

Parametric bootstrap estimators for hybrid inference in forest inventories

Mathieu Fortin^{1,*}, Rubén Manso^{1,3}, and Robert Schneider²

¹INRA/AgroParisTech, UMR 1092 LERFoB, 54000 Nancy, France.

²Département de Biologie, Chimie et Géographie, , Université du Québec à Rimouski, Rimouski, Canada.

³Current address: Forestry Research, Northern Research Station, Forestry Commission, Roslin, Midlothian, United Kingdom

*Corresponding author: mathieu.fortin@agroparistech.fr

22 Nov 2017

Abstract

In forestry, the variable of interest is not always directly available from forest inventories. Consequently, practitioners have to rely on models to obtain predictions of this variable of interest. This context leads to hybrid inference, which is based on both the probability design and the model. Unfortunately, the current analytical hybrid estimators for the variance of the point estimator are mainly based on linear or nonlinear models and their use is limited when the model reaches a high level of complexity.

An alternative consists of using a variance estimator based on resampling methods, as suggested in Rubin (1987). However, it turns out that a parametric bootstrap estimator of the variance can be biased in contexts of hybrid inference. In this study, we designed and tested a corrected bootstrap estimator for the variance of the point estimator, which can easily be implemented as long as all of the stochastic components of the model can be properly simulated. Like previous estimators, this corrected variance estimator also makes it possible to distinguish the contribution of the sampling and the model to the variance of the point estimator.

The results of three simulation studies of increasing complexity showed no evidence of bias for this corrected variance estimator, which clearly outperformed the bootstrap variance estimator used in previous studies. Since the implementation of this corrected variance estimator is not much more complicated, we recommend its use in contexts of hybrid inference based on complex models.

Keywords. Estimator bias; Variance estimator; Estimate of the total; Horvitz-Thompson estimator; Log grade; Lumber recovery model.

Introduction

In forestry, it often happens that the variable of interest is not directly available from forest inventories. For example, volume is a key variable in forest management planning, but its measurement on each individual tree, even in a small sample, is absolutely impossible without destructive sampling. In such contexts, practitioners usually rely on predictive models that are based on auxiliary variables that can easily be measured. In the case of volume, tree diameter and height are used to predict tree volume.

When it comes to estimating the total or mean volume of a population, design-based point estimators such as the Horvitz-Thompson (HT) estimator (Horvitz and Thompson, 1952) are unbiased. However, their associated variance estimators tend to underestimate the true variance because they do not take the variance related to model predictions into account. Model-assisted estimators require some observations of the response variable (Mandallaz, 2008, p. 109). They can be used under the assumption that the model predictions are close to the observations (Mandallaz and Massey, 2012), but there is still a risk of underestimating the variance. Pure model-dependent estimators, also known as model-based estimators, assume the existence of a super-population model and require a census of the auxiliary variables of the population (Mandallaz, 2008, p. 98). Model-dependent estimators are particularly suited to remote sensing data since the auxiliary variables are then available for large areas (e.g. Saarela et al., 2016a,b; Breidenbach et al., 2016).

When the response variable is unobserved and the auxiliary variables are only available in the probability sample, which is a common context in forestry, the statistical inferences require hybrid estimators, i.e., estimators that rely on both the probability design and the super-population model. The term “hybrid” was coined by Fattorini (2012) and Corona et al. (2014). According to Ståhl et al. (2016), the lack of recognition of the design-based component of these estimators in studies prior to Corona et al. (2014) had led to different names such as model-based inference (Ståhl et al., 2011), model-dependent inference (Gobakken et al., 2012) and pseudo-synthetic estimator (Mandallaz, 2013). Analytical hybrid estimators have been mainly developed with linear and nonlinear models in contexts of volume or biomass estimation (Ståhl et al., 2011, 2014; Corona et al., 2014; Margolis et al., 2015; Saarela et al., 2015). Here, we want to emphasize that the term “analytical” refers to the fact that these estimators are treated using algebra, in contrast to those based on resampling methods such as the bootstrap estimator (see Efron, 1977; Efron and Tibshirani, 1994). Analytical hybrid estimators have been recently extended to models with heteroscedastic variances (Saarela et al., 2016a) and mixed-effects models (Fortin et al., 2016a), but at the cost of tedious statistical developments.

The analytical hybrid estimators of the variance of the point estimator share the same requirement: the variance induced in the model predictions by the uncertainty in the parameter estimates, the residual error and, eventually, the random effects must be derived or closely approximated using Taylor series. This is possible when the fixed-effects part of the model is expressed as a single equation or as a system of simultaneous equations. However, when the model reaches a higher level of complexity, these analytical derivations or approximations quickly become intractable, hindering the use of current analytical hybrid estimators. An example of “complex” models would be those models that are composed of various submodels, with the predictions of some of them being considered as predictors in others.

These complex models are common in forestry. Tree-level growth models are a good

example of this complexity (e.g. Pretzsch et al., 2002). Some tree- and stand-level models of log grade volumes or lumber product recovery also reach similar levels of complexity (e.g. Schneider, 2007). Basically, when some models are coupled to create decision-support tools (e.g. Kotamaa et al., 2010), the error propagation usually becomes so complex that there is no analytical hybrid estimator of the variance of the point estimator.

An alternative consists of using variance estimators based on resampling methods, which have been widely used in forestry for uncertainty assessment and error propagation (Gertner and Dzialowy, 1984; McRoberts, 1996; Berger et al., 2014). In the context of hybrid inference, a parametric bootstrap estimator of the variance of the point estimator has been used by some authors (McRoberts and Westfall, 2014, 2016; McRoberts et al., 2016), based on the work of Rubin (1987). However, to the best of our knowledge, no study has ever tested the reliability of this variance estimator in the context of hybrid inference. It can be mathematically demonstrated that this estimator tends to overestimate the true variance of the point estimator.

Thus, our objective was to develop a corrected variance estimator, which could be easily used with any model regardless of its complexity. To do this, we built on what is currently known about hybrid inference with linear and nonlinear models, and developed a new bootstrap estimator for the variance of the point estimator.

This paper is structured as follows: it first presents the statistical developments that show the bias of the current bootstrap variance estimator and how this bias can be avoided by using a corrected variance estimator. Secondly, the corrected variance estimator is tested in three simulation studies with models of increasing complexity. Thirdly, the corrected variance estimator is applied in a real-world case study, namely, white birch lumber recovery in the Bas-Saint-Laurent region, Québec, Canada. Finally, the strengths and weaknesses of this corrected variance estimator are discussed.

Material and Methods

Statistical developments

In this paper, we will adopt the notation of Corona et al. (2014): the subscripts d and m stand for design- and model-dependent features, respectively, while hybrid features bear both subscripts. Like the model-dependent inference framework, the hybrid approach assumes that the variable of interest is the realization of a general model that describes the whole population. This model is referred to as a super-population model (Gregoire, 1998). Mathematically speaking, it can be defined as follows:

$$y_i = f(\mathbf{x}_i; \boldsymbol{\beta}) + \epsilon_i \quad (1)$$

where y_i is the variable of interest of population unit i , \mathbf{x}_i is a vector of auxiliary variables, $\boldsymbol{\beta}$ is a vector of unknown model parameters, and ϵ_i is a residual error. Since we assume that $\mathbb{E}_m[\epsilon_i | \mathbf{x}_i] = 0$, the function $f(\mathbf{x}_i; \boldsymbol{\beta})$ is actually the expectation of y_i conditional on the auxiliary variables. For the sake of simplicity, we will define this expectation as $\mathbb{E}_m[y_i | \mathbf{x}_i] = f(\mathbf{x}_i; \boldsymbol{\beta}) \equiv \dot{y}_i$.

Because the auxiliary variables are known in the sample s and not for the whole population, the inferences cannot be purely model-dependent. In fact, they have to rely on the probability design as well, which is known as hybrid inference (Corona et al., 2014; Ståhl et al., 2016). McRoberts et al. (2016) listed the four key features of hybrid inference

as: (i) a probability sample of the auxiliary variables; (ii) a super-population model that can predict the variable of interest in the sample based on the auxiliary variables; (iii) a design-based point estimator of a population parameter, which is usually the mean or the total; and (iv) inferential methods that are based on both the model and the probability design in order to estimate the variance of the point estimator.

In this study, we will focus on the total of the population (τ) and not the mean. As outlined in Gregoire and Valentine (2008, p. 40), the estimator of the total of the population is unbiased under all probability designs, as opposed to the estimator of the mean, and the mean can easily be estimated from the estimate of the total. Throughout this paper, we will also assume that the probability design is one of simple random sampling without replacement (SRSWoR).

The total of the y_i in the N units that compose the population, i.e., $\tau_y = \sum_{i=1}^N y_i$, is the quantity of interest. The HT estimator is an unbiased estimator of this quantity:

$$\hat{\tau}_y = \sum_{i \in s} \frac{y_i}{\pi_i} \quad (2)$$

where π_i is the inclusion probability of unit i . The estimated sampling variance of estimator 2 is (Horvitz and Thompson, 1952):

$$\hat{\mathbb{V}}_d(\hat{\tau}_y) = \sum_{i \in s} \left(\frac{1 - \pi_i}{\pi_i^2} \right) y_i^2 + \sum_{i \in s} \sum_{i' \neq i} \left(\frac{\pi_{ii'} - \pi_i \pi_{i'}}{\pi_i \pi_{i'}} \right) \frac{y_i y_{i'}}{\pi_{i'}} \quad (3)$$

where $\pi_{ii'}$ stands for the joint inclusion probability, i.e., the probability that both units i and i' are part of the sample.

Because the variable y_i is unobserved, we rely on the model in order to obtain its expectation. If we define $\tau_{\hat{y}}$ as the total of the expectations of y_i throughout the population, then:

$$\tau_y = \sum_{i=1}^N \hat{y}_i + \sum_{i=1}^N \epsilon_i = \tau_{\hat{y}} + \sum_{i=1}^N \epsilon_i \quad (4)$$

It must be stressed that τ_y is a random variable even though the parameters of the super-population model are known (Gregoire, 1998). Under the assumption that $\mathbb{E}_m[\epsilon_i | \mathbf{x}_i] = 0$, then the model-dependent variance of τ_y is:

$$\begin{aligned} \mathbb{V}_m(\tau_{\hat{y}} - \tau_y) &= \mathbb{V}_m\left(\sum_{i=1}^N \epsilon_i\right) \\ &= \sum_{i=1}^N \mathbb{V}_m(\epsilon_i) + \sum_{i=1}^N \sum_{i' \neq i}^N \text{COV}_m(\epsilon_i, \epsilon_{i'}) \end{aligned} \quad (5)$$

In practice, the values of the parameters are unknown and have to be estimated from a sample of the population. If this sample is actually a subset of the sample s , the super-population model is said to be internal, as opposed to an external model that would be fitted to observations that are not part of s (Mandallaz, 2013). Since we assume that the response variable is not observed in s , the super-population model is necessarily external. To avoid any confusion between the sample from the probability design and the sample to which the super-population model was fitted, we will systematically refer to the second as the dataset.

Once the parameters are replaced by their estimates, predictions of y_i are obtained as $\hat{y}_i = f(\mathbf{x}_i, \hat{\boldsymbol{\beta}})$. Relying on both the probability design and the model, τ_y can be estimated using the following unbiased hybrid (HYB) estimator:

$$\hat{\tau}_{\text{HYB}} = \sum_{i \in s} \frac{\hat{y}_i}{\pi_i} \quad (6)$$

only under the assumption that the model is correct.

Calculating the variance of the estimator presented in Eq. 6 requires the variances of the model-dependent and the design-based approach to be combined:

$$\mathbb{V}_{md}(\hat{\tau}_{\text{HYB}} - \tau_y) = \mathbb{V}_{md}(\hat{\tau}_{\text{HYB}} - \tau_{\hat{y}}) + \mathbb{V}_m(\tau_{\hat{y}} - \tau_y) \quad (7)$$

The first term on the right-hand side of Eq. 7 accounts for the design variance and the variance due to the parameter estimates. The second term accounts for the residual errors, as shown in Eq. 5. The development of this variance in the context of linear models is shown in the Supplementary Material SM1.

If the function f in Eq. 1 is differentiable, then the variance components in Eq. 7 can be estimated using analytical hybrid estimators (Ståhl et al., 2014, 2016; Fortin et al., 2016a). However, functions that are not differentiable or for which the derivatives can only be calculated with great difficulty considerably hinder the use of these analytical variance estimators. An alternative consists of using a variance estimator based on resampling methods. As we already mentioned, some authors have used a bootstrap variance estimator inspired by Rubin (1987, p. 77) in the context of hybrid inference (e.g. McRoberts and Westfall, 2014, 2016).

Basically, the idea is to generate random deviates for the residual error as well as for the errors in the parameter estimates in order to obtain a realized sample. Let $y_{i,b}$ be the realized value of y_i in the sample s for realization b . By using the HT estimator as defined in Eq. 2 and its variance estimator as defined in Eq. 3 with this sample of $y_{i,b}$, it is possible to obtain the realized point estimate $\hat{\tau}_{y,b}$ and its realized variance estimate $\hat{\mathbb{V}}_d(\hat{\tau}_{y,b})$. After B realizations, the bootstrap (BS) point estimator is obtained by averaging all the realized point estimates:

$$\hat{\tau}_{\text{BS}} = \frac{\sum_{b=1}^B \hat{\tau}_{y,b}}{B} \quad (8)$$

The bootstrap variance estimator is given by:

$$\hat{\mathbb{V}}_{md}(\hat{\tau}_{\text{BS}} - \tau_y) = \frac{\sum_{b=1}^B (\hat{\tau}_{y,b} - \hat{\tau}_{\text{BS}})^2}{B} + \frac{\sum_{b=1}^B \hat{\mathbb{V}}_d(\hat{\tau}_{y,b})}{B} \quad (9)$$

If the number of realizations B is large, the estimator found in Eq. 9 reduces to the law of total variance (see Casella and Berger, 2002, p. 167):

$$\lim_{B \rightarrow \infty} \hat{\mathbb{V}}_{md}(\hat{\tau}_{\text{BS}} - \tau_y) = \hat{\mathbb{V}}_m(\hat{\tau}_{y,b}) + \hat{\mathbb{E}}_m[\hat{\mathbb{V}}_d(\hat{\tau}_{y,b})] \quad (10)$$

These bootstrap estimators can be seen as the average of all the possible realizations in the sample s . Under the assumption that the model is correct, the estimator in Eq. 8 is unbiased. However, it can be shown that the variance estimator in Eq. 9 overestimates the true variance by this quantity (see the Supplementary Material SM2):

$$\begin{aligned} \text{Bias}[\hat{\mathbb{V}}_{md}(\hat{\tau}_{\text{BS}} - \tau_y)] &= 2\text{Tr}(\mathbb{V}_d(\hat{\boldsymbol{\tau}}_x)\mathbb{V}_m(\hat{\boldsymbol{\beta}})) + \sum_{i=1}^N \left(\frac{2}{\pi_i} - 2 \right) \mathbb{V}_m(\epsilon_i) \\ &+ \sum_{i=1}^N \sum_{i' \neq i}^N \left(\frac{2\pi_{ii'} - 2\pi_i\pi_{i'}}{\pi_i\pi_{i'}} \right) \mathbb{C}\mathbb{O}\mathbb{V}_m(\epsilon_i, \epsilon_{i'}) \end{aligned} \quad (11)$$

where $\text{Tr}(\cdot)$ is the trace of a matrix argument and $\boldsymbol{\tau}_x$ is the total of the auxiliary variables across the population.

The first term of the bias is described in Fortin et al. (2016a) and arises from the substitution of estimates for $\mathbb{V}_d(\hat{\boldsymbol{\tau}}_x)$ and $\mathbb{V}_m(\hat{\boldsymbol{\beta}})$. As the sample size increases, the variance $\mathbb{V}_d(\hat{\boldsymbol{\tau}}_x)$ converges to 0, and this first term becomes negligible. In the context of SRSWoR, $\pi_{ii'} \approx \pi_i\pi_{i'}$ for large populations, and the last term of Eq. 11 converges to 0 as well. The overestimation mainly comes from the term involving the variance of ϵ_i , which only cancels out when the population is censused, i.e., $\pi_i = 1$. In all other cases, the result of $\frac{2}{\pi_i} - 2$ will be greater than 0. Moreover, the smaller the inclusion probability is, the greater this overestimation will be. The quantification of this bias is further developed in the Supplementary Material SM3.

The following corrected variance estimator is unbiased (see the Supplementary Material SM2):

$$\begin{aligned} \hat{\mathbb{V}}_{md,\text{CORR}}(\hat{\tau}_{\text{BS}} - \tau_y) &= \hat{\mathbb{V}}_m(\hat{\tau}_{y,b}) + 2\hat{\mathbb{V}}_d(\hat{\tau}_{\bar{y}}) - \hat{\mathbb{E}}_m[\hat{\mathbb{V}}_d(\hat{\tau}_{y,b})] \\ &= \frac{\sum_{b=1}^B (\hat{\tau}_{y,b} - \hat{\tau}_{\text{BS}})^2}{B} + 2\hat{\mathbb{V}}_d(\hat{\tau}_{\bar{y}}) - \frac{\sum_{b=1}^B \hat{\mathbb{V}}_d(\hat{\tau}_{y,b})}{B} \end{aligned} \quad (12)$$

where $\hat{\mathbb{V}}_d(\hat{\tau}_{\bar{y}})$ can be obtained by substituting $\bar{y}_i = \sum_{b=1}^B \frac{y_{i,b}}{B}$ for y_i in the HT variance estimator found in Eq. 3. Like some previously developed analytical or bootstrap estimators (e.g. Breidenbach et al., 2014; McRoberts et al., 2016; Fortin et al., 2016a), the variance estimator presented in Eq. 12 makes it possible to distinguish the sampling-related variance from the model-related variance, which are estimated as $\hat{\mathbb{V}}_d(\hat{\tau}_{\bar{y}})$ and $\hat{\mathbb{V}}_m(\hat{\tau}_{y,b}) + \hat{\mathbb{V}}_d(\hat{\tau}_{\bar{y}}) - \hat{\mathbb{E}}_m[\hat{\mathbb{V}}_d(\hat{\tau}_{y,b})]$, respectively (see the Supplementary Material SM2). The implementation of this new variance estimator is summarized in Fig. 1. For the sake of clarity, we will refer to this new variance estimator as the corrected bootstrap variance estimator, while the estimator found in Eq. 9 will be referred to as the uncorrected bootstrap variance estimator. Both of them are parametric bootstrap estimators as defined in Efron and Tibshirani (1994, p. 53).

While the mathematical developments in the Supplementary Material SM2 provide a corrected bootstrap variance estimator for linear models, there is no certainty that the estimator is still unbiased with more complex models. Since nonlinear models can be linearly approximated using a first-order Taylor series, the corrected bootstrap variance estimator in Eq. 12 should be asymptotically unbiased with these models. For more complex models, we relied on simulation studies, which are detailed in the next section.

Realization	Sampling unit i					A B	
	1	2	3	4	...	Total*	Variance**
1	$y_{1,1}$	$y_{2,1}$	$y_{3,1}$	$y_{4,1}$...	$\hat{\tau}_{y,1}$	$\hat{V}_d(\hat{\tau}_{y,1})$
2	$y_{1,2}$	$y_{2,2}$	$y_{3,2}$	$y_{4,2}$...	$\hat{\tau}_{y,2}$	$\hat{V}_d(\hat{\tau}_{y,2})$
3	$y_{1,3}$	$y_{2,3}$	$y_{3,3}$	$y_{4,3}$...	$\hat{\tau}_{y,3}$	$\hat{V}_d(\hat{\tau}_{y,3})$
\vdots	\vdots	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
B	$y_{1,B}$	$y_{2,B}$	$y_{3,B}$	$y_{4,B}$...	$\hat{\tau}_{y,B}$	$\hat{V}_d(\hat{\tau}_{y,B})$
C	\bar{y}_i^{***}	\bar{y}_1	\bar{y}_2	\bar{y}_3	\bar{y}_4	...	

* The estimates are obtained using the HT estimator in Eq. 2

** The variance estimates are obtained using the HT variance estimator in Eq. 3

*** The \bar{y}_i are calculated as $\bar{y}_i = \sum_{b=1}^B y_{i,b}/B$

Estimate of the total

Calculate $\hat{\tau}_{BS} = \sum_{b=1}^B \hat{\tau}_{y,b}/B$ using the values in column **A**.

Estimate of the variance

Step 1: Calculate $\frac{\sum_{b=1}^B (\hat{\tau}_{y,b} - \hat{\tau}_{BS})^2}{B}$ using the values in column **A**.

Step 2: Calculate $\frac{\sum_{b=1}^B \hat{V}_d(\hat{\tau}_{y,b})}{B}$ using the values in column **B**.

Step 3: Calculate $\sum_{i \in s} \left(\frac{1 - \pi_i}{\pi_i^2} \right) \bar{y}_i^2 + \sum_{i \in s} \sum_{i' \neq i} \left(\frac{\pi_{ii'} - \pi_i \pi_{i'}}{\pi_i \pi_{i'}} \right) \frac{\bar{y}_i \bar{y}_{i'}}{\pi_{ii'}}$ using the values in row **C**.

Step 4: Calculate the variance estimate as the result of Step 1 plus twice the result of Step 3 minus the result of Step 2.

Figure 1: Schematic implementation of the corrected bootstrap variance estimator. The values in the upper part of the table are the realizations $y_{i,b}$.

Simulation studies

The estimators defined in Eqs. 8 and 12 were first tested with a simple linear model. We assumed that the population contained 1000 units. For each unit i , we defined the auxiliary variable x_i , which was attributed a random value uniformly distributed between 2 and 8. The super-population model was set to $y_i = 4 + 3x_i + \epsilon_i$ with $\epsilon_i \sim N(0, \sigma^2 x_i)$ and $\sigma^2 = 2$. We tested three sample sizes, i.e., $n \in 10, 25, 50$. For this linear model, we

proceeded as follows:

- Step 1: The error term ϵ_i was generated for each unit of the population, and the total τ_y was calculated.
- Step 2: Deviates were generated to create parameter estimates under the assumption that their true distribution was:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \sim N \left(\begin{pmatrix} 4 \\ 3 \end{pmatrix}, \begin{pmatrix} 0.02500 & 0.00035 \\ 0.00035 & 0.00050 \end{pmatrix} \right)$$

as if the model had been fitted to an external dataset. An estimated variance-covariance matrix for the parameter estimates and an estimated residual variance ($\hat{\sigma}^2$) were also generated assuming a Wishart (1928) distribution and a χ^2 distribution, respectively, with 98 degrees of freedom in both cases.

- Step 3: A sample of n units was selected following a SRSWoR probability design.
- Step 4: We used the super-population model with the parameter estimates, their estimated variance-covariance matrix and the estimated residual variance obtained in Step 2 to simulate the realized values $y_{i,b}$ in the sample s .
- Step 5: Step 4 was repeated 1000 times, $\hat{\tau}_{\text{BS}}$ was calculated using Eq. 8, and its variance was estimated using the corrected bootstrap variance estimator found in Eq. 12.
- Step 6: Steps 1 to 5 were repeated 10 000 times in order to obtain the empirical distribution of the differences $\hat{\tau}_{\text{BS}} - \tau_y$. The empirical bias of the point estimator $\hat{\tau}_{\text{BS}}$ was calculated as the average of these differences. The average estimated variance was also compared to the empirical variance of the differences to test whether the corrected bootstrap variance estimator was biased or not.

We tested the corrected bootstrap variance estimator in a second and a third simulation study based on two models, both of which predict the volumes of different log grades in individual standing trees. The second simulation study is based on the model of Fortin et al. (2009). This model aims at predicting the volumes of five log grades, namely F1, F2, F3, F4 and pulpwood, in standing trees of sugar maple (*Acer saccharum* Marsh.) and yellow birch (*Betula alleghaniensis* Britton). The first four are sawlog grades where F1 is the best grade and F4 the worst. The fixed-effects part of the model consists of two systems of five equations. The first system predicts the presence of a particular grade in the tree, whereas the second system provides a prediction of the volume of each log grade, conditional on its presence in the tree. Further details about the structure of the model can be found in the Supplementary Material SM4.

We ran this second simulation study with a setup similar to the first one. The population was assumed to contain 1000 plots. Each plot was randomly attributed a discrete number of trees uniformly distributed between two and 22. All the stems were assumed to be sugar maple trees with a diameter at breast height randomly drawn from a uniform distribution between 24 and 50 cm. We ran the simulation according to the steps described above. The super-population model in Step 1 was assumed to be the one fitted by Fortin et al. (2009). In Step 2, realized parameter estimates were drawn from a multivariate Gaussian distribution with the mean and variance-covariance set to those

estimated in the original study. Likewise, the realized estimated variance-covariance matrix of the parameter estimates and the realized estimated residual variances were drawn from Wishart (1928) distributions.

We also included an intermediate step between Step 4 and Step 5, which consisted of aggregating the tree-level realizations in order to obtain a realized value at the population unit level, namely the plot. Note that these plot-level realizations are actually vectors that contain the sum of the volumes per log grade for all of the trees in a particular plot.

The third simulation study was based on the model developed by Schneider (2007). This model predicts the volume of five log grades in standing trees of white birch (*Betula papyrifera* Marsh.). These different log grades are pulpwood, sawlog, low-grade sawlog, veneer and low-grade veneer. For any white birch tree, the model first predicts the height to a small-end diameter of 20 cm as well as the commercial volume, which is defined as the volume from 0.15 m in height up to a small-end diameter of 9 cm. Depending on the diameter at breast height and the visual quality of standing trees, the commercial volume is then split into the different log grades. Some grades are eligible only if the visual quality is good enough and the diameter at breast height is larger than a particular threshold. These features make the model more complex since some components are enabled only if these criteria of quality and diameter are met. The model is described in the Supplementary Material SM5.

Again, we assumed a population of 1000 plots. The number of white birch trees in each plot ranged from two to 22. Each tree was attributed a random diameter uniformly distributed between 18 and 45 cm. We then followed the steps described above. The original model developed by Schneider (2007) was used as a super-population model in Step 1. The realized parameter estimates were randomly drawn from a multivariate Gaussian distribution with the mean and variance-covariance equal to those in the original study. The realized estimated variance-covariance matrix of the parameter estimates and the realized estimated residual variances were also drawn from Wishart (1928) distributions. The intermediate step of aggregation between Step 4 and Step 5 was also required in order to obtain plot-level realizations.

For the sake of the comparison, the uncorrected bootstrap variance estimator found in Eq. 9 was also tested in each simulation study. The average estimated variance obtained with this uncorrected bootstrap variance estimator was also compared to this empirical variance in order to quantify the bias.

Real-world case study

To illustrate the applicability of the corrected bootstrap variance estimator, we used a subset of the provincial forest inventory with the model of Schneider (2007). This inventory programme partly relies on a network of permanent sample plots that were established in the 1970s. In 2003, 393 plots in this network were measured in the Bas-Saint-Laurent region, which is exactly the region for which the model was fitted. At that time, the forested area under survey was 22 681 km² (Parent and Fortin, 2003). Each plot covered an area of 400 m². We assumed a SRSWoR design with an inclusion probability $\forall i : \pi_i = 6.930911 \times 10^{-6}$. The estimation was based on 1000 realizations, like in the simulation studies. A summary of the 393 plots used in this case study is shown in Table 1.

Table 1: Summary of the 393 permanent plots measured in the Bas-Saint-Laurent region in 2003.

	All species	White birch only
Basal area (m^2ha^{-1})		
Minimum	0.0	0.0
Mean	15.4	1.4
Maximum	61.2	18.5
Tree density (trees ha^{-1})		
Minimum	0.0	0.0
Mean	551	71
Maximum	2425	825

Results

The results of the simulation studies are shown in Tables 2, 3 and 4. In most cases, the biases of the point estimator $\hat{\tau}_{\text{BS}}$ were smaller than 1% in absolute value, except for some log grades such as F1 and F4 in the second simulation study (Table 3), as well as for the sawlog, low-grade sawlog and veneer grades in the third simulation study (Table 4). For a particular grade, the biases were constant, regardless of the sample size.

Table 2: Results of the simulation study with the linear model. Relative biases are shown in parentheses.

	Sample size (n)		
	10	25	50
Average difference $\hat{\tau}_{\text{BS}} - \tau_y$	1.7 (0.0%)	6.5 (0.0%)	-2.5 (-0.0%)
Empirical variance	2.74	1.15	0.57
Average uncorrected variance	4.66 (70.2%)	1.93 (67.7%)	0.96 (68.4%)
Average corrected variance	2.71 (-1.1%)	1.14 (-1.0%)	0.58 (1.7%)
Sampling-related	2.66	1.09	0.53
Model-related	0.05	0.05	0.05

Variance estimates ($\times 10^6$)

The corrected bootstrap variance estimator yielded estimates that were always close to the average empirical variances, whereas the uncorrected bootstrap variance estimator overestimated these empirical variances. For the simulation study with the simple linear model, the uncorrected bootstrap variance estimator yielded estimates that were 70% greater than the empirical variances (Table 2). For the two other simulation studies, the uncorrected bootstrap variance estimator overestimated the empirical variance by more than 50% in most cases. The pulpwood grade in the third simulation study was the case where the uncorrected bootstrap variance estimator showed the smallest bias (Table 4). Nevertheless, it still overestimated the empirical variance by 20%. Even in that particular case, the uncorrected bootstrap variance estimator was clearly outperformed by the corrected bootstrap variance estimator.

Table 3: Results of the simulation study with the model of Fortin et al. (2009). Relative biases are shown in parentheses.

	Sample size (n)		
	10	25	50
F1 grade ^a			
Average difference $\hat{\tau}_{BS} - \tau_y$	-9.8 (-4.3%)	-9.6 (-4.3%)	-9.0 (-3.9%)
Empirical variance	4.6	2.8	2.7
Average uncorrected variance	29.2 (535.4%)	12.3 (337.5%)	7.4 (179.3%)
Average corrected variance	4.5 (-1.0%)	2.9 (4.0%)	2.7 (1.7%)
Sampling-related	2.3	0.8	0.4
Model-related	2.3	2.2	2.3
F2 grade ^a			
Average difference $\hat{\tau}_{BS} - \tau_y$	-4.7 (-0.4%)	-1.6 (-0.1%)	-1.7 (-0.1%)
Empirical variance	49.3	20.4	13.5
Average uncorrected variance	112.1 (127.5%)	45.1 (121.5%)	26.2 (94.0%)
Average corrected variance	48.2 (-2.2%)	20.5 (0.3%)	13.7 (1.7%)
Sampling-related	42.8	15.2	8.1
Model-related	5.4	5.2	5.6
F3 grade ^a			
Average difference $\hat{\tau}_{BS} - \tau_y$	0.7 (0.0%)	3.8 (0.2%)	0.9 (0.0%)
Empirical variance	113.4	45.8	25.9
Average uncorrected variance	184.0 (62.2%)	73.4 (60.2%)	39.9 (54.0%)
Average corrected variance	111.0 (-2.1%)	45.1 (-1.7%)	25.6 (-1.2%)
Sampling-related	105.5	39.7	19.9
Model-related	5.5	5.4	5.7
F4 grade ^b			
Average difference $\hat{\tau}_{BS} - \tau_y$	-1.6 (-1.4%)	-1.7 (-1.5%)	-1.9 (-1.6%)
Empirical variance	5.5	3.2	2.7
Average uncorrected variance	29.8 (444.8%)	12.7 (296.3%)	7.4 (177.7%)
Average corrected variance	5.5 (0.6%)	3.2 (-0.0%)	2.6 (-1.6%)
Sampling-related	3.6	1.4	0.7
Model-related	1.9	1.8	1.9
Pulpwood grade ^c			
Average difference $\hat{\tau}_{BS} - \tau_y$	5.4 (0.1%)	9.9 (0.3%)	8.0 (0.2%)
Empirical variance	41.7	15.8	8.7
Average uncorrected variance	54.2 (30.1%)	20.9 (32.1%)	11.1 (27.1%)
Average corrected variance	40.7 (-2.4%)	15.7 (-1.0%)	8.5 (-3.2%)
Sampling-related	39.8	14.8	7.5
Model-related	0.9	0.9	0.9

^a Variance estimates ($\times 10^3$)^b Variance estimates ($\times 10^2$)^c Variance estimates ($\times 10^4$)

The sampling-related variance, as estimated through the corrected bootstrap variance estimator, decreased with increasing sample sizes, whereas the model component remained approximately constant. With the smallest sample size, the sampling-related variances were larger than the model-related ones in most cases. There were some exceptions such as

the F1 grade and the veneer grade in the second and third simulation studies, respectively. For the largest sample size, there was no clear pattern.

Table 4: Results of the simulation study with the model of Schneider (2007). Relative biases are shown in parentheses.

	Sample size (n)		
	10	25	50
Pulpwood^a			
Average difference $\hat{\tau}_{BS} - \tau_y$	42.0 (0.9%)	42.8 (0.9%)	40.3 (0.9%)
Empirical variance	60.3	26.3	15.0
Average uncorrected variance	77.9 (29.1%)	32.5 (23.7%)	18.0 (20.3%)
Average corrected variance	61.7 (2.2%)	26.4 (0.4%)	15.0 (0.1%)
Sampling-related	57.5	22.4	11.0
Model-related	4.2	4.0	4.0
Sawlog^a			
Average difference $\hat{\tau}_{BS} - \tau_y$	-68.2 (-5.8%)	-63.2 (-5.6%)	-67.2 (-5.9%)
Empirical variance	8.4	5.5	4.6
Average uncorrected variance	22.8 (170.3%)	11.0 (100.0%)	7.4 (61.0%)
Average corrected variance	9.0 (7.1%)	5.7 (3.8%)	4.8 (4.3%)
Sampling-related	5.0	1.9	0.9
Model-related	4.0	3.8	3.9
Low-grade sawlog^b			
Average difference $\hat{\tau}_{BS} - \tau_y$	-3.3 (-3.4%)	-3.4 (-3.5%)	-3.1 (-3.3%)
Empirical variance	9.5	6.4	4.7
Average uncorrected variance	28.6 (201.3%)	13.9 (116.5%)	8.2 (75.5%)
Average corrected variance	9.8 (3.0%)	6.5 (1.7%)	4.7 (1.9%)
Sampling-related	6.0	2.6	1.3
Model-related	3.8	3.9	3.5
Veneer^c			
Average difference $\hat{\tau}_{BS} - \tau_y$	28.0 (6.9%)	26.8 (6.8%)	25.9 (6.4%)
Empirical variance	19.5	14.2	13.7
Average uncorrected variance	88.0 (352.0%)	40.4 (185.1%)	27.2 (99.0%)
Average corrected variance	20.8 (6.9%)	15.2 (7.1%)	14.6 (6.7%)
Sampling-related	7.5	2.9	1.5
Model-related	13.3	12.3	13.1
Low-grade veneer^c			
Average difference $\hat{\tau}_{BS} - \tau_y$	0.9 (0.1%)	1.0 (0.1%)	2.9 (0.3%)
Empirical variance	51.8	28.3	21.7
Average uncorrected variance	86.6 (67.0%)	41.3 (46.2%)	29.2 (34.6%)
Average corrected variance	53.1 (2.5%)	28.9 (2.3%)	22.9 (5.6%)
Sampling-related	37.3	14.3	7.3
Model-related	15.8	14.6	15.6

^a Variance estimates ($\times 10^4$)

^b Variance estimates ($\times 10^2$)

^c Variance estimates ($\times 10^3$)

Regarding the case study, the pulpwood log grade was estimated to be the most abun-

dant one, followed by the sawlog, the low-grade sawlog, the low-grade veneer and, finally, the veneer grade (Table 5). The uncorrected bootstrap variance estimator systematically yielded larger estimates than the corrected bootstrap variance estimator. The variance component that largely predominated was the sampling, which represented more than 80% of the variance for all grades except the low-grade sawlog (Fig. 2).

Table 5: Estimates of total volumes (m^3) per log grades for white birch in the Bas-Saint-Laurent region in 2003 based on the hybrid estimator and the model of Schneider (2007). Relative differences between the uncorrected and the corrected bootstrap variance estimators are shown in parentheses.

Estimate	Pulpwood	Sawlog	Low-grade sawlog	Veneer	Low-grade veneer
Point estimate ($\times 10^3$)	7 492	1 573	1 119	93	160
Uncorrected variance	1 026 708 (6.7%)	111 757 (36.7%)	60 333 (27.5%)	3 815 (163.4%)	8 530 (86.2%)
Corrected variance	962 575	81 733	47 324	1 449	4 582
Model-related	92 864	13 811	20 880	111	131
Sampling-related	869 711	67 922	26 444	1 338	4 451
Variance estimates ($\times 10^6$)					

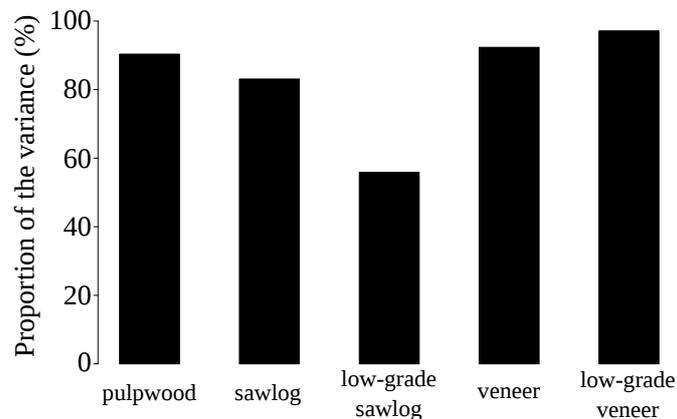


Figure 2: Contribution of the sampling-related variance to the total variance of the estimator of the total volume for the different log grades in the 2003 inventory in the Bas-Saint-Laurent region.

Discussion

Bias

In the context of hybrid inference, it has always been a challenge to merge the variance that stems from the probability design with that of the super-population model. It usually results in complex mathematical developments, even though the variances induced by the uncertainty in the parameter estimates, the random effects and the residual error in the model predictions can be explicitly formulated (Ståhl et al., 2011; Fortin et al., 2016a).

However, in some circumstances, the super-population model can be more complex than those that were dealt with in previous studies on hybrid inference. This complexity arises when the model includes several components with eventual logical choices based on intermediate predicted values. In such a context, the contribution of the uncertainty in the parameter estimates, the random effects and the residual error to the variance of the point estimator can rarely be derived or approximated, which hampers the use of current analytical variance estimators.

A common method to overcome this problem consists of using estimators based on resampling methods. A variance estimator based on a parametric bootstrap, which we described earlier as the uncorrected bootstrap variance estimator, has already been used in the context of hybrid inference. Our statistical developments show that this variance estimator overestimates the true variance of the point estimator when the super-population model is linear. Our second and third simulation studies confirmed that this overestimation also exists when the hybrid estimator is based on a more complex super-population model.

The major part of the bias of the uncorrected bootstrap variance estimator arises from the second term on the right-hand side of Eq. 11, namely $\sum_{i=1}^N \left(\frac{2}{\pi_i} - 2 \right) \mathbb{V}_m(\epsilon_i)$ (see the Supplementary Material SM3). It turns out that there are three factors that affect the magnitude of this bias: the population size, the sample size and the average residual variance of the super-population model (see the Supplementary Material SM3). This bias never tends to 0 unless the whole population is censused, i.e., $n = N$. As a matter of fact, this bias will increase along with the population size. If we focus on the first simulation study, where $\mathbb{V}_m(\epsilon_i) = 2x_i$ and $\forall i : \pi_i = \pi$, this bias can be approximated to 1.98×10^6 . This approximation is close to the difference of 1.92×10^6 that we obtained between the average uncorrected variance estimate and the empirical variance (Table 2). Repeating the same calculation for $n = 25$ and $n = 50$ also yields values that are very close to the observed differences.

Bias estimation

The unknown variance $\mathbb{V}_m(\epsilon_i)$ can be replaced by its estimate in order to obtain a feasible correction factor. If the variance $\mathbb{V}_m(\epsilon_i)$ is heterogeneous, then the total $\sum_{i=1}^N \mathbb{V}_m(\epsilon_i)$ can also be estimated using a HT estimator (see Fortin et al., 2016a). Nevertheless, the implementation of such a correction factor is more tedious than implementing the corrected bootstrap variance estimator (Fig. 1).

Many studies address the estimate of the mean and not of the total. The mean estimate and its estimated variance are easily obtained by dividing estimators 8 and 12 by N and N^2 , respectively (Gregoire and Valentine, 2008, p. 40). Whether it is the total or the mean of the population, it can be shown that the relative bias of the uncorrected bootstrap variance estimator is the same for both quantities (see the Supplementary Material SM3). This relative bias is proportional to the average residual variance of the super-population model. It is also independent of the population size. There are two patterns that can be distinguished. The first one is when the model-related variance is large compared to the sampling-related variance, in which case a large sample size then induces a decrease in the relative bias. The second pattern occurs when the sampling-related variance is found to be much greater than the model-related variance. In such cases, the decrease in the relative bias induced by an increasing sample size remains marginal (see Eq. S18 in the Supplementary Material SM3).

Both patterns can be found in the simulation studies. For instance, the first simulation study clearly belongs to the second pattern: the sampling-related variance was much greater than the model-related variance and the relative bias did not decrease with increases in the sample size (Table 2). The F1 log grade in the second simulation study is a typical case of the first pattern: the model-related variance was greater than the sampling-related variance and the relative bias decreased as the sample size increased (Table 3).

In practice, the first pattern is rather uncommon with the exception of the study of Fortin et al. (2016a). In most cases, the sampling-related variance is much greater than the model-related variance (McRoberts and Westfall, 2014; Ståhl et al., 2014; Fortin et al., 2016b) and the relative bias can be roughly approximated from the coefficient of determination (R^2) of the super-population model as follows (see the Supplementary Material SM3):

$$\frac{\text{Bias}[\hat{\mathbb{V}}_{md}(\hat{\tau}_{\text{BS}})]}{\mathbb{V}_{md}(\hat{\tau}_{\text{BS}})} \approx \frac{2(1 - R^2)}{R^2} \quad (13)$$

In the first simulation study, the R^2 was 0.7297 and the relative bias could be approximated to 74.1%, which is close to the observed relative biases that ranged from 67.7 to 70.2% (Table 2). Note that this approximation does not directly apply when the super-population model provides tree-level predictions because this R^2 should be calculated at the population unit level, namely the plot. Nevertheless, this approximation indicates that there would still be a relative bias of about 10%, even though the super-population model has a R^2 of 0.95. We ran the first simulation study again with a smaller residual variance in order to have a super-population model with a R^2 close to 0.95 and we obtained relative biases that ranged from 9.9% to 13.8%.

In the context of volume estimation, Fortin and DeBlois (2010) found that the standard deviation of the prediction errors represented approximately 15% of the plot-level volume predictions. This level of precision leads to a R^2 of 0.978. Using the approximation shown in Eq. 13, the bias of the uncorrected bootstrap variance estimator can be estimated at 4.5%, which is probably not a great concern. However, there is no guarantee that such precise super-population models are the rule.

Ståhl et al. (2016) suggested that the residual variance of the super-population model makes a relatively small contribution to the total variance, which was empirically demonstrated by some authors (Breidenbach et al., 2014; Fortin et al., 2016a,b). If the residual error is omitted in the uncorrected bootstrap variance estimator, the bias can be greatly reduced. While this option seems to be a valid alternative for linear and nonlinear models, which are the most common situation, it may not be satisfactory for complex models. Some authors have shown that complex models may yield predictions that are significantly different depending on the stochastic components that are enabled (Higgins et al., 1997; Zhou and Buongiorno, 2004; Fortin and Langevin, 2012), which is essentially due to Jensen's (1906) inequality. With such models, omitting the residual error terms could induce a bias in the point estimator of the total or the mean.

In all three simulation studies, our corrected bootstrap variance estimator yielded estimates that were very close to the empirical variances. These results support the conclusion that this corrected bootstrap variance estimator, which was developed in the context of linear models, also works with complex models. The implementation of this corrected bootstrap variance estimator is not much more complicated than that of the uncorrected bootstrap variance estimator (Fig. 1). It only requires an additional term to

be calculated, i.e., $\hat{\mathbb{V}}_d(\hat{\tau}_{\bar{y}})$. The calculation of the \bar{y}_i implies that the realizations $y_{i,b}$ are saved in the memory until the B realizations have been generated. In all our simulations, that additional memory requirement was not an issue.

Sampling vs model-related variance

Like some previous hybrid estimators, this corrected bootstrap variance estimator makes it possible to distinguish the variance that is induced by the sampling from that of the model. While the sampling variance decreased with increasing sample sizes, the model-related variance remained stable. This result was also highlighted in Fortin et al. (2016a). In reality, this model-related variance is only slightly impacted by the sample size through term $\text{Tr}(\mathbb{V}_d(\hat{\tau}_x)\mathbb{V}_m(\hat{\beta}))$ (see the Supplementary Material SM1), which is relatively small compared to terms $\beta^T \mathbb{V}_d(\hat{\tau}_x)\beta$ and $\tau_x \mathbb{V}_m(\hat{\beta})\tau_x^T$. Even though a census of the whole population was taken, the variance related to the model remained because the responses are not observed but only predicted. The only way to reduce this variance component is to fit a more accurate super-population model, which is possible by either adding explanatory variables in the model or increasing the size of the dataset to which the model is fitted.

In our case study, the distinction between these two sources of variance clearly indicated that the sampling-related variance was by far the most important source of variability. For most grades, it accounted for more than 80% of the total variance of the estimate (Fig. 2). It clearly appears that any effort to increase the precision of the estimates should first focus on the sampling.

This large sampling variance was due to the scarcity of white birch trees in the sample. Many plots actually had no white birch at all, whereas the species accounted for a significant part of the basal area and tree density in some others. A simple way to reduce this sampling-related variance would be to use a stratified design instead of a simple random design like the one we considered here. If the strata that have a higher proportion of white birch trees could be delineated on the map, the overall sampling-related variance could be decreased. This sampling scheme is largely described in Gregoire and Valentine (2008, Ch. 5).

Hybrid estimators are both model-dependent and design-based. Being model-dependent, they rely on the usual distributional assumptions related to model fitting (Gregoire, 1998). If these assumptions hold, the estimator of the variance-covariance matrix is usually unbiased or, at least, asymptotically unbiased. With linear models, the variability of the estimated variance-covariance matrix should not be a concern. With more complex models that include nonlinear sub models with strong curvature, the corrected bootstrap variance estimator could be biased. For further details on this curvature issue, the reader is referred to Seber and Wild (2003, Ch. 12). In such a case, the distribution of the parameter estimates could be better approximated by refitting the super-population model to bootstrapped samples of the original dataset, as suggested in McRoberts and Westfall (2014, 2016). This however assumes that this original dataset is still available. This approach remains to be tested.

Confidence intervals

It can be reasonably assumed that the uncertainty of the estimated variance-covariance matrix also impacts the coverage of the confidence intervals. When working in a design-based framework, confidence intervals are based on Student's t distribution instead of a

Gaussian distribution. Student's t distribution was designed to take the fact that the true variance remains unknown into account and that only an estimate is available. The corrected bootstrap variance estimator in Eq. 12 implies the estimation of three variance components, i.e., $\hat{\mathbb{V}}_m(\hat{\tau}_{y,b})$, $\hat{\mathbb{V}}_d(\hat{\tau}_{\bar{y}})$ and $\hat{\mathbb{E}}_m[\hat{\mathbb{V}}_d(\hat{\tau}_{y,b})]$. Student's t distribution cannot be used because the degrees of freedom in the first term are probably different from those of the other two. An approximation of the total degrees of freedom could be calculated using the Welch-Satterthwaite equation (Satterthwaite, 1946). The empirical coverage of confidence intervals based on this approximation should be further investigated. In the meantime, confidence intervals based on a normal distribution should achieve their nominal coverage only if both the dataset used to fit the super-population model and the sample are large.

Another possibility that we did not test is that of spatial correlations among the residual errors of the super-population model. In all our simulation studies, the term $\text{COV}_m(\epsilon_i, \epsilon_{i'})$ was assumed to be null. Theoretically, the corrected bootstrap variance estimator is unbiased even if these covariances are not null. However, this should be empirically tested in future simulation studies.

Practical implications

In terms of practical implications, this corrected bootstrap variance estimator opens the way for large-area growth predictions using tree-level models. As we mentioned in the introduction, these models are complex by nature. They include nonlinear components and they rely on an iterative process in which the predictions are reinserted in the model in order to obtain long-term projections. When using this type of model for large-area predictions, there exist two strategies of upscaling. The first one consists of averaging the input data, which is known as the simple averaging method (Wu and Li, 2006). However, with this method, the large-area predictions can be biased because the nonlinear components of the model are likely affected by Jensen's (1906) inequality.

The second strategy is known as the direct extrapolation method (Wu and Li, 2006) and it makes it possible to avoid the prediction issue due to Jensen's inequality. Each plot is simulated individually and the predictions are then averaged or summed in order to obtain a prediction for the whole population, which is a setup identical to that of this study. In order to obtain a reliable estimate of the variance, the options are limited. Using the uncorrected bootstrap variance estimator leads to overestimating the variance of the point estimator. Omitting the residual error term would reduce the bias in the uncorrected bootstrap variance estimator, but it would induce a bias in the point estimator (see Fortin and Langevin, 2012). If the tree-level growth model benefits from a full stochastic implementation, the prediction errors can be propagated as well as the sampling error using the corrected bootstrap variance estimator of this study, which is, to the best of our knowledge, the only reliable option in this context.

Conclusions

On the basis of this study, we can draw the following conclusions:

1. A variance estimator based on a parametric bootstrap and the law of total variance tends to overestimate the true variance in contexts of hybrid inference.

2. When working with the estimate of the total, the quantity $\sum_{i=1}^N \left(\frac{2}{\pi_i} - 2\right) \mathbb{V}_m(\epsilon_i)$ accounts for the major part of this bias. If $\forall i : \pi_i = \pi$, this quantity can be estimated using a HT estimator. The population size, the sample size and the average residual variance of the super-population model affect the magnitude of the bias. This bias never tends to 0 unless the whole population is censused, i.e., $n = N$, or the residual variances $\mathbb{V}_m(\epsilon_i)$ tend to 0.
3. Whether it is the total or the mean of the population, the relative bias of this uncorrected bootstrap variance estimator is the same for both quantities. This relative bias is independent of the population size and it is rather insensitive to the sample size when the model-related variance is much smaller than the sampling-related variance, which is often the case. In such a context, the relative bias can be roughly estimated from the coefficient of determination (R^2) of the super-population model as follows: $2(1 - R^2)/R^2$.
4. If the residual error is omitted under the assumption that the residual variance has a small contribution to the total variance of the point estimator, the bias of the uncorrected bootstrap variance estimator is greatly reduced. This works well with simple linear or nonlinear models as in most situations. However, when dealing with a complex model, there is a risk of causing a bias in the point estimator, which is essentially due to Jensen's (1906) inequality.
5. In comparison, the corrected bootstrap variance estimator in Eq. 12 is unbiased and its implementation only requires the computation of an additional term once the B realizations have been generated (Fig. 1). In all our simulation studies, the bootstrap point estimator and the corrected bootstrap variance estimator were based on 1000 realizations, which ensured the stabilization of the total and variance estimates.
6. The corrected bootstrap variance estimator works even though the aggregation of the predictions is required. Even if the model applies at a lower level than the population unit, this corrected bootstrap variance estimator showed no evidence of bias.
7. Like some other variance estimators, the corrected bootstrap variance estimator also makes it possible to distinguish the model-related variance from the sampling-related variance. As shown in our case study, it was possible to identify the most important source of variability and to focus on the methods that are the most likely to increase the precision of the estimates.

From a theoretical point of view, this corrected bootstrap variance estimator can apply to any super-population model in the context of hybrid inference as long as its nonlinear components do not exhibit strong curvature (see Seber and Wild, 2003, Ch. 12). However, it should be stressed that hybrid inference is tied to the assumptions underlying the model. As a consequence, the estimator is not robust to model misspecification and overparameterisation. Moreover, the estimator requires that all the stochastic components of the super-population model be properly simulated.

In summary, we make the following recommendations. If the super-population model is linear or nonlinear and it applies at the plot level, then use some analytical hybrid variance estimators such as those in Ståhl et al. (2011, 2016) and Fortin et al. (2016a). If the super-population model is linear or nonlinear but it applies at a lower hierarchical

level, analytical hybrid variance estimators can still be used (Fortin et al., 2016a). If the super-population model is hardly differentiable or not differentiable at all, then use the corrected bootstrap variance estimator of this study, which is also a valid alternative to all of the cases mentioned above, although it implies a greater computational burden and it assumes that all the stochastic components of the super-population model can be properly simulated.

Funding

This work was supported by a grant overseen by the French National Research Agency (ANR) as part of the “Investissements d’Avenir” programme (ANR-11-LABX-0002-01, Lab of Excellence ARBRE). The development of the models found in Schneider (2007) and Fortin et al. (2009) was funded by the Quebec Ministry of Forests, Wildlife and Parks (*Ministère des Forêts, de la Faune et des Parcs du Québec*). This study is part of the WOHTOC project funded by LabEx ARBRE.

Acknowledgements

Thanks to Ronald McRoberts (USDA Forest Service), two anonymous reviewers and Dave Auty (Northern Arizona University) for their constructive comments on a preliminary version of this paper. The permanent-plot data were provided by the Forest inventory branch (*Direction des inventaires forestiers*) of the Quebec Ministry of Forests, Wildlife and Parks (*Ministère des Forêts, de la Faune et des Parcs du Québec*).

References

- Berger, A., Gschwantner, T., McRoberts, R. E., and Schadauer, K. (2014). Effects of measurement errors on individual tree stem volume estimates for the Austrian national forest inventory. *Forest Science*, 60(1):14–24.
- Breidenbach, J., Antón-Fernández, C., Petersson, H., McRoberts, R. E., and Astrup, R. (2014). Quantifying the model-related variability of biomass stock and change estimates in the Norwegian national forest inventory. *Forest Science*, 60(1):25–33.
- Breidenbach, J., McRoberts, R. E., and Astrup, R. (2016). Empirical coverage of model-based variance estimators for remote sensing assisted estimation of stand-level timber volume. *Remote Sensing of Environment*, 173:274–281.
- Casella, G. and Berger, R. L. (2002). *Statistical inference*. Duxbury, Thomson Learning, Pacific Grove, California, USA.
- Corona, P., Fattorini, L., Franceschi, S., Scrinzi, G., and Torresan, C. (2014). Estimation of standing wood volume in forest compartments by exploiting airborne laser scanning information: model-based, design-based, and hybrid perspectives. *Canadian Journal of Forest Research*, 44:1303–1311.
- Efron, B. (1977). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7:1–26.
- Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Chapman & Hall/CRC, Boca Raton, FL, USA.
- Fattorini, L. (2012). *Studies in Honor of Claudio Scala*, chapter Design-based or model-based inference? The role of hybrid approaches in environmental surveys, pages 173–214. Department of Economics and Statistics, University of Siena.
- Fortin, M. and DeBlois, J. (2010). A statistical estimator to propagate height prediction errors into a general volume model. *Canadian Journal of Forest Research*, 40:1930–1939.
- Fortin, M., Guillemette, F., and Bédard, S. (2009). Predicting volumes by log grades in standing sugar maple and yellow birch trees in southern Quebec, Canada. *Canadian Journal of Forest Research*, 39:1928–1938.
- Fortin, M. and Langevin, L. (2012). Stochastic or deterministic single-tree models: is there any difference in growth prediction? *Annals of Forest Science*, 69:271–282.
- Fortin, M., Manso, R., and Calama, R. (2016a). Hybrid estimation based on mixed-effects models in forest inventories. *Canadian Journal of Forest Research*, 46:1310–1319.
- Fortin, M., Robert, N., and Manso, R. (2016b). Uncertainty assessment of large-scale forest growth predictions based on a transition-matrix model in Catalonia. *Annals of Forest Science*, 73:871–883.
- Gertner, G. and Dzialowy, P. J. (1984). Effects of measurement errors on an individual tree-based growth projection system. *Canadian Journal of Forest Research*, 14:311–316.

- Gobakken, T., Næsset, E., Nelson, R., Bollandsås, O. M., Gregoire, T. G., Ståhl, G., Holm, S., Ørka, H. O., and Astrup, R. (2012). Estimating biomass in Hedmark County, Norway using national forest inventory field plots and airborne laser scanning. *Remote Sensing of Environment*, 123:443–456.
- Gregoire, T. G. (1998). Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research*, 28:1429–1447.
- Gregoire, T. G. and Valentine, H. T. (2008). *Sampling techniques for natural and environmental resources*. Chapman & Hall/CRC, Boca Raton, Florida, USA.
- Higgins, K., Hasting, A., Sarvela, J. N., and Botsford, L. W. (1997). Stochastic dynamics and deterministic skeletons: Population behavior of dungeness crab. *Science*, 276:1431–1435.
- Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47:663–685.
- Jensen, J. L. W. V. (1906). Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30:175–193.
- Kotamaa, E., Tokola, T., Maltamo, M., Packalén, P., Jurtilla, M., and Mäkinen, A. (2010). Integration of remote-based bioenergy inventory data and optimal bucking for stand-level decision making. *European Journal of Forest Research*, 129:875–886.
- Mandallaz, D. (2008). *Sampling techniques for forest inventories*. Chapman & Hall/CRC, London.
- Mandallaz, D. (2013). Design-based properties of some small-area estimators in forest inventory with two-phase sampling. *Canadian Journal of Forest Research*, 43:441–449.
- Mandallaz, D. and Massey, A. (2012). Comparison of estimators in one-phase two-stage Poisson sampling in forest inventories. *Canadian Journal of Forest Research*, 42:1865–1871.
- Margolis, H. A., Nelson, R. F., Montesano, P. M., Beaudoin, A., Sun, G., Andersen, H.-E., and Wulder, M. A. (2015). Combining satellite lidar, airborne lidar and ground plots to estimate the amount and distribution of aboveground biomass in the boreal forest of North America. *Canadian Journal of Forest Research*, 45:838–855.
- McRoberts, R. E. (1996). Estimating variation in field crew estimates of site index. *Canadian Journal of Forest Research*, 26:560–565.
- McRoberts, R. E., Chen, Q., Domke, G. M., Ståhl, G., Saarela, S., and Westfall, J. A. (2016). Hybrid estimators for mean aboveground carbon per unit area. *Forest Ecology and Management*, 378:44–56.
- McRoberts, R. E. and Westfall, J. A. (2014). Effects of uncertainty in model predictions of individual tree volume on larger area volume estimates. *Forest Science*, 60(1):34–42.
- McRoberts, R. E. and Westfall, J. A. (2016). Propagating uncertainty through individual tree volume model predictions to large-area volume estimates. *Annals of Forest Science*, 73:625–633.

- Parent, B. and Fortin, C. (2003). Ressources et industries forestières – Portrait statistique Edition 2003. Technical report, Ministère des Ressources naturelles, de la Faune et des Parcs du Québec, Direction du développement de l'industrie des produits forestiers.
- Pretzsch, H., Biber, P., and Ďurský, J. (2002). The single tree-based stand simulator SILVA: construction, application and evaluation. *Forest Ecology and Management*, 162:3–21.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse surveys*. John Wiley & Sons, Hoboken, New Jersey, USA.
- Saarela, S., Holm, S., Grafström, A., Schnell, S., Næsset, E., Gregoire, T. G., Nelson, R. F., and Ståhl, G. (2016a). Hierarchical model-based inference for forest inventory utilizing three sources of information. *Annals of Forest Science*, 73:895–910.
- Saarela, S., Schnell, S., Grafström, A., Tuominen, S., Nordkvist, K., Hyypä, J., Kangas, A., and Ståhl, G. (2015). Effects of sample size and model form on the accuracy of model-based estimators of growing stock volume. *Canadian Journal of Forest Research*, 45:1524–1534.
- Saarela, S., Schnell, S., Tuominen, S., Balázs, A., Hyypä, Grafström, A., and Ståhl, G. (2016b). Effects of positional errors in model-assisted and model-based estimation of growing stock volume. *Remote Sensing of Environment*, 172:101–108.
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2:110–114.
- Schneider, R. (2007). Matrice de répartition par produits du bouleau à papier du Bas-Saint-Laurent. Technical report, Ministère des Ressources naturelles et de la Faune du Québec.
- Seber, G. A. F. and Wild, C. J. (2003). *Nonlinear regression*. John Wiley & Sons, Hoboken, New Jersey, USA.
- Ståhl, G., Heikkinen, J., Petersson, H., Repola, J., and Holm, S. (2014). Sample-based estimation of greenhouse gas emissions from forests - a new approach to account for both sampling and model errors. *Forest Science*, 60(1):3–13.
- Ståhl, G., Holm, S., Gregoire, T. G., Gobakken, T., Næsset, E., and Nelson, R. (2011). Model-based inference for biomass estimation in a LiDAR sample survey in Hedmark County, Norway. *Canadian Journal of Forest Research*, 41:96–107.
- Ståhl, G., Saarela, S., Schnell, S., Holm, S., Breidenbach, J., Healey, S. P., Patterson, P. L., Magnussen, S., Næsset, E., McRoberts, R. E., and Gregoire, T. G. (2016). Use of models in large-area forest surveys: comparing model-assisted, model-based and hybrid estimation. *Forest Ecosystems*, 3:5.
- Wishart, J. (1928). The generalized product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A:32–52.
- Wu, J. and Li, H. (2006). *Scaling and uncertainty analysis in ecology*, chapter Perspectives and methods of scaling, pages 17–44. Springer, Dordrecht, The Netherlands.

Zhou, M. and Buongiorno, J. (2004). Nonlinearity and noise interaction in a model of forest growth. *Ecological Modelling*, 180:291–304.