HH estimator under RBS (Table 2). In addition, for some trees SRS or STRS coupled with $\hat{\tau}_{yr}$, $\hat{\tau}_{yr-}$ and $\hat{\tau}_{yr+}$ (in univariate form) was as accurate as or even more accurate than RBS coupled with those same estimators. That is, for some trees, the use of q_k in setting selection probabilities for RBS appeared to be superfluous: the same or improved accuracies were achieved by stratified or unstratified equal probability sampling followed by ratio estimation. However, for the majority of trees, accuracy was improved through the use of auxiliary information first at the design stage (via RBS selection probabilities) and then again at the estimation stage (via one of $\hat{\tau}_{yr}$, $\hat{\tau}_{yr-}$ and $\hat{\tau}_{yr+}$).

In terms of variance estimation, the estimator (11) was unbiased for both the univariate and multivariate forms of $\hat{\tau}_{yr-}$ in accordance with theory. Also as expected, this variance estimator consistently understated the MSE of $\hat{\tau}_{yr}$. On average across the 20 trees, the mean variance estimate for the univariate form of the estimator fell below the observed MSE of $\hat{\tau}_{yr}$ by 9.5 per cent ($n = 5$) and 5.5 per cent ($n = 20$) under RBS. Variance estimates were considerably worse for the multivariate form $\hat{\tau}_{yr}$ estimator, underestimating its observed MSE by 69 per cent under RBS at $n = 20$. In contrast, at $n = 5$ the list sampling estimator of the variance of the univariate form $\hat{\tau}_{yr+}$ differed by no more than ± 2 per cent from that estimator's observed MSE for all trees. At this same sample size, MSEs of the multivariate form $\hat{\tau}_{yr+}$ under RBS were underestimated consistently, and by 7.6 per cent on average.

Discussion

In felled-tree studies, the goal of field sampling is to appraise the mass of each individual tree selected for study. The numbers so obtained provide the raw material for identifying sources of variation among trees and developing predictive equations. This argues for regarding the set of felled trees and their biomass totals as fixed, and for holding these tree totals as the target parameters for field sampling strategies. Estimating the parameters of superpopulation models linking linear branch dimensions to mass can allow for improved precision (Figures 6 and 7), but the sampling strategies adopted should ensure that tree totals are estimated without substantial bias.

Many felled-tree studies have relied on stratified but non-random branch samples to estimate crown biomass. The credibility of such estimates rests in part on the validity of models, implicit or explicit, for branch moisture content or mass. Credibility rests also on the likelihood of obtaining balanced samples when branches are selected arbitrarily (or by convenience) across crown strata and/or trees of similar characteristics. Thompson (2002, §8.5) discusses the advantages of balanced sampling when model-based ratio or regression estimators are anticipated, and these are apparent in our results in the improved performance of the GREG estimators at small sample sizes under STRS compared with SRS (Tables 1 and 2). Yet, Thompson also advises against eschewing randomization altogether. The risk of model misspecification, leading to imbalance and biased estimation under purposive selection, should be appreciated. The lack of randomization in some past crown sampling efforts may be attributable to the fact that frame development for SRS or PPS branch sampling designs can be time-consuming and logistically challenging in the field. If so, it is important to recognize that RBS strategies permit probabilistic branch selection without having to first count or measure all the branches on a tree. Also, if a sample is selected entirely from the lower crown, there is never a need to complete a count of the branches. In these ways, RBS allows for incremental (and perhaps only partial) frame development while at the same time allowing the practitioner to draw on advance knowledge of branch allometry and crown architecture at both the design and estimation stages.

The simulations conducted in this study show how branch models that might assist design-based estimators can also lead

to large biases in tree totals when synthetic estimators are applied (i.e. $\hat{\tau}_{ym}$ in Figures 4 and 5). These biases likely could be reduced if more refined branch models were developed. Ritchie *et al.* (2013), for example, aggregate branch-level predictions from models that are nonlinear in branch diameter and position. More generally, models developed specifically for the species and region of interest, and possibly incorporating additional characteristics such as branch length, should provide estimates with less bias than was observed for $\hat{\tau}_{ym}$ in our simulations. Nonetheless, structural crown characteristics and branch allometry can vary substantially within and across trees (Figure 2), such that model-based estimators can be highly precise but heavily biased for some tree totals. Qualitatively, this bias will result in felled tree biomass estimates that incorporate an overoptimistic perspective on homogeneity in branch allometry (or moisture content). It cannot be determined how this might influence the subsequent development of crown biomass equations, except that it is likely to supply overoptimistic assessments of the goodness-of-fit of those equations.

Owing to the emphasis on RBS, this research did not focus on the utility of delineating vertical crown strata of the form frequently used in felled-tree studies. In particular, stratified branch sampling within a crown creates the opportunity to apply combined or separate ratio/regression coefficients to estimate tree totals. Given the small sample sizes under consideration, only the former was considered in our simulations. Generally, stratum-specific ratio/regression estimators allow for improved precision but can be heavily biased when stratum sample sizes are small (Cochran, 1977, §§6.12, 7.11; Meng, 1993). The comparisons of synthetic vs standard GREG estimators for RBS in the present study have parallels to the question of combined vs separate ratio estimation, but the context is different. In particular, our evaluation of synthetic and direct GREG estimators is in regard to their performance in estimating strata-level (tree-level) totals, not a combined (multi-tree) total. Thus, our evaluations are properly in the context of small area or domain estimation in that information from outside each domain of interest (each individual felled tree) is considered in the estimation of domain characteristics. Yet, unlike classical small area estimation problems the domains in question also constitute sampling strata. This alignment of individual tree crowns as both sampling strata and small area domains is what allows for the design-unbiased synthetic estimator $\hat{\tau}_{yr-}$.

In our RBS simulations, when employing multiple auxiliary variables as predictors both synthetic estimators $\hat{\tau}_{yr-}$ and $\hat{\tau}_{yr+}$ were clearly superior to the direct GREG estimator $\hat{\tau}_{yr}$ (Figure 6). The former had either no or low bias at $n = 5$, and $\hat{\tau}_{yr+}$ had lower MSE than the simple HH estimator for all trees. For trees or species beyond the small set considered in our simulations, the bias of $\hat{\tau}_{yr+}$ could increase, though its analytic properties suggest it will remain small in relation to that of $\hat{\tau}_{yr}$. In terms of accuracy, while the asymptotic results of Estavio and Särndal (2004) show that out-of-sample data cannot improve direct estimators when n is large, our simulations indicate that such data can be useful when domain sample sizes are small. At the same time, after accounting for aggregate branch basal area, the utility of branch numbers and aggregate crown depths as linear predictors of total branch mass was limited. Using branch basal area as the sole predictor yielded univariate-form regression estimators with generally

lower average RMSE (RMSE increased only for $\hat{\tau}_{yr+}$; Tables 1 and 2). Moreover, among the univariate GREG estimators, there were smaller differences between the RMSEs of $\hat{\tau}_{yr}$, $\hat{\tau}_{yr+}$, and $\hat{\tau}_{yr-}$ across our 20 trees. Taken together, these results suggest that pooling branch data from multiple trees is effective in stabilizing partial regression coefficient estimates when sample and effect sizes are small. But given the dominant effect of branch basal area (relative to the crown depth and intercept terms) for the measured trees, this stabilization effect was of limited utility in estimating their crown masses. Nonetheless, since its variance can be estimated unbiasedly, it is worth noting that with a_k as the sole predictor the unbiased synthetic estimator $\hat{\tau}_{yr-}$ did show precision comparable with both $\hat{\tau}_{yr}$ and $\hat{\tau}_{yr+}$.

Although the simulation results stem from a set of only 20 trees on two species, they do accord with the sampling properties developed above. We conjecture also that for dry crown mass estimation, the synthetic estimators $\hat{\tau}_{yr+}$ and $\hat{\tau}_{yr-}$ will show improved relative accuracy since more consistent (across trees) relationships between branch dry mass and other branch dimensions can be anticipated. It is less clear how the results may translate to component biomass estimation (e.g. estimation of crown foliage mass or of whole crown 1-h fuel mass). Yet, the utility of GREG estimation is enhanced when interest attaches to multiple attributes, as when several crown biomass components are to be estimated from the same sample. Since one can expect higher levels of variability in each component of branch mass given overall branch size, there may be advantages to borrowing strength from other trees to stabilize regression relationships if only a small number of branches are drawn from each tree. Finally, in trees with very large first-order branches or decurrent crowns, it may be necessary to continue RBS within selected branches. Doing so will add variability to branch-level mass estimates and will likely weaken associations with other whole-branch dimensions. For these reasons, it is expected that ratio/regression estimation and the census of first-order branches will be less efficient for trees with these crown forms or characteristics.

Overall, in applying RBS to estimate crown mass parameters for trees with excurrent branching habits, it appears prudent to collect diameter measurements on all first-order branches for use at the estimation stage. In the Douglas-fir and western larch trees examined, the utility of branch height information appeared to be secondary while branch lengths were not collected. Secondly, if branch data from several felled trees of the same species or crown form are available, it should be recognized that these allow for design-unbiased (or nearly unbiased) regression estimation as well as unbiased variance estimation in the case of $\hat{\tau}_{yr-}$. The latter property suggests direct comparison, and selection, of HH and $\hat{\tau}_{yr-}$ estimators on the basis of estimated precision. This study did not evaluate the accuracy improvements that can be obtained using auxiliary information against the associated measurement costs. Yet, in our experience working with coniferous species in the Pacific Northwest, it has proved economical to complete a first-order branch diameter census by measuring beyond the uppermost RBS selection. Notably, such a census would be required a priori for PPS, SRS or STRS with or without ratio/regression estimation, whereas RBS allows branch sampling and any auxiliary variable census to take place concurrently.

Conflict of interest statement

None declared.

Funding

This work was partially supported by the USDA Forest Service (11-JV-11242305-120), the Joint Fire Science Program (10-1-02-13) and the Inland Northwest Growth & Yield Cooperative.

References

- Bartelink, H.H. 1996 Allometric relationships on biomass and needle area of Douglas-fir. *For. Ecol. Manage.* **86**, 193–203.
- Brown, J.K. 1978 Weight and density of crowns of Rocky Mountain conifers. Research Paper INT-197, USDA Forest Service, Intermountain Forest and Range Experiment Station, Ogden, UT, 56 pp.
- Cochran, W.G. 1977 *Sampling Techniques*. 3rd edn. Wiley.
- Estavio, V.M. and Särndal, C.E. 2004 Borrowing strength is not the best technique within a wide class of design-consistent domain estimators. *J. Official Stat.* **20**, 645–669.
- Gregoire, T.G. and Valentine, H.T. 2008 *Sampling Strategies for Natural Resources and the Environment*. Chapman and Hall/CRC, pp. 35–38, 62–65, 405–421.
- Jiménez, E., Vega, J.A., Fernández-Alonso, J.M., Vega-Nieva, D., Álvarez-González, J.G. and Ruiz-González, A.D. 2013 Allometric equations for estimating canopy fuel load and distribution of pole-size maritime pine trees in five Iberian provenances. *Can. J. For. Res.* **43**, 149–158.
- Kershaw, J.A. Jr and Maguire, D. (1995) Crown structure in western hemlock, Douglas-fir, and grand fir in western Washington: trends in branch-level mass and leaf area. *Can. J. For. Res.* **25**, 1897–1912.
- Lepik, N. (2007) On the bias of the generalized regression estimator in survey sampling. *Acta Appl. Math.* **97**, 41–52.
- Loomis, R.M., Phares, R.E. and Crosby, J.S. 1966 Estimating foliage and branchwood quantities in shortleaf pine. *For. Sci.* **12**, 30–39.
- Meng, X. 1993 On the absolute bias ratio of ratio estimators. *Statist. Probab. Lett.* **18**, 345–348.
- Monserud, R. and Marshall, J. 1999 Allometric crown relations in three northern Idaho conifer species. *Can. J. For. Res.* **29**, 521–535.
- Rao, J.N.K. 2003 *Small Area Estimation*. Wiley.
- Ritchie, M.W., Zhang, J. and Hamilton, T.A. 2013 Aboveground tree biomass for *Pinus ponderosa* in Northeastern California. *Forests* **4**, 179–196.
- Särndal, C.E., Swensson, B. and Wretman, J. 1992 *Model Assisted Survey Sampling*. Springer.
- Schlecht, R.M. and Affleck, D.L.R. 2014 Branch aggregation and crown allometry condition the precision of randomized branch sampling estimators of conifer crown mass. *Can. J. For. Res.* **44**, 499–508.
- Schneider, R., Berninger, F., Ung, C.H., Bernier, P.Y., Swift, D.E. and Zhang, S.Y. 2008 Calibrating jack pine allometric relationships with simultaneous regression. *Can. J. For. Res.* **38**, 2566–2578.
- Temesgen, H., Monleon, V., Weiskittel, A. and Wilson, D. 2011 Sampling strategies for efficient estimation of tree foliage biomass. *For. Sci.* **57**, 153–163.
- Thompson, S.K. 2002 *Sampling*. 2nd edn. Wiley.
- Valentine, H.T., Tritton, L.M. and Furnival, G.M. 1984 Subsampling trees for biomass, volume, or mineral content. *For. Sci.* **30**, 673–681.
- Williams, R.A. 1989 Use of randomized branch and importance sampling to estimate loblolly pine biomass. *Southern J. Appl. For.* **13**, 181–184.
- Zeng, W.S. and Tang, S.Z. (2012) Modeling compatible single-tree above-ground biomass equations for masson pine (*Pinus massoniana*) in southern China. *J. For. Res.* **23**, 593–598.