Parametric bootstrap estimators for hybrid inference in forest inventories Supplementary Material

Mathieu Fortin[1,*], Rubén Manso[1,3], and Robert Schneider[2]

[1]INRA/AgroParisTech, UMR 1092 LERFoB, 54000 Nancy, France.
[2]Département de Biologie, Chimie et Géographie, , Université du Québec à Rimouski, Rimouski, Canada.
[3]Current address: Forestry Research, Northern Research Station, Forestry Commission, Roslin, Midlothian, United Kingdom
[*]Corresponding author: mathieu.fortin@agroparistech.fr

22 Nov 2017

SM1 Variance of $\hat{\tau}_{\text{HYB}}$ based on linear models

Let us consider a linear model such that $y_i = \boldsymbol{x}_i \boldsymbol{\beta} + \epsilon_i$. Then, $\hat{\tau}_{\text{HYB}} = \hat{\boldsymbol{\tau}}_x \hat{\boldsymbol{\beta}}$ and $\tau_{\dot{y}} = \boldsymbol{\tau}_x \boldsymbol{\beta}$, where $\boldsymbol{\tau}_x$ is a vector that contains the totals of the auxiliary variables in the population. The variance of $\hat{\tau}_{\text{HYB}}$ can be expressed as:

$$
\begin{aligned}
\mathbb{V}_{md}(\hat{\tau}_{\text{HYB}} - \tau_y) &= \mathbb{V}_{md}(\hat{\tau}_{\text{HYB}} - \tau_{\dot{y}}) + \mathbb{V}_m(\tau_{\dot{y}} - \tau_y) \\
&= \boldsymbol{\beta}^T \mathbb{V}_d(\hat{\boldsymbol{\tau}}_x)\boldsymbol{\beta} + \boldsymbol{\tau}_x \mathbb{V}_m(\hat{\boldsymbol{\beta}})\boldsymbol{\tau}_x^T + \text{Tr}(\mathbb{V}_d(\hat{\boldsymbol{\tau}}_x)\mathbb{V}_m(\hat{\boldsymbol{\beta}})) \\
&\quad + \sum_{i=1}^{N} \mathbb{V}_m(\epsilon_i) + \sum_{i=1}^{N}\sum_{i' \neq i}^{N} \mathbb{COV}_m(\epsilon_i, \epsilon_{i'})
\end{aligned}
\tag{S1}
$$

where $\text{Tr}(\cdot)$ is the trace of the matrix argument and N is the population size. The first three terms are actually the variance of the product of the two independent random variables $\hat{\boldsymbol{\tau}}_x \hat{\boldsymbol{\beta}}$, which is well known in statistics (see Goodman, 1960). The last two terms come from the residual errors of the super-population model. Similar mathematical development applied to the forestry context can be found in Ståhl et al. (2016) and Fortin et al. (2016a).

SM2 Bias of the variance estimator of $\hat{\tau}_{\text{BS}}$

Throughout the following mathematical developments, we will assume that the number of realizations B tends to infinity. Let us consider a linear model such that $y_i = \boldsymbol{x}_i \boldsymbol{\beta} + \epsilon_i$,

with $\epsilon_i \sim N(0, \sigma^2)$. Since the vector $\boldsymbol{\beta}$ is unknown, it is replaced by its estimate $\hat{\boldsymbol{\beta}}$, which is assumed to follow a multivariate normal distribution, i.e., $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \mathbb{V}_m(\hat{\boldsymbol{\beta}}))$. For each realization, random deviates are drawn to account for the residual errors and the errors in the parameter estimates. These deviates would be drawn from the two normal distributions mentioned above if $\boldsymbol{\beta}$, $\mathbb{V}_m(\hat{\boldsymbol{\beta}})$ and σ^2 were known. To obtain a feasible estimator, the parameters of these distributions are replaced by their estimates. If we define $\boldsymbol{\epsilon_\beta} = \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ as the errors in the parameter estimates such that $\boldsymbol{\epsilon_\beta} \sim N(\mathbf{0}, \hat{\mathbb{V}}_m(\hat{\boldsymbol{\beta}}))$, then:

$$y_{i,b} = \boldsymbol{x}_i\hat{\boldsymbol{\beta}} + \boldsymbol{x}_i\boldsymbol{\epsilon}_{\boldsymbol{\beta},b} + \epsilon_{i,b} \tag{S2}$$

The variance estimator in Eq. 9 can be re-expressed as:

$$\hat{\mathbb{V}}_{md}(\hat{\tau}_{\text{BS}} - \tau_y) = \hat{\mathbb{V}}_m(\hat{\tau}_{y,b}) + \hat{\mathbb{E}}_m[\hat{\mathbb{V}}_d(\hat{\tau}_{y,b})]$$
$$= \hat{\mathbb{V}}_m\left(\sum_{i \in s} \frac{y_{i,b}}{\pi_i}\right) + \hat{\mathbb{E}}_m\left[\sum_{i \in s}\left(\frac{1 - \pi_i}{\pi_i^2}\right)y_{i,b}^2 + \sum_{i \in s}\sum_{i' \neq i}\left(\frac{\pi_{ii'} - \pi_i\pi_{i'}}{\pi_i\pi_{i'}}\right)\frac{y_{i,b}y_{i',b}}{\pi_{ii'}}\right] \tag{S3}$$

The term $\boldsymbol{x}_i\hat{\boldsymbol{\beta}} + \boldsymbol{x}_i\boldsymbol{\epsilon}_{\boldsymbol{\beta},b} + \epsilon_{i,b}$ can be substituted for $y_{i,b}$ in Eq. S3, and the first term on the right-hand side then becomes:

$$\hat{\mathbb{V}}_m\left(\sum_{i \in s} \frac{y_{i,b}}{\pi_i}\right) = \hat{\mathbb{V}}_m\left(\sum_{i \in s} \frac{\boldsymbol{x}_i\hat{\boldsymbol{\beta}} + \boldsymbol{x}_i\boldsymbol{\epsilon}_{\boldsymbol{\beta},b} + \epsilon_{i,b}}{\pi_i}\right)$$
$$= \hat{\mathbb{V}}_m\left(\sum_{i \in s} \frac{\boldsymbol{x}_i\boldsymbol{\epsilon}_{\boldsymbol{\beta},b}}{\pi_i} + \sum_{i \in s} \frac{\epsilon_{i,b}}{\pi_i}\right)$$
$$= \hat{\boldsymbol{\tau}}_x\hat{\mathbb{V}}_m(\hat{\boldsymbol{\beta}})\hat{\boldsymbol{\tau}}_x^T + \hat{\mathbb{V}}_m\left(\sum_{i \in s} \frac{\epsilon_{i,b}}{\pi_i}\right)$$
$$= \hat{\boldsymbol{\tau}}_x\hat{\mathbb{V}}_m(\hat{\boldsymbol{\beta}})\hat{\boldsymbol{\tau}}_x^T + \sum_{i \in s}\left(\frac{\hat{\mathbb{V}}_m(\epsilon_i)}{\pi_i^2}\right) + \sum_{i \in s}\sum_{i' \neq i} \frac{\widehat{\text{COV}}_m(\epsilon_i, \epsilon_{i'})}{\pi_i\pi_{i'}} \tag{S4}$$

where $\hat{\boldsymbol{\tau}}_x$ is the estimated total of the auxiliary variables throughout the population.

With a little algebra, the second term on the right-hand side of Eq. S3 can be re-expressed as:

$$\hat{\mathbb{E}}_m\left[\sum_{i \in s}\left(\frac{1 - \pi_i}{\pi_i^2}\right)y_{i,b}^2 + \sum_{i \in s}\sum_{i' \neq i}\left(\frac{\pi_{ii'} - \pi_i\pi_{i'}}{\pi_i\pi_{i'}}\right)\frac{y_{i,b}y_{i',b}}{\pi_{ii'}}\right] = \hat{\boldsymbol{\beta}}^T\hat{\mathbb{V}}_d(\hat{\boldsymbol{\tau}}_x)\hat{\boldsymbol{\beta}}$$
$$+ \text{Tr}(\hat{\mathbb{V}}_d(\hat{\boldsymbol{\tau}}_x)\hat{\mathbb{V}}_m(\hat{\boldsymbol{\beta}})) + \sum_{i \in s}\left(\frac{1 - \pi_i}{\pi_i^2}\right)\hat{\mathbb{V}}_m(\epsilon_i) + \sum_{i \in s}\sum_{i' \neq i}\left(\frac{\pi_{ii'} - \pi_i\pi_{i'}}{\pi_i\pi_{i'}}\right)\frac{\widehat{\text{COV}}_m(\epsilon_i, \epsilon_{i'})}{\pi_{ii'}} \tag{S5}$$

Combining Eqs. S4 and S5 yields:

$$\hat{\mathbb{V}}_{md}(\hat{\tau}_{\text{BS}} - \tau_y) = \hat{\boldsymbol{\beta}}^T \hat{\mathbb{V}}_d(\hat{\boldsymbol{\tau}}_x) \hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\tau}}_x \hat{\mathbb{V}}_m(\hat{\boldsymbol{\beta}}) \hat{\boldsymbol{\tau}}_x^T + \text{Tr}(\hat{\mathbb{V}}_d(\hat{\boldsymbol{\tau}}_x) \hat{\mathbb{V}}_m(\hat{\boldsymbol{\beta}}))$$

$$+ \sum_{i \in s} \left(\frac{2 - \pi_i}{\pi_i^2} \right) \hat{\mathbb{V}}_m(\epsilon_i) + \sum_{i \in s} \sum_{i' \neq i} \left(\frac{2\pi_{ii'} - \pi_i \pi_{i'}}{\pi_i \pi_{i'}} \right) \frac{\widehat{\mathbb{COV}}_m(\epsilon_i, \epsilon_{i'})}{\pi_{ii'}} \qquad \text{(S6)}$$

In the same conditions, i.e. with the same sample and the same super-population model, the limit of $\hat{\tau}_{\text{BS}}$ when B approaches infinity is actually $\hat{\tau}_{\text{HYB}}$:

$$\lim_{B \to \infty} \hat{\tau}_{\text{BS}} = \frac{\sum_{b=1}^B \hat{\tau}_{y,b}}{B}$$

$$= \frac{1}{B} \sum_{b=1}^B \sum_{i \in s} \frac{\boldsymbol{x}_i \hat{\boldsymbol{\beta}} + \boldsymbol{x}_i \boldsymbol{\epsilon}_{\boldsymbol{\beta},b} + \epsilon_{i,b}}{\pi_i}$$

$$= \sum_{i \in s} \frac{1}{\pi_i} \sum_{b=1}^B \frac{\boldsymbol{x}_i \hat{\boldsymbol{\beta}} + \boldsymbol{x}_i \boldsymbol{\epsilon}_{\boldsymbol{\beta},b} + \epsilon_{i,b}}{B}$$

$$= \sum_{i \in s} \frac{\boldsymbol{x}_i \hat{\boldsymbol{\beta}}}{\pi_i}$$

$$= \hat{\tau}_{\text{HYB}} \qquad \text{(S7)}$$

Consequently, the variance of the estimator is exactly the same as the one shown in Section SM1:

$$\mathbb{V}_{md}(\hat{\tau}_{\text{BS}} - \tau_y) = \boldsymbol{\beta}^T \mathbb{V}_d(\hat{\boldsymbol{\tau}}_x) \boldsymbol{\beta} + \boldsymbol{\tau}_x \mathbb{V}_m(\hat{\boldsymbol{\beta}}) \boldsymbol{\tau}_x^T + \text{Tr}(\mathbb{V}_d(\hat{\boldsymbol{\tau}}_x) \mathbb{V}_m(\hat{\boldsymbol{\beta}}))$$

$$+ \sum_{i=1}^N \mathbb{V}_m(\epsilon_i) + \sum_{i=1}^N \sum_{i' \neq i}^N \mathbb{COV}_m(\epsilon_i, \epsilon_{i'}) \qquad \text{(S8)}$$

where N is the population size.
Given that:

$$\mathbb{E} \left[\sum_{i \in s} \left(\frac{2 - \pi_i}{\pi_i^2} \right) \hat{\mathbb{V}}_m(\epsilon_i) \right] = \sum_{i=1}^N \left(\frac{2 - \pi_i}{\pi_i} \right) \mathbb{V}_m(\epsilon_i) \qquad \text{(S9)}$$

and:

$$\mathbb{E} \left[\sum_{i \in s} \sum_{i' \neq i} \left(\frac{2\pi_{ii'} - \pi_i \pi_{i'}}{\pi_i \pi_{i'}} \right) \frac{\widehat{\mathbb{COV}}_m(\epsilon_i, \epsilon_{i'})}{\pi_{ii'}} \right] = \sum_{i=1}^N \sum_{i' \neq i}^N \left(\frac{2\pi_{ii'} - \pi_i \pi_{i'}}{\pi_i \pi_{i'}} \right) \mathbb{COV}_m(\epsilon_i, \epsilon_{i'}) \qquad \text{(S10)}$$

it can be shown that the estimator in Eq. S6 overestimates the true variance (Eq. S8) by the quantity:

$$\mathbb{E}[\hat{\mathbb{V}}_{md}(\hat{\tau}_{\mathrm{BS}} - \tau_y) - \mathbb{V}_{md}(\hat{\tau}_{\mathrm{BS}} - \tau_y)] = 2\mathrm{Tr}(\mathbb{V}_d(\hat{\boldsymbol{\tau}}_x)\mathbb{V}_m(\hat{\boldsymbol{\beta}})) + \sum_{i=1}^{N}\left(\frac{2}{\pi_i} - 2\right)\mathbb{V}_m(\epsilon_i)$$

$$+ \sum_{i=1}^{N}\sum_{i'\neq i}^{N}\left(\frac{2\pi_{ii'} - 2\pi_i\pi_{i'}}{\pi_i\pi_{i'}}\right)\mathbb{COV}(\epsilon_i, \epsilon_{i'}) \qquad (\mathrm{S}11)$$

This estimator can be corrected as follows:

$$\hat{\mathbb{V}}_{md,\mathrm{CORR}}(\hat{\tau}_{\mathrm{BS}} - \tau_y) = \hat{\mathbb{V}}_m(\hat{\tau}_{y,b}) + 2\hat{\mathbb{V}}_d(\hat{\tau}_{\bar{y}}) - \hat{\mathbb{E}}_m[\hat{\mathbb{V}}_d(\hat{\tau}_{y,b})]$$

$$= \hat{\boldsymbol{\tau}}_x\hat{\mathbb{V}}_m(\hat{\boldsymbol{\beta}})\hat{\boldsymbol{\tau}}_x^T + \hat{\boldsymbol{\beta}}^T\hat{\mathbb{V}}_d(\hat{\boldsymbol{\tau}}_x)\hat{\boldsymbol{\beta}} - \mathrm{Tr}(\hat{\mathbb{V}}_d(\hat{\boldsymbol{\tau}}_x)\hat{\mathbb{V}}_m(\hat{\boldsymbol{\beta}}))$$

$$+ \sum_{i\in s}\left(\frac{\hat{\mathbb{V}}_m(\epsilon_i)}{\pi_i}\right) + \sum_{i\in s}\sum_{i'\neq i}\left(\frac{\widehat{\mathbb{COV}}_m(\epsilon_i, \epsilon_{i'})}{\pi_{ii'}}\right) \qquad (\mathrm{S}12)$$

where $\hat{\mathbb{V}}_d(\hat{\tau}_{\bar{y}})$ is obtained by substituting $\bar{y}_i = \sum_{b=1}^{B}\frac{y_{i,b}}{B}$ for y_i in the HT variance estimator defined in Eq. 3. When the model is linear, $\lim_{B\to\infty}\hat{\mathbb{V}}_d(\hat{\tau}_{\bar{y}}) = \hat{\boldsymbol{\beta}}^T\hat{\mathbb{V}}_d(\hat{\boldsymbol{\tau}}_x)\hat{\boldsymbol{\beta}}$, and the bias of this corrected estimator is then:

$$\mathbb{E}[\hat{\mathbb{V}}_{md,\mathrm{CORR}}(\hat{\tau}_{\mathrm{BS}} - \tau_y) - \mathbb{V}_{md}(\hat{\tau}_{\mathrm{BS}} - \tau_y)] = 0 \qquad (\mathrm{S}13)$$

The only sampling-related variance component in this variance estimator (Eq. S12) is $\hat{\mathbb{V}}_d(\hat{\tau}_{\bar{y}})$, which is equal to $\hat{\boldsymbol{\beta}}^T\hat{\mathbb{V}}_d(\hat{\boldsymbol{\tau}}_x)\hat{\boldsymbol{\beta}}$ when the model is linear. All the other terms, i.e., $\hat{\mathbb{V}}_m(\hat{\tau}_{y,b}) + \hat{\mathbb{V}}_d(\hat{\tau}_{\bar{y}}) - \hat{\mathbb{E}}_m[\hat{\mathbb{V}}_d(\hat{\tau}_{y,b})]$, compose the model-related variance. The sampling-related variance is only included in the model-related component as a correction factor.

SM3 Quantification of the bias of the variance estimator

As shown in Section SM2, a variance estimator based on the law of total variance and a parametric bootstrap has the following bias:

$$\mathrm{Bias}[\hat{\mathbb{V}}_{md}(\hat{\tau}_{\mathrm{BS}} - \tau_y)] = 2\mathrm{Tr}(\mathbb{V}_d(\hat{\boldsymbol{\tau}}_x)\mathbb{V}_m(\hat{\boldsymbol{\beta}})) + \sum_{i=1}^{N}\left(\frac{2}{\pi_i} - 2\right)\mathbb{V}_m(\epsilon_i)$$

$$+ \sum_{i=1}^{N}\sum_{i'\neq i}^{N}\left(\frac{2\pi_{ii'} - 2\pi_i\pi_{i'}}{\pi_i\pi_{i'}}\right)\mathbb{COV}_m(\epsilon_i, \epsilon_{i'}) \qquad (\mathrm{S}14)$$

where $\hat{\tau}_{\mathrm{BS}}$ is the estimated total, τ_y is the true (but unknown) total, $\mathrm{Tr}(\cdot)$ is the trace of a matrix argument, $\boldsymbol{\tau}_x$ is the total of the auxiliary variables across the population, $\hat{\boldsymbol{\beta}}$ is the vector of parameter estimates of the super-population model, N is number of units in the population, π_i is the inclusion probability of population unit i, ϵ_i is the residual error term

of the super-population model for unit i, and $\pi_{ii'}$ is the joint inclusion probability, i.e., the probability that both units i and i' are part of the sample. Indices m and d stand for model and probability design features, respectively.

Term $\mathrm{Tr}(\mathbb{V}_d(\hat{\boldsymbol{\tau}}_x)\mathbb{V}_m(\hat{\boldsymbol{\beta}}))$ is asymptotically equal to 0. In the context of simple random sampling without replacement (SRSWoR) of large populations, $\pi_{ii'} \approx \pi_i \pi_{i'}$ and term $\frac{2\pi_{ii'} - 2\pi_i \pi_{i'}}{\pi_i \pi_{i'}}$ tends to 0. Thus, this bias can be approximated as follows:

$$\mathrm{Bias}[\hat{\mathbb{V}}_{md}(\hat{\tau}_{\mathrm{BS}} - \tau_y)] \approx \sum_{i=1}^{N} \left(\frac{2}{\pi_i} - 2 \right) \mathbb{V}_m(\epsilon_i) \tag{S15}$$

Now, let us assume that the inclusion probability is equal across the population units. With SRSWoR, $\pi_i = n/N$ for $\forall i$ (Gregoire and Valentine, 2008, p. 37), and the bias is then:

$$\begin{aligned}
\mathrm{Bias}[\hat{\mathbb{V}}_{md}(\hat{\tau}_{\mathrm{BS}} - \tau_y)] &\approx \sum_{i=1}^{N} \left(\frac{2}{\pi_i} - 2 \right) \mathbb{V}_m(\epsilon_i) \\
&= \left(\frac{2N}{n} - 2 \right) \sum_{i=1}^{N} \mathbb{V}_m(\epsilon_i) \\
&= \left(\frac{N}{n} - 1 \right) 2N\bar{\mathbb{V}}_{m,\epsilon}
\end{aligned} \tag{S16}$$

where $\bar{\mathbb{V}}_{m,\epsilon}$ is the average variance of the residual error term across the population, i.e., $\bar{\mathbb{V}}_{m,\epsilon} = \sum_{i}^{N} \mathbb{V}_m(\epsilon_i)/N$.

Thus, there are three factors that affect the magnitude of the bias: the population size, the sample size and the average residual variance of the super-population model. 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.

Since the variance $\mathbb{V}_{md}(\hat{\tau}_{\mathrm{BS}} - \tau_y)$ also increases with the population size, the relative bias can provide a better idea of the impact of this bias. Let us assume that the super-population model is linear and that the covariance between the error terms is negligible. Building on the link between the estimate of the total and that of the mean $\hat{\mu} = \hat{\tau}/N$ (Gregoire and Valentine, 2008, p. 40), the relative bias can be approximated as:

$$\begin{aligned}
\frac{\mathrm{Bias}[\hat{\mathbb{V}}_{md}(\hat{\tau}_{\mathrm{BS}} - \tau_y)]}{\mathbb{V}_{md}(\hat{\tau}_{\mathrm{BS}} - \tau_y)} &\approx \frac{\left(\frac{N}{n} - 1 \right) 2N\bar{\mathbb{V}}_{m,\epsilon}}{\mathbb{V}_{md}(\hat{\tau}_{\mathrm{BS}} - \tau_y)} \\
&= \frac{\left(\frac{N}{n} - 1 \right) 2N\bar{\mathbb{V}}_{m,\epsilon}}{\boldsymbol{\beta}^T \mathbb{V}_d(\hat{\boldsymbol{\tau}}_x)\boldsymbol{\beta} + \boldsymbol{\tau}_x \mathbb{V}_m(\hat{\boldsymbol{\beta}})\boldsymbol{\tau}_x^T + \mathrm{Tr}(\mathbb{V}_d(\hat{\boldsymbol{\tau}}_x)\mathbb{V}_m(\hat{\boldsymbol{\beta}})) + \sum_{i=1}^{N} \mathbb{V}_m(\epsilon_i)} \\
&= \frac{\left(\frac{N}{n} - 1 \right) 2N\bar{\mathbb{V}}_{m,\epsilon}}{N^2 \left(\boldsymbol{\beta}^T \mathbb{V}_d(\hat{\boldsymbol{\mu}}_x)\boldsymbol{\beta} + \boldsymbol{\mu}_x \mathbb{V}_m(\hat{\boldsymbol{\beta}})\boldsymbol{\mu}_x^T + \mathrm{Tr}(\mathbb{V}_d(\hat{\boldsymbol{\mu}}_x)\mathbb{V}_m(\hat{\boldsymbol{\beta}})) + \bar{\mathbb{V}}_{m,\epsilon}/N \right)} \\
&= \frac{2(1 - n/N)\bar{\mathbb{V}}_{m,\epsilon}}{n \left(\boldsymbol{\beta}^T \mathbb{V}_d(\hat{\boldsymbol{\mu}}_x)\boldsymbol{\beta} + \boldsymbol{\mu}_x \mathbb{V}_m(\hat{\boldsymbol{\beta}})\boldsymbol{\mu}_x^T + \mathrm{Tr}(\mathbb{V}_d(\hat{\boldsymbol{\mu}}_x)\mathbb{V}_m(\hat{\boldsymbol{\beta}})) + \bar{\mathbb{V}}_{m,\epsilon}/N \right)}
\end{aligned} \tag{S17}$$

In the context of a large population with a low sampling intensity, terms n/N and $\bar{\mathbb{V}}_{m,\epsilon}/N$ are both close to 0. At first glance, it could be assumed that the relative bias asymptotically decreases with increasing sample sizes. However, term $\mathbb{V}_d(\hat{\boldsymbol{\mu}}_x)$ does the same. Under the assumption of equal inclusion probabilities across the population units, then $\mathbb{V}_d(\hat{\boldsymbol{\mu}}_x) \approx \mathbb{V}_d(\boldsymbol{x}_i)/n$ and the approximation of the relative bias further simplifies to:

$$\frac{\text{Bias}[\hat{\mathbb{V}}_{md}(\hat{\tau}_{\text{BS}})]}{\mathbb{V}_{md}(\hat{\tau}_{\text{BS}})} \approx \frac{2\bar{\mathbb{V}}_{m,\epsilon}}{\boldsymbol{\beta}^T\mathbb{V}_d(\boldsymbol{x}_i)\boldsymbol{\beta} + n\boldsymbol{\mu}_x\mathbb{V}_m(\hat{\boldsymbol{\beta}})\boldsymbol{\mu}_x^T + \text{Tr}(\mathbb{V}_d(\boldsymbol{x}_i)\mathbb{V}_m(\hat{\boldsymbol{\beta}}))} \tag{S18}$$

The approximation S18 makes it possible to distinguish two different patterns. On the one hand, if the model-related variance is greater than the sampling-related variance, then term $n\boldsymbol{\mu}_x\mathbb{V}_m(\hat{\boldsymbol{\beta}})\boldsymbol{\mu}_x^T$ contributes to a greater part of the denominator and the relative bias will decrease with an increasing sample size. On the other hand, if the sampling-related variance, mainly represented by term $\boldsymbol{\beta}^T\mathbb{V}_d(\boldsymbol{x}_i)\boldsymbol{\beta}$, is the most important component of the denominator, then the increasing sample size should have a negligible effect on the relative bias. In both cases, the relative bias proportionally decreases to any decrease in the average residual variance of the super-population model.

In practice, the sampling-related variance is much greater than the model-related variance in most cases (e.g. McRoberts and Westfall, 2014; Ståhl et al., 2014; Fortin et al., 2016b), which implies that the increasing sample size should have a negligible effect on the relative bias. If the super-population model was fitted to a large dataset and the observations of this dataset were collected under SRSWoR, then:

$$\frac{\bar{\mathbb{V}}_{m,\epsilon}}{\boldsymbol{\beta}^T\mathbb{V}_d(\boldsymbol{x}_i)\boldsymbol{\beta}} \approx \frac{1-R^2}{R^2} \tag{S19}$$

where R^2 is the coefficient of determination of the super-population model.

Given this large dataset, we can assume that $\boldsymbol{\mu}_x\mathbb{V}_m(\hat{\boldsymbol{\beta}})\boldsymbol{\mu}_x^T$ and $\text{Tr}(\mathbb{V}_d(\boldsymbol{x}_i)\mathbb{V}_m(\hat{\boldsymbol{\beta}}))$ are negligible and that the relative bias could be roughly approximated as:

$$\frac{\text{Bias}[\hat{\mathbb{V}}_{md}(\hat{\tau}_{\text{BS}})]}{\mathbb{V}_{md}(\hat{\tau}_{\text{BS}})} \approx \frac{2(1-R^2)}{R^2} \tag{S20}$$

Note that all these approximations of the relative bias also apply to the estimator of the mean since both the numerator and the denominator of the relative bias are divided by N^2.

SM4 Model of Fortin et al. (2009)

Fortin et al. (2009) designed a model to predict the presence and volume conditional on the presence for five log grades in standing trees of sugar maple (*Acer saccharum* Marsh.) and yellow birch (*Betula alleghaniensis* Britton). The minimum diameter at breast height (DBH, 1.3 m in height) for commercial use of these two species is set at 23.1 cm in the province of Québec, Canada. Trees below this threshold are not considered in the model.

In brief, the model has two parts: a first one based on logistic regression, which aims at predicting the presence of the log grades in a particular tree, and a second one based on

linear regression, which predicts the volumes of the log grades conditional on their presence. Each part consists of a system of five simultaneous equations, one for each log grade.

Let i, j, and k be the indices of the plots, the trees and the log grades, respectively. In Fortin et al. (2009), the model was fitted with the species in interaction with the other fixed effects, but for the sake of simplicity, we will omit the species index here. The first part of the model can be described as follows:

$$\Pr(v_{ijk} > 0) = \frac{e^{\lambda_{0,k}+\lambda_{1,k}d_{ij}+\lambda_{2,k}d_{ij}^2}}{1+e^{\lambda_{0,k}+\lambda_{1,k}d_{ij}+\lambda_{2,k}d_{ij}^2}} \tag{S21}$$

where v_{ijk} is the volume of log grade k in tree j of plot i and d_{ij} is the tree DBH (cm) and $\lambda_{0,k}$, $\lambda_{1,k}$, and $\lambda_{2,k}$ are parameters to be estimated.

The second part of the model predicts the volume conditional on the presence of the log grade in the tree and can be expressed as:

$$\ln(v_{ijk} \mid v_{ijk} > 0) = \gamma_{0,k} + \gamma_{1,k}d_{ij} + \epsilon_{ijk} \tag{S22}$$

where $\gamma_{0,k}$ and $\gamma_{1,k}$ are parameters to be estimated and ϵ_{ijk} is a residual error term. The vector of within-tree residual error terms is assumed to follow a multivariate normal distribution so that:

$$\boldsymbol{\epsilon_{ij}} = \begin{pmatrix} \epsilon_{ij1} \\ \epsilon_{ij2} \\ \epsilon_{ij3} \\ \epsilon_{ij4} \\ \epsilon_{ij5} \end{pmatrix} \sim N_5(\mathbf{0}, \mathbf{R}) \tag{S23}$$

According to the data structure reported in Fortin et al. (2009), the fit of the first part of the model had 1590 degrees of freedom. The fit of the second part could be estimated at 599 degrees of freedom. Given the dimensions of the two vectors of parameters and their variance-covariance matrices, the estimates are not listed here but they can be provided upon request to the first author.

SM5 Model of Schneider (2007)

The model of Schneider (2007) aims at predicting the volume of five log grades – pulpwood, sawlog, low-grade sawlog, veneer, and low-grade veneer – in standing trees of white birch (*Betula papyrifera* Marsh.). The model consists of a series of submodels, with the predictions of some being used as predictors in others, and requires an assessment of the tree quality. This classification, which can be found in MRN (1995), is essentially inspired by the work of Hanks (1976). Based on the presence of external defects, there are four possible tree quality classes, A, B, C, and D, ranging from the highest to the lowest quality. Trees below 23.1 cm in DBH do not qualify for this classification and are therefore, given none (N). The flowchart of the model is presented in Fig. S1.

Let i and j be the indices of the plots and the trees, respectively. Let us also define k as the index of the submodel. The first submodel predicts the height (m) to a small-end diameter of 20 cm ($\mathrm{h}20_{ij}$):
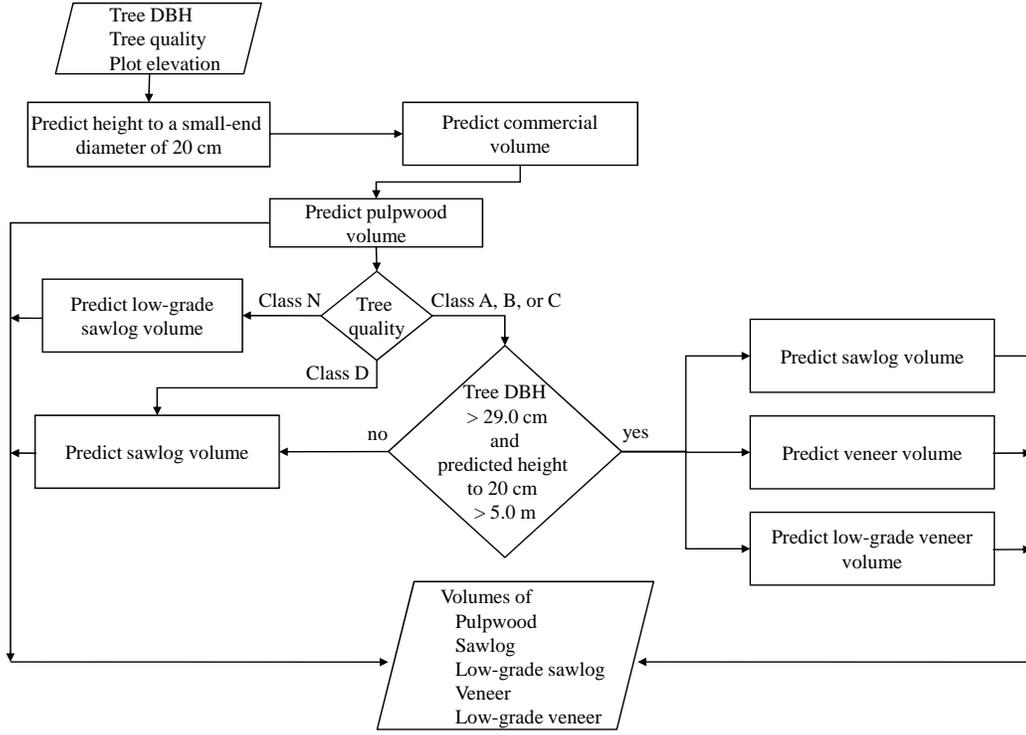
Figure S1: Flowchart of the model of Schneider (2007). DBH: diameter at breast height (1.3 m).

$$h20_{ij} = (\lambda_1 + \lambda_2 \text{elev}_i) \cdot d_{ij} \cdot (1 - e^{\lambda_3(d_{ij}-15)})^{\lambda_4} + \epsilon_{ij1} \tag{S24}$$

where elev_i is the elevation (m) of plot i, d_{ij} is the diameter (cm) measured at 1.3 m in height (DBH), λ_1, λ_2, λ_3, and λ_4 are parameters to be estimated, and ϵ_{ij1} is a residual error term.

The second submodel predicts the commercial volume as follows:

$$\text{vcom}_{ij} = \lambda_5 + \lambda_6(d_{ij} - 9) \cdot \widehat{h20}_{ij} + \lambda_7 \text{isAB}_{ij} + \epsilon_{ij2} \tag{S25}$$

where vcom_{ij} is the commercial volume (m^3) of tree j in plot i, $\widehat{h20}_{ij}$ is the prediction of the height to a small-end diameter of 20 cm as provided by the previous submodel (Eq. S24), isAB_{ij} is a dummy variable that takes the value of 1 if the tree quality belongs to class A or B according to the current quality classification or 0 otherwise, λ_5, λ_6, and λ_7 are parameters to be estimated, and ϵ_{ij2} is a residual error term.

The third submodel predicts the volume of pulpwood in the tree:

$$\text{vpulp}_{ij} = \frac{\widehat{\text{vcom}}_{ij}}{1 + e^{\lambda_8 - \lambda_9 \frac{\widehat{h20}_{ij}}{d_{ij}}}} + \epsilon_{ij3} \tag{S26}$$

8

where vpulp$_{ij}$ is the pulpwood volume (m^3) of tree j in plot i, $\widehat{\text{vcom}}_{ij}$ is the prediction of the commercial volume according to the previous submodel (Eq. S25), λ_8 and λ_9 are parameters to be estimated, and ϵ_{ij3} is a residual error term.

If the tree has no quality class, then the only possible log grades are pulpwood and low-grade sawlog (Fig. S1). The volume of low-grade sawlog is then predicted as the difference between the commercial volume and the pulpwood volume.

For trees with a quality class, if one of these three conditions is true:

1. The tree belongs to the lowest quality class, i.e., D.

2. Its DBH is smaller than 29 cm.

3. Its predicted height to a small-end diameter of 20 cm is smaller than or equal to 5 m.

then the only possible log grades are pulpwood and sawlog. The sawlog volume is then predicted as the difference between the commercial volume and the pulpwood volume.

For all the other trees, a fourth submodel predicts the volume of sawlog grade:

$$\text{vsaw}_{ij} = \frac{\widehat{\text{vcom}}_{ij} - \widehat{\text{vpulp}}_{ij}}{1 + e^{\lambda_{10} - \lambda_{11} d_{ij}}} + \epsilon_{ij4} \tag{S27}$$

where vsaw$_{ij}$ is the volume (m^3) of sawlog grade, $\widehat{\text{vpulp}}_{ij}$ is the prediction of pulpwood volume according the previous submodel (Eq. S26), and λ_{10} and λ_{11} are parameters to be estimated, and ϵ_{ij4} is a residual error term.

A fifth submodel predicts the volume of low-grade veneer:

$$\text{vlgven}_{ij} = \frac{\widehat{\text{vcom}}_{ij} - \widehat{\text{vpulp}}_{ij} - \widehat{\text{vsaw}}_{ij}}{1 + \lambda_{12} d_{ij}} + \epsilon_{ij5} \tag{S28}$$

where vlgven$_{ij}$ is the volume (m^3) of low-grade veneer, $\widehat{\text{vsaw}}_{ij}$ is the prediction of sawlog volume according to the previous submodel (Eq. S27), λ_{12} is a parameter to be estimated, and ϵ_{ij5} is a residual error term. The volume of veneer is then predicted as the residual volume once the predictions of pulpwood, sawlog and low-grade veneer volumes have been subtracted from the commercial volume.

The submodels were simultaneously fitted and the vector of within-tree residual error terms was assumed to follow a multivariate normal distribution:

$$\boldsymbol{\epsilon}_{ij} = \begin{pmatrix} \epsilon_{ij1} \\ \epsilon_{ij2} \\ \epsilon_{ij3} \\ \epsilon_{ij4} \\ \epsilon_{ij5} \end{pmatrix} \sim N_5(\mathbf{0}, \mathbf{R}_{ij}) \tag{S29}$$

In order to account for heteroscedasticity, $\mathbf{R}_{ij}$ can be re-expressed as:

$$\mathbf{R}_{ij} = \sigma^2 \boldsymbol{\Gamma}_{ij}^{1/2} \boldsymbol{\Psi} \boldsymbol{\Gamma}_{ij}^{1/2} \tag{S30}$$

where σ^2 is the residual variance, matrix $\boldsymbol{\Psi}$ is a correlation matrix and $\boldsymbol{\Gamma}_{ij}$ is diagonal matrix with its elements being defined by variance functions:

$$\mathbf{\Gamma}_{ij} = \begin{pmatrix} \widehat{\mathrm{h20}}_{ij}^{2\theta_1} & 0 & 0 & 0 & 0 \\ 0 & \widehat{\mathrm{vcom}}_{ij}^{2\theta_2} & 0 & 0 & 0 \\ 0 & 0 & \widehat{\mathrm{vpulp}}_{ij}^{2\theta_3} & 0 & 0 \\ 0 & 0 & 0 & \widehat{\mathrm{vsaw}}_{ij}^{2\theta_4} & 0 \\ 0 & 0 & 0 & 0 & \widehat{\mathrm{vlgven}}_{ij}^{2\theta_5} \end{pmatrix} \tag{S31}$$

where $\widehat{\mathrm{vlgven}}_{ij}$ is the prediction of low-grade veneer volume according to the submodel shown in Eq. S28, θ_1, θ_2, θ_3, θ_4, and θ_5 are the parameters of the variance functions that are estimated during the model fit as well as the elements of the correlation matrix $\mathbf{\Psi}$.

According to the data structure reported in Schneider (2007), the model fit had 607 degrees of freedom. Given the dimensions of the vector of parameters and its variance-covariance matrix, the estimates are not listed here but they can be provided upon request to the first author.

References

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., 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.

Goodman, L. A. (1960). On the exact variance of products. *Journal of the American Statistical Association*, 55:708–713.

Gregoire, T. G. and Valentine, H. T. (2008). *Sampling techniques for natural and environmental resources*. Chapman & Hall/CRC, Boca Raton, Florida, USA.

Hanks, L. (1976). Hardwood tree grades for factory lumber. Technical report, USDA Forest Service. Northeastern Forest Experiment Station. Research Paper NE-333.

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.

MRN (1995). Classification des tiges d'essences feuillues. Technical report, Ministère des Ressources naturelles du Québec. Service des inventaires forestiers.

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.

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., 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.