



HAL
open science

Sampling Intensification for Forest Inventories within a specific domain

Trinh H.K. Duong, Guillaume Chauvet, Olivier Bouriaud

► **To cite this version:**

Trinh H.K. Duong, Guillaume Chauvet, Olivier Bouriaud. Sampling Intensification for Forest Inventories within a specific domain. 2025. hal-04934215

HAL Id: hal-04934215

<https://hal.science/hal-04934215v1>

Preprint submitted on 7 Feb 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sampling Intensification for Forest Inventories within a specific domain

Trinh H.K. Duong^{1†}, Guillaume Chauvet^{2†}, Olivier Bouriaud^{1,3*}

^{1*}Laboratoire d'Inventaire Forestier (LIF), Institut National de l'Information Géographique et Forestière (IGN), 54000 Nancy, France.

²Ecole Nationale de la Statistique et de l'Analyse de l'Information (ENSAI), CNRS, IRMAR-UMR 6625, Université de Rennes, F-35000 Rennes, France.

³Ștefan cel Mare University of Suceava, Str. Universității 13, 720229 Suceava, Romania.

*Corresponding author(s). E-mail(s): obouriaud@gmail.com;

Contributing authors: kimtrinh0311@gmail.com; guillaume.chauvet@ensai.fr;

[†]These authors contributed equally to this work.

Abstract

National Forest Inventories (NFIs) are large-scale surveys that typically employ low sampling intensity, sufficient for national-level estimations. However, this low sampling intensity can make it difficult to produce reliable estimates for specific domains of interest under a design-based approach due to limited sample sizes. NFIs use models (model-assisted or model-based approaches) for small area estimation to make estimations in the domain of interest with minimal or no sample. However the reduced sample size can also be challenging for fitting models. Increasing the sampling intensity would represent resolve these issues. In this paper, we propose solutions to complement an existing NFI sample in order to improve estimation. We compare several sampling designs of intensification. This intensification poses the issue of integrating two dependent and non-overlapping samples with varying sampling intensities: the regular NFI sample and the intensified sample. We provide estimators of totals and ratios, and associated variance estimators for the domain of interest and the entire territory using a conditional approach. Our results show that intensification reduces the variance for an estimation at the level of both the domain of interest and the whole territory, that the choice of sampling designs considered has a limited impact on the estimation of the outcome.

Keywords: two-phase sampling, domain of interest, conditional approach, intensified sample, national forest inventory, ratio estimation

Acknowledgements

This work was supported by a grant of Advanced Research on the Biology of Tree and Forest Ecosystems (LabEx ARBRE) [MAESTRO 21PN03], the National Institute of Geographic and Forest Information (IGN), and the French institute of Mathematics for Planet Earth (iMPT) [CONIFER]. OB acknowledges support from the Ministry of Research, Innovation and Digitalization within Program 1 - Development of national research and development system, Subprogram 1.2 - Institutional Performance - RDI excellence funding projects [10PFE/2021]. This work was also supported by project "Interdisciplinary Cloud and Big Data Center at Ștefan cel Mare University of Suceava" [POC/398/1/1, 343/390019], co-founded by the European Union.

This work was acknowledged with support by the Laboratory of Forest Inventory (LIF) with the administration of the University of Gustave Eiffel (UGE). We would like to express our sincere gratitude to Jean-Daniel Bontemps and Jean-Pierre Renauds for their invaluable assistance in reviewing and providing constructive feedback on this article. His expertise and insightful comments greatly contributed to the overall quality of this work.

1 Introduction

National Forest Inventories (NFIs) are large-scale surveys that produce statistics on a great variety of forest resource attributes, including growing stock volume, forest area, biodiversity, and deadwood volume (Tomppo, 2004; Kangas and Maltamo, 2006). To collect the data in the space, sampling grids are used to distribute the samples regularly across the territory, ensuring a spatial balance. NFIs primarily use design-based approaches on this sampling grid, relying on sampling designs for their estimations, which ensures the unbiasedness of the estimations. Sometimes, NFIs improve their estimations using auxiliary information through multi-phase sampling. Two-phase sampling (Neyman, 1934) is commonly used for efficiency: the first phase gathers auxiliary information (e.g., forest/non-forest classification) through photo-interpretation or satellite data, and the second phase involves field surveys (Opsomer et al, 2007; Tomppo et al, 2010), accommodating the dynamic nature of forest locations. Moreover, the auxiliary data can be used either for stratification at the second phase (Duong et al, 2024) or for post-stratification (Scott et al, 2005).

Under the sampling grid, NFIs typically have low sampling intensity, for example, one unit per 4 km² in the Danish NFI (Tomppo et al, 2010, chapter 9), or one unit per km² in the Spanish NFI (Tomppo et al, 2010, chapter 34). This intensity is sufficient to produce accurate national-level estimates (Guldin, 2021), often extendable to regions. However, the limited sampling intensity poses challenges in using design-based estimation for restricted domains of interest (DOIs), which range from large areas like provinces to small forest districts (Mauro et al, 2016; Hill, 2018; McRoberts, 2006) with minimal to no samples (Goerndt et al, 2013; Fabrizi and Żądło, 2018). While NFIs optimize sampling for national to regional needs, local estimates are subject to a large variability when sample sizes drop below critical levels, leading to unreliable design-based estimators.

Obtaining accurate DOI estimates for forest management or protected areas is challenging because of the limited number of observations (Breidenbach and Astrup, 2012), requiring advanced methodologies like SAE. SAE is used by NFIs using models (model-assisted or model-based estimation). Models generated by integrating NFI data with remote sensing (RS) (e.g., Landsat, ALS, LiDAR) have been effective in estimating wood volume and canopy height in several districts in Norway and Switzerland (Breidenbach and Astrup, 2012; Astrup et al, 2019; Breidenbach et al, 2020). For model-assisted estimation, NFI data and auxiliary data are taken from within the DOIs (Särndal et al, 1992; McConville et al, 2020). An example is post-stratification, which has been a key strategy for forest inventory estimation in the U.S. for decades (Bechtold and Patterson, 2005). The model-assisted estimators are approximately unbiased (Magnussen et al, 2014; Magnussen, 2015; Costa Saenz de San Pedro et al, 2003) for large sample sizes (McRoberts, 2010; Magnussen et al, 2014; Breidenbach et al, 2016). Meanwhile, for model-based estimation, the auxiliary data are mainly taken from outside the DOIs to "borrow strength" (Mauro et al, 2016; Rao and Molina, 2015; Goerndt et al, 2013; Fabrizi and Żądło, 2018; Best et al, 2019; Dettmann et al, 2022) to estimate variables with sparse or even no observations inside the DOI itself. Model-based techniques, despite potential biases (Georgakis, 2019), achieve high accuracy when well-fitted (Magnussen et al, 2014). However, even though using remote-sensed data improves DOI spatial resolution in both cases, it remains inherently biased (Ståhl et al, 2024) and supposes the availability of a sufficiently large sample size within the DOI to improve the model fitting (Hill et al, 2018a), avoiding overfitting (Hastie et al, 2005), and minimizing extrapolation errors (White et al, 2017; Renaud et al, 2022).

To solve the problem of insufficient sample size in the DOI, increasing local sampling intensity remains an efficient way to produce estimations for many variables within a DOI (Fuller, 1999), regardless of the estimation method. Some attributes, such as biodiversity characteristics, are indeed often measurable on the ground only, or not well captured from RS, making it challenging to model and predict. For example, RS has a limited ability to penetrate vegetation canopies by using satellites (Amani et al, 2019; Sirin et al, 2018). The saturation of the model was discovered in certain forest types, like young stands or forest transitions (Sagar et al, 2022) despite a huge amount of available auxiliary data. This indicates the necessity of extensive field data to train the model effectively. In particular, increasing the sampling intensity would enable design-based approaches, which are the preferred approaches in NFIs due to their unbiasedness and ability to support the additivity of estimations. However, increasing local intensity creates challenges in defining design weights when combining two different samples resulting from two samplings: the regular NFI sample and the locally intensified sample. This is because intensification is an unplanned event that happens after NFI sampling; it can only be added at the second phase of the NFI’s two-phase sampling design. This makes the computation harder. Two techniques, single-count (Horvitz-Thompson) and multiple-count (Hansen-Hurwitz) estimators, have been applied to independent samples (Grafström et al, 2019; Prentius et al, 2021). However, the intensification within a domain examined in this paper may result in the situation of non-overlapping and non-independent samples.

NFIs also require ratio estimates, e.g., species proportion in mixed forests (Dirnberger and Sterba, 2014) or volume by forest type (McRoberts, 2008), which remain under-explored in complex designs. This article contributes by: (1) proposing a local intensification approach for two-phase NFI designs to enhance estimations; (2) offering a method to combine dependent and non-overlapping samples with defined design weights; and (3) developing total and ratio estimators, along with variance estimators, and evaluating how intensified samples improve overall territory-level accuracy. Different methods can be used to increase sampling intensity within an NFI grid-based design; however, to our best knowledge, this has received little attention. This article includes a comparison to determine which method performs better than the others.

2 Grid-based two-phase sampling approach in NFIs

2.1 Selection of the NFI sample

We outline the two-phase sampling approach commonly used in NFIs using grid-based sampling ((Stevens Jr, 1997; Duong et al, 2024), which ensures a spatially balanced distribution of the sample over the territory studied (Stevens Jr and Olsen, 2004; Kermorvant et al, 2019). The NFI first-phase sample (NFI-P1 sample) can be selected by systematic aligned sampling (Opsomer et al, 2007; Gschwantner et al, 2010) or unaligned sampling (Cochran, 1977; McRoberts et al, 2015), or by means of two-stage sampling (Duong et al, 2024).

The NFI-P1 sample of points, denoted as S^a , of size n^a is selected from the territory \mathcal{T} with area $A_{\mathcal{T}}$ using any form of grid-based sampling, as illustrated in Figure 1. Photo-interpretation is then used to classify each unit of S^a according to its vegetation type (forest/non-forest) thus forming post-strata.

The NFI second-phase sample (NFI-P2 sample), denoted as S^{1b} , is subsampled from S^a using two-phase systematic sampling as outlined by Opsomer et al (2007), which applies a constant sampling

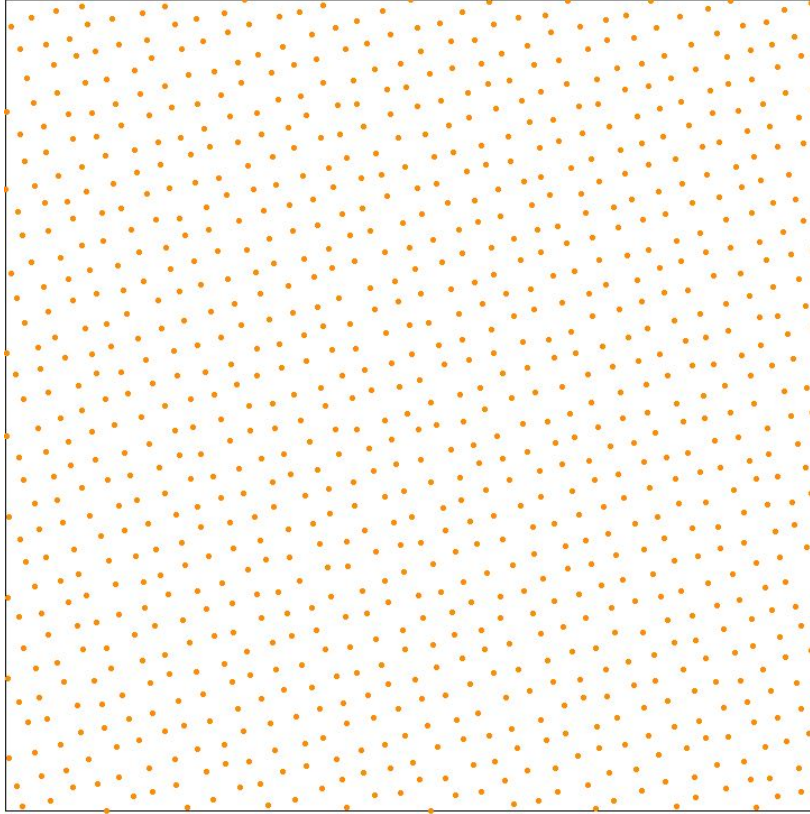


Fig. 1: The illustration of the NFI-P1 sample in the territory.

Lecture note: The large square represents the territory \mathcal{T} , and the orange dots represent the NFI-P1 sample S^a in the territory.

rate of f^{1b} and ensures regular sample distribution over the territory (see Figure 2a). Duong et al (2024) described this sampling design as *spatially systematic sampling*, and extended it to obtain a sample with varying sampling intensities for the French NFI. Since most NFIs use this design with a constant sampling rate, we refer to it as NFI design in this study. Although this article simplifies the second-phase selection using a constant sampling intensity across the territory, the sample size n^{1b} of S^{1b} is seen as random.

2.2 Selection of the intensified sample

In the following, \mathcal{D} represents the DOI (pink zone in Figure 2b) for which it is hypothesized that the surface area $A_{\mathcal{D}}$ is known. This situation is the most probable given that the delineation of this DOI will be made from geographical information, for instance management maps or cadastral data. $\bar{\mathcal{D}}$ represents the area outside the DOI. Intensification increases the second-phase sampling intensity within \mathcal{D} through taking an additional sample.

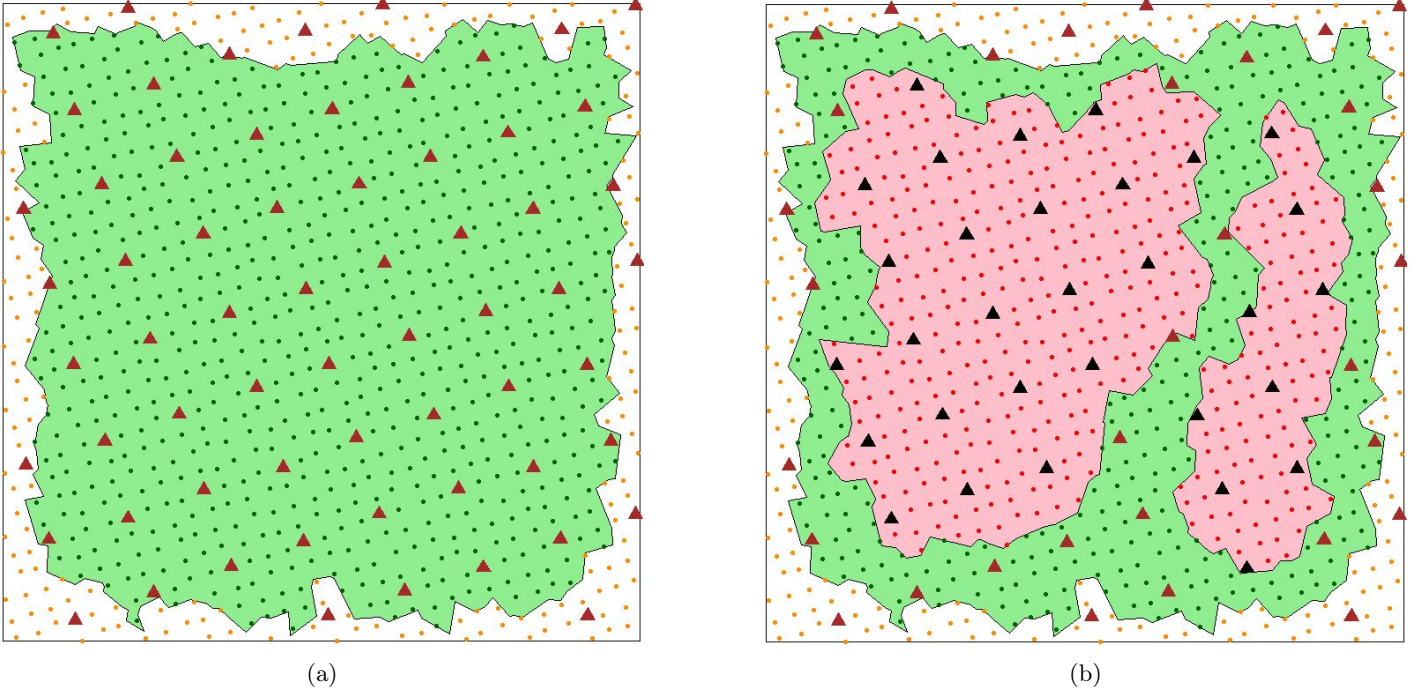


Fig. 2: The illustration of the first-phase classification, and the NFI-P2 sample selection

Lecture note: The NFI-P1 sample is classified into forest (the green zone in a) and non-forest. The triangles represent the NFI-P2 sample S^{1b} . The pink zone in (b) represents the DOI \mathcal{D} . The red points and the black triangles represent the NFI-P1 sample and the NFI-P2 sample which belongs to the DOI.

Figure 3 illustrates the notation used. In both S^a and S^{1b} , the points are categorized as being inside or outside \mathcal{D} according to their geographical location. Let S_D^a and S_D^{1b} denote the sub-samples within \mathcal{D} , with random sample sizes n_D^a and n_D^{1b} , respectively. The size will vary according to the sampling design and will vary from one drawing to the other. Meanwhile, S_D^{1b} denotes the sub-samples outside \mathcal{D} , with random size n_D^{1b} .

In \mathcal{D} , let $S_D^a \setminus S_D^{1b}$ denote the portion of S_D^a not selected in S_D^{1b} , with size $n_D^{a, \overline{1b}} = n_D^a - n_D^{1b}$. Intensification selects a new sample S_D^{2b} from this subset using a sampling rate $f_D^{2b|a, \overline{1b}}$. Figure 4 shows the intensified selection using various designs: simple random sampling without replacement (SRS_{swor}, a), systematic sampling (SYS, b), Poisson sampling (c), and NFI design (d, e).

In \mathcal{D} , two samples with different sampling intensities, S_D^{1b} and S_D^{2b} , are combined to form the union sample S_D^b with size $n_D^b = n_D^{1b} + n_D^{2b}$. The global second-phase sample (global P2 sample) S^b in \mathcal{T} includes two samples, S_D^b within \mathcal{D} and S_D^{1b} outside \mathcal{D} . The global P2 sample size is $n^b = n_D^b + n_D^{1b}$.

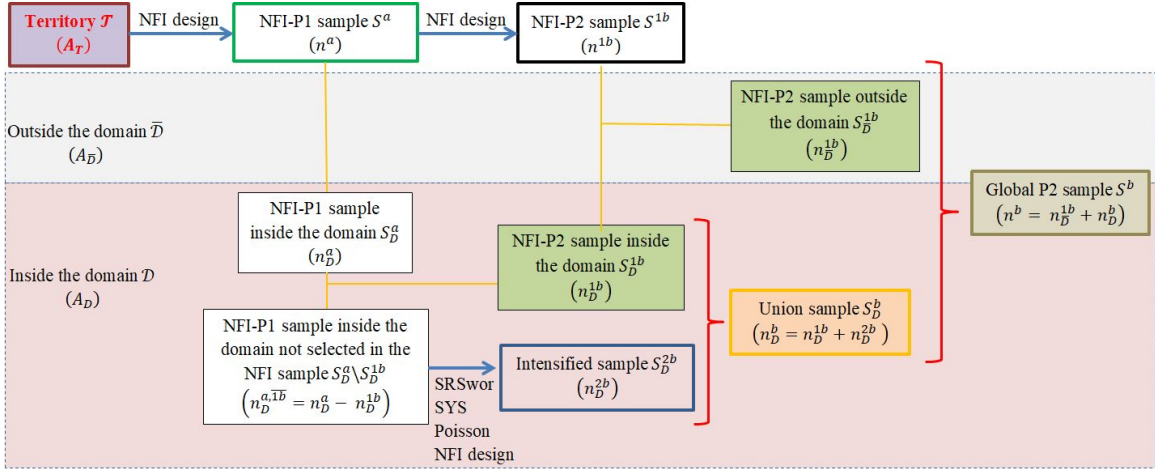


Fig. 3: The scheme to illustrate the notation of the populations with their area (in the parentheses) and the samples with their sample size (in the parentheses).

3 Estimation and variance estimation for population totals

The Weight Share Method (WSM) was presented by [Chauvet et al \(2023\)](#) and [Bouriaud et al \(2024\)](#) to formalize the relationship between trees and points, transforming tree-related variables into local variables defined at the point level, thereby simplifying the estimation of tree variables to point-value estimation.

For a local variable y defined for any point in \mathcal{T} , we are interested in estimating the total value τ_y for \mathcal{T} and τ_{Dy} for \mathcal{D} where, respectively

$$\tau_y = \int_{\mathcal{T}} y(x) dx,$$

$$\tau_{Dy} = \int_{\mathcal{D}} y(x) dx.$$

Under the design-based approach, to account for the sampling mechanism, estimators incorporate inclusion probabilities in finite populations ([McConville et al, 2020](#)), extended to inclusion density functions for continuous populations ([Cordy, 1993](#)). By selecting S^a via grid-based sampling, all points in \mathcal{T} have the same inclusion density in S^a

$$\pi^a(x) = \frac{n^a}{A_T} \text{ for } x \in \mathcal{T}.$$

In survey statistics, the design weight, derived as the inverse of a unit's inclusion probability (or density for continuous populations), reflects the probability of selecting that particular unit for the sample and represents the population units represented by this unit in a finite population, or the

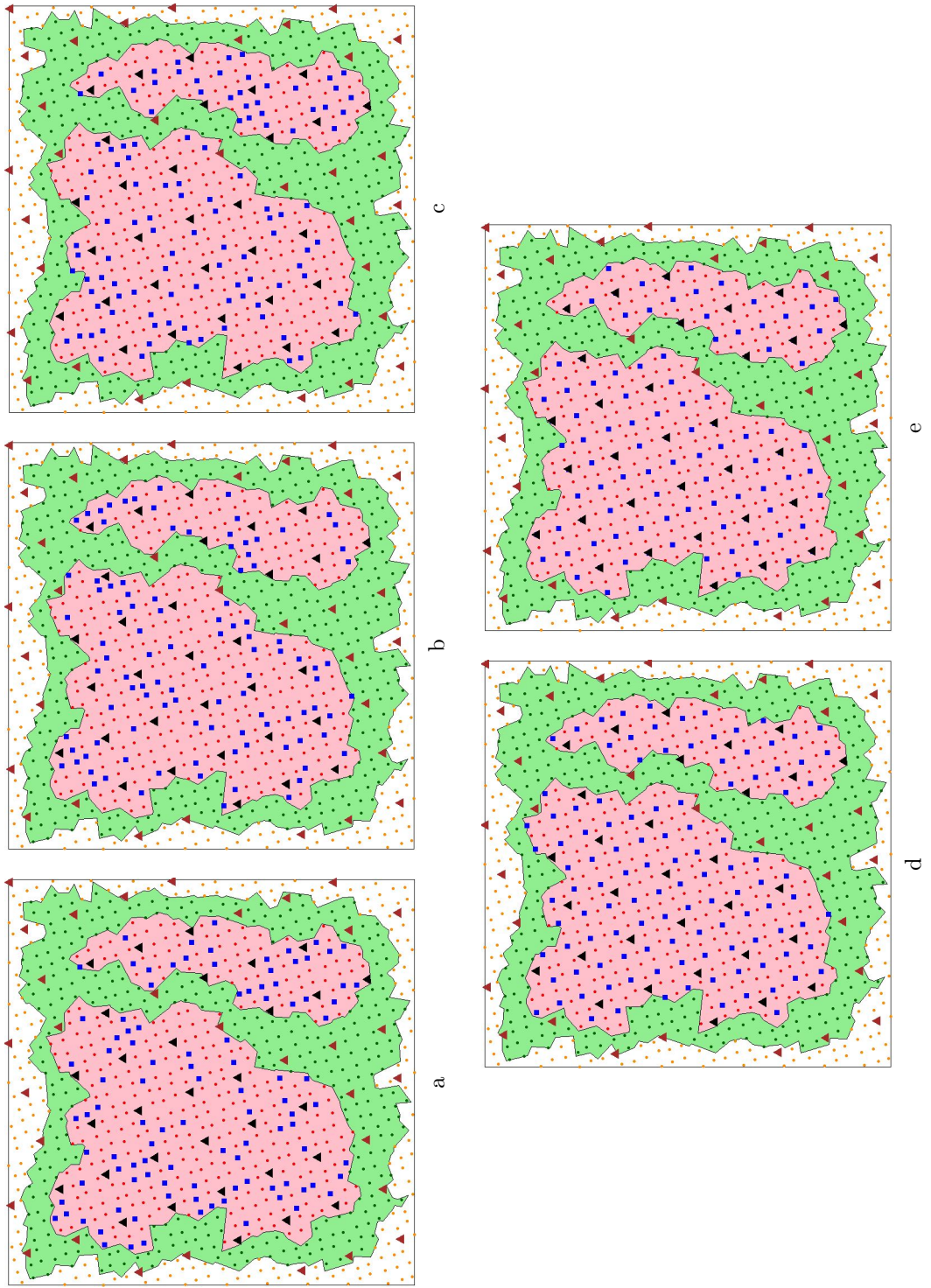


Fig. 4: The illustration of taking the sub-sample S_D^{2b} in the DOI.

Lecture note: The blue squares represent the additional points taken by (a) simple random sampling without replacement, (b) systematic sampling, (c) Poisson sampling, and (d,e) NFI design.

area in a continuous population (Prentius et al, 2021; Dettmann et al, 2022). For a point $x \in S^a$, the design weight is

$$d^a(x) = \frac{1}{\pi^a(x)} = \frac{A_T}{n^a} \text{ for } x \in S^a.$$

Using only the first-phase sample, the Horvitz-Thompson (HT) estimator based on the first-phase sample for τ_y is

$$\hat{\tau}_y^a = \sum_{x \in S^a} d^a(x) \times y(x) = \frac{A_T}{n^a} \sum_{x \in S^a} y(x).$$

However, the post-stratification divides the selected points into two post-strata: \mathcal{D} and $\bar{\mathcal{D}}$. While n^a is fixed, the post-strata sizes n_D^a and $n_{\bar{D}}^a$ are random. According to Rao (1985), inferences should be conditional on observed sample sizes if the population distribution is known and the sample sizes are random. Therefore, we provide total and variance estimators for \mathcal{D} and \mathcal{T} conditionally on the observed sample sizes.

3.1 Estimation without intensification

In absence of the intensification sample, the estimation for \mathcal{T} is based on S^{1b} only. The grid-based sampling results in equal inclusion densities for all the points of S^{1b} in \mathcal{T} . The random sample size n^{1b} has an expected value $E(n^{1b}) = n^a \times f^{1b}$. Conditionally on n^{1b} , the inclusion density function and design weights are

$$\begin{aligned} \pi^{1b|n^{1b}}(x) &= \frac{n^{1b}}{A_T} \text{ for } x \in \mathcal{T}, \\ d^{1b|n^{1b}}(x) &= \frac{A_T}{n^{1b}} \text{ for } x \in S^{1b}. \end{aligned}$$

The HT estimator based on the second-phase sample for τ_y is

$$\hat{\tau}_y^{1b} = \sum_{x \in S^{1b}} d^{1b|n^{1b}}(x) \times y(x) = \frac{A_T}{n^{1b}} \sum_{x \in S^{1b}} y(x). \quad (1)$$

The total estimator $\hat{\tau}_y^{1b}$ in (1) is conditionally design-unbiased since

$$E(\hat{\tau}_y^{1b} | n^{1b}) = \tau_y. \quad (2)$$

Appendix A proves equation (2).

Standard variance estimators require positive joint inclusion density function to be unbiased, a condition which systematic approaches like grid-based sampling do not satisfy since the grid-based

sampling precludes the spatial superposition of units in the space. However, as noted by [Mandallaz \(2007, p.53-54\)](#), forest inventories often assimilate points on systematic grids as resulting from a uniform sampling, leading to the approximating variance under SYS by the variance under SRS ([Särndal et al, 1992](#); [Wolter, 2007](#); [Westfall et al, 2018](#); [Duong et al, 2024](#)). This implies that assumptions of randomness for variance estimation are required to obtain the variance estimators.

To obtain the variance estimator of $\hat{\tau}_y^{1b}$, two assumptions are required. Firstly, knowing that S^a is obtained by the NFI strategy, it is considered as being selected by uniform sampling of size n^a from \mathcal{T} . Secondly, conditionally on n^{1b} , knowing that S^{1b} is obtained using the NFI strategy, it is treated as if it had been selected by SRS_{swr} from S^a with a size n^{1b} . Under these assumptions, S^{1b} can be seen as selected through a uniform sampling of size n^{1b} from \mathcal{T} , conditionally on n^{1b} . Therefore, the conditional unbiased variance estimator is given by

$$\hat{V}(\hat{\tau}_y^{1b}) = \frac{(A_T)^2}{n^{1b}} s_{y,1b}^2,$$

$$\text{with } s_{y,1b}^2 = \frac{1}{n^{1b} - 1} \sum_{x \in S^{1b}} \left\{ y(x) - \frac{1}{n^{1b}} \sum_{x' \in S^{1b}} y(x') \right\}^2.$$

Estimating within \mathcal{D} can be challenging with a limited sample size in S_D^{1b} , but can be improved by the intensification presented in Section 3.2. Table 1 presents the formulas for the total estimators and their corresponding variance estimators used in the DOI and the entire territory with and without intensification. The formulas for the DOI without intensification, which do not require detail, can be found in this table. The process for acquiring these formulas parallels those outlined previously for the entire territory.

3.2 Estimation with intensification

In \mathcal{D} , intensification increases the DOI sample size with the union sample S_D^b . The units in S_D^b have varying sampling intensities based on their origin. However, all points in \mathcal{D} have the same inclusion density in S_D^b (see Appendix B), leading the conditional inclusion density function and design weights by conditioning on n_D^b

$$\pi_D^{b|n_D^b}(x) = \frac{n_D^b}{A_D} \text{ for } x \in \mathcal{D}.$$

$$d_D^{b|n_D^b}(x) = \frac{A_D}{n_D^b} \text{ for } x \in S_D^b.$$

The conditional unbiased HT estimator is

$$\hat{\tau}_{Dy}^b = \frac{A_D}{n_D^b} \sum_{x \in S_D^b} y(x). \quad (3)$$

To obtain a variance estimator of $\hat{\tau}_{Dy}^b$, the two assumptions listed above are required but must be applied specifically to the samples within the DOI, as the target population has been changed. Note that the second assumption assumes S_D^b selected by SRS_{wo} of size n_D^b from S_D^a conditionally, even though it consists of two samples. This assumption neglects the differences in point features (from different samples, taken by different sampling designs, with different sampling intensities) inside the union sample, resulting in a straightforward method to calculate the variance estimator. Based on these assumptions, S_D^b is seen as selected by uniform sampling of size n_D^b from \mathcal{D} , conditionally on n_D^b , resulting in the conditional unbiased variance estimator

$$\hat{V}(\hat{\tau}_{Dy}^b) = \frac{(A_D)^2}{n_D^b} s_{yD,b}^2, \quad (4)$$

$$\text{with } s_{yD,b}^2 = \frac{1}{n_D^b - 1} \sum_{x \in S_D^b} \left\{ y(x) - \frac{1}{n_D^b} \sum_{x' \in S_D^b} y(x') \right\}^2. \quad (5)$$

With intensification, we provide an alternative solution for estimating in \mathcal{T} using the sample S^b , which consists of S_D^b within \mathcal{D} and $S_{\bar{D}}^{1b}$ within $\bar{\mathcal{D}}$. The conditional inclusion density function and design weights are determined for each post-stratum (\mathcal{D} and $\bar{\mathcal{D}}$), conditionally on n_D^b and $n_{\bar{D}}^{1b}$,

$$\pi^{b|n_D^b, n_{\bar{D}}^{1b}}(x) = \begin{cases} \frac{n_D^b}{A_D} & \text{if } x \in \mathcal{D}, \\ \frac{A_D}{A_T - A_D} & \text{if } x \in \bar{\mathcal{D}}, \end{cases} \quad \text{for } x \in \mathcal{T},$$

$$d^{b|n_D^b, n_{\bar{D}}^{1b}}(x) = \begin{cases} \frac{A_D}{n_D^b} & \text{if } x \in S_D^b, \\ \frac{A_T - A_D}{n_{\bar{D}}^{1b}} & \text{if } x \in S_{\bar{D}}^{1b}. \end{cases}$$

The conditional unbiased estimator for τ_y is then obtained according to the post-strata \mathcal{D} and $\bar{\mathcal{D}}$

$$\hat{\tau}_y^b = \frac{A_D}{n_D^b} \sum_{x \in S_D^b} y(x) + \frac{A_T - A_D}{n_{\bar{D}}^{1b}} \sum_{x \in S_{\bar{D}}^{1b}} y(x). \quad (6)$$

Conditionally on n_D^b and $n_{\bar{D}}^{1b}$ and under our assumptions, S^b may be seen as selected by stratified uniform sampling from \mathcal{T} , which would be stratified according to \mathcal{D} and $\bar{\mathcal{D}}$. An unbiased variance estimator is also obtained based on the post-strata \mathcal{D} and $\bar{\mathcal{D}}$

$$\hat{V}(\hat{\tau}_y^b) = \frac{(A_D)^2}{n_D^b} s_{yD,b}^2 + \frac{(A_T - A_D)^2}{n_{\bar{D}}^{1b}} s_{y\bar{D},1b}^2, \quad (7)$$

with $s_{yD,b}^2$ is presented in formula (5),

$$\text{and } s_{y\overline{D},1b}^2 = \frac{1}{n_D^{1b} - 1} \sum_{x \in S_D^{1b}} \left\{ y(x) - \frac{1}{n_D^{1b}} \sum_{x' \in S_D^{1b}} y(x') \right\}^2.$$

Intensification	Total estimator	Variance estimator
Domain of interest		
No Sample: S_D^{1b}	$\hat{\tau}_{Dy}^{1b} = \frac{A_D}{n_D^{1b}} \sum_{x \in S_D^{1b}} y(x)$	$\hat{V}(\hat{\tau}_{Dy}^{1b}) = \frac{(A_D)^2}{n_D^{1b}} s_{yD,1b}^2$ with $s_{yD,1b}^2 = \frac{1}{n_D^{1b} - 1} \sum_{x \in S_D^{1b}} \left\{ y(x) - \frac{1}{n_D^{1b}} \sum_{x' \in S_D^{1b}} y(x') \right\}^2$
Yes Sample: S_D^b	$\hat{\tau}_{Dy}^b = \frac{A_D}{n_D^b} \sum_{x \in S_D^b} y(x)$	$\hat{V}(\hat{\tau}_{Dy}^b) = \frac{(A_D)^2}{n_D^b} s_{yD,b}^2$ with $s_{yD,b}^2 = \frac{1}{n_D^b - 1} \sum_{x \in S_D^b} \left\{ y(x) - \frac{1}{n_D^b} \sum_{x' \in S_D^b} y(x') \right\}^2$
Territory		
No Sample: S^{1b}	$\hat{\tau}_y^{1b} = \frac{A_T}{n^{1b}} \sum_{x \in S^{1b}} y(x)$	$\hat{V}(\hat{\tau}_y^{1b}) = \frac{(A_T)^2}{n^{1b}} s_{y,1b}^2$ with $s_{y,1b}^2 = \frac{1}{n^{1b} - 1} \sum_{x \in S^{1b}} \left\{ y(x) - \frac{1}{n^{1b}} \sum_{x' \in S^{1b}} y(x') \right\}^2$
Yes Sample: S^b	$\hat{\tau}_y^b = \frac{A_D}{n_D^b} \sum_{x \in S_D^b} y(x) + \frac{A_T - A_D}{n_D^b} \sum_{x \in S_D^{1b}} y(x)$	$\hat{V}(\hat{\tau}_y^b) = \frac{(A_D)^2}{n_D^b} s_{yD,b}^2 + \frac{(A_T - A_D)^2}{n_D^{1b}} s_{y\overline{D},1b}^2$ with $s_{yD,b}^2 = \frac{1}{n_D^b - 1} \sum_{x \in S_D^b} \left\{ y(x) - \frac{1}{n_D^b} \sum_{x' \in S_D^b} y(x') \right\}^2$ and $s_{y\overline{D},1b}^2 = \frac{1}{n_D^{1b} - 1} \sum_{x \in S_D^{1b}} \left\{ y(x) - \frac{1}{n_D^{1b}} \sum_{x' \in S_D^{1b}} y(x') \right\}^2$

Table 1: The total estimators and their variance estimators used to estimate in the domain of interest and the entire territory with and without the intensification.

4 Estimation and variance estimation for population ratios

This section focuses on estimating the ratio of two unknown population totals for both \mathcal{D} and \mathcal{T} , respectively

$$R_D = \frac{\tau_{Dz}}{\tau_{Dy}} \text{ with } \tau_{Dz} = \int_{\mathcal{D}} z(x) dx,$$

$$R = \frac{\tau_z}{\tau_y} \text{ with } \tau_z = \int_{\mathcal{T}} z(x) dx.$$

Table 2 summarizes the formulas for the ratio estimators and their corresponding variance estimators used in the DOI and the entire territory with and without intensification. In this section, we present the ratio estimators for the case of combining two samples, i.e. intensification case, for the DOI and the entire territory. The formulas without intensification can be obtained similarly, and they can be found in Table 2.

With intensification, the substitution estimator of R_D is

$$\hat{R}_D^b = \frac{\hat{\tau}_{Dz}^b}{\hat{\tau}_{Dy}^b},$$

While the ratio estimators \hat{R}_D^b is biased for R_D though composed of two unbiased components (Särndal et al 1992, Chapter 5, p.163), the bias is negligible in large samples. Therefore, \hat{R}_D^b is *approximately* unbiased for R_D under large sample size. Since \hat{R}_D^b is a nonlinear estimator of R_D , its exact variance and variance estimator are difficult to obtain. Särndal et al (1992) suggested to use the *Taylor linearization technique* to estimate its variance, and we apply this method to derive the variance estimator. Let define $l_D^b(x)$ as

$$l_D^b(x) = \frac{1}{\hat{\tau}_{Dy}^b} \left\{ z(x) - \hat{R}_D^b \times y(x) \right\}.$$

Conditionally on n_D^b , the linearized variance estimator for \hat{R}_D^b is

$$\hat{V}(\hat{R}_D^b) = \frac{(A_D)^2}{n_D^b} s_{lD,b}^2,$$

where

$$s_{lD,b}^2 = \frac{1}{n_D^b - 1} \sum_{x \in S_D^b} \left\{ l_D^b(x) - \frac{1}{n_D^b} \sum_{x' \in S_D^b} l_D^b(x') \right\}^2.$$

For \mathcal{T} , the substitution estimator of R is

$$\hat{R}^b = \frac{\hat{\tau}_z^b}{\hat{\tau}_y^b}$$

where $\hat{\tau}_z^b$ is obtained from (6) by replacing $y(x)$ with $z(x)$. The linearized variable is

$$u^b(x) = \frac{1}{\hat{\tau}_y^b} \left\{ z(x) - \hat{R}^b \times y(x) \right\}.$$

The linearized variance estimator for \hat{R}^b can be obtained according to the post-strata \mathcal{D} and $\overline{\mathcal{D}}$, similar to the total estimation. The linearized variance estimator is

$$\hat{V}(\hat{R}^b) = \frac{(A_D)^2}{n_D^b} s_{uD,b}^2 + \frac{(A_T - A_D)^2}{n_{\overline{\mathcal{D}}}^{1b}} s_{u\overline{\mathcal{D}},1b}^2$$

where

$$s_{uD,b}^2 = \frac{1}{n_D^b - 1} \sum_{x \in S_D^b} \left\{ u^b(x) - \frac{1}{n_D^b} \sum_{x' \in S_D^b} u^b(x') \right\}^2,$$

$$s_{uD,1b}^2 = \frac{1}{n_D^{1b} - 1} \sum_{x \in S_D^{1b}} \left\{ u^b(x) - \frac{1}{n_D^{1b}} \sum_{x' \in S_D^{1b}} u^b(x') \right\}^2.$$

Intensification	Ratio estimator	Variance estimator
Domain of interest		
No	$\hat{R}_D^{1b} = \frac{\hat{\tau}_D^{1b} z}{\hat{\tau}_D^{1b} y}$	$\hat{V}(\hat{R}_D^{1b}) = \frac{(A_D)^2}{n_D^{1b}} s_{lD,1b}^2$
Sample: S_D^{1b}		with $s_{lD,1b}^2 = \frac{1}{n_D^{1b} - 1} \sum_{x \in S_D^{1b}} \left\{ l_D^{1b}(x) - \frac{1}{n_D^{1b}} \sum_{x' \in S_D^{1b}} l_D^{1b}(x') \right\}^2$ and $l_D^{1b}(x) = \frac{1}{\hat{\tau}_D^{1b}} \left\{ z(x) - \hat{R}_D^{1b} \times y(x) \right\}$
Yes	$\hat{R}_D^b = \frac{\hat{\tau}_D^b z}{\hat{\tau}_D^b y}$	$\hat{V}(\hat{R}_D^b) = \frac{(A_D)^2}{n_D^b} s_{lD,b}^2$
Sample: S_D^b		with $s_{lD,b}^2 = \frac{1}{n_D^b - 1} \sum_{x \in S_D^b} \left\{ l_D^b(x) - \frac{1}{n_D^b} \sum_{x' \in S_D^b} l_D^b(x') \right\}^2$ and $l_D^b(x) = \frac{1}{\hat{\tau}_D^b} \left\{ z(x) - \hat{R}_D^b \times y(x) \right\}$
Territory		
No	$\hat{R}^{1b} = \frac{\hat{\tau}^{1b} z}{\hat{\tau}^{1b} y}$	$\hat{V}(\hat{R}^{1b}) = \frac{(A_T)^2}{n^{1b}} s_{l,1b}^2$
Sample: S^{1b}		with $s_{l,1b}^2 = \frac{1}{n^{1b} - 1} \sum_{x \in S^{1b}} \left\{ u^{1b}(x) - \frac{1}{n^{1b}} \sum_{x' \in S^{1b}} u^{1b}(x') \right\}^2$ and $u^{1b}(x) = \frac{1}{\hat{\tau}^{1b}} \left\{ z(x) - \hat{R}^{1b} \times y(x) \right\}$
Yes	$\hat{R}^b = \frac{\hat{\tau}^b z}{\hat{\tau}^b y}$	$\hat{V}(\hat{R}^b) = \frac{(A_D)^2}{n_D^b} s_{uD,b}^2 + \frac{(A_T - A_D)^2}{n^{1b}} s_{uD,1b}^2$
Sample: S^b		with $s_{uD,b}^2 = \frac{1}{n_D^b - 1} \sum_{x \in S_D^b} \left\{ u^b(x) - \frac{1}{n_D^b} \sum_{x' \in S_D^b} u^b(x') \right\}^2$ with $s_{uD,1b}^2 = \frac{1}{n^{1b} - 1} \sum_{x \in S_D^{1b}} \left\{ u^b(x) - \frac{1}{n^{1b}} \sum_{x' \in S_D^{1b}} u^b(x') \right\}^2$ and $u^b(x) = \frac{1}{\hat{\tau}^b} \left\{ z(x) - \hat{R}^b \times y(x) \right\}$

Table 2: The ratio estimators and their variance estimators used to estimate in the domain of interest and the entire territory with and without the intensification.

5 Simulation study

5.1 Simulation set-up

A simulation dataset from [Duong et al \(2023\)](#) is used to illustrate numerically the formulas in Sections 3 and 4. The territory \mathcal{T} is a 100 km² square. The forest zone is located within \mathcal{T} (see Figure 2a). The population contains $N = 631\,714$ trees with features including species, height, breast height diameter, and volume. The DOI of area $A_D \approx 45.46$ km² is located inside the forest zone (see Figure 2b). We aim to estimate the total wood volume and the ratio of *Abies alba* volume within \mathcal{D} and \mathcal{T} .

5.2 Sampling design and estimators applied in the simulation study

The NFI sample was based on a two-phase design (see Section 2.1) with two-stage sampling in the first phase, as described by [Duong et al \(2024\)](#). The sample S^a of $n^a = 1000$ points was selected from \mathcal{T} , and S^{1b} was drawn with $f^{1b} = 1/16$ from S^a .

In \mathcal{D} , we investigated four sampling designs from the subset $S_D^a \setminus S_D^{1b}$ to obtain S_D^{2b} : SRSwor, SYS (by listing the points in ascending coordinate order), Poisson sampling ([Särndal et al 1992](#), Chapter 3), and NFI design. Within each design, S_D^{2b} was selected with $f_D^{2b|a,1b} = 1/4$. Meanwhile, SRSwor selected S_D^{2b} of size n_D^{2b} , ensuring constant union sample size n_D^b during iterations and consistent intensified fraction across all four sampling procedures.

The sampling design from which we produce the variance estimators under two assumptions - the uniform sampling at the first phase and SRSwor at the second phase - is referred to as the *benchmark strategy*. If the benchmark strategy is exactly implemented, the variance estimators should be unbiased. Therefore, we simulated the benchmark strategy to verify variance estimation assumptions in Sections 3 and 4. Uniform sampling from \mathcal{T} is used to sample S^a , with $n^a = 1000$. S^{1b} is then selected through SRSwor from S^a , with $n^{1b} = 62$ (see Table 3). Finally, S_D^{2b} is selected by SRSwor from $S_D^a \setminus S_D^{1b}$, ensuring the union sample S_D^b has the same size as other strategies (see Table 3).

Phase 1			Phase 2				Intensification			
S^a	n^a	n_D^a	S^{1b}	n^{1b}	n_D^{1b}	n_D^{1b}	S_D^{2b}	n_D^{2b}	n_D^b	n^b
NFI	1000	454.60 (3.68)	NFI	62.50 (0.81)	34.06 (2.14)	28.44 (1.84)	SRSwor	106.56 (1.84)	135	169.06 (2.14)
							SY \bar{S}	106.54 (1.05)	134.98 (1.79)	169.04 (1.38)
							Poisson	106.47 (9.01)	134.91 (9.12)	168.97 (9.05)
							NFI	106.54 (12.51)	134.98 (12.59)	169.04 (12.54)
Uniform	1000	454.43 (15.75)	SRSwor	62	33.83 (3.91)	28.18 (3.91)	SRSwor	106.83 (3.91)	135	168.83 (3.91)

Table 3: The Monte Carlo mean and standard deviation (in parentheses) of the sample size of the NFI-P1 sample, the NFI-P2 sample, the intensified sample, and the union samples in the DOI and the whole territory over 50 000 iterations.

The WSM (Chauvet et al, 2023; Bouriaud et al, 2024; Duong et al, 2024) is applied to convert the volume of trees into the selected points through a virtual circle centering on those points to connect the trees and the defined locations. We compared the total, ratio, and variance estimations for \mathcal{D} and \mathcal{T} with and without intensification. Formulas applied in this simulation study are described in Sections 3 and 4, and summarized in Tables 1 and 2.

5.3 Results from the simulation study

The sampling and estimation steps were repeated $M = 50\,000$ times, with conditional Monte Carlo (MC) computation detailed in Appendix C. Let $RB_{MC}(\hat{\theta})$ and $V_{MC}(\hat{\theta})$ denote the MC relative bias and variance of an estimator $\hat{\theta}$ for a parameter θ . Table 3 presents the MC results for the sample sizes, with standard deviation in parentheses. Table 3 indicates that the sampling designs have approximately the same expected sample sizes, though some variability was observed in the effective sample sizes, particularly in Poisson sampling and the NFI design.

Table 4 presents the simulation results for total estimation. The first lines for each main strategy (without intensification) are based only on the NFI-P2 sample. As expected, the Horvitz-Thompson estimator remains approximately unbiased across all scenarios. For \mathcal{D} , intensification reduces variance by about five times, with the NFI design for intensification showing the smallest variance (2.71×10^8), demonstrating high efficiency. Variance estimation is nearly unbiased for the benchmark strategies, as expected, but shows moderate positive bias (11 – 12%) for others and strong positive bias (37%) for the NFI design for intensification. These findings highlight that the intensification is highly effective for DOI estimation. For \mathcal{T} , similar trends are observed; however, the variance only decreases by 25% with intensification. This is linked to the fact that the DOI represents a large share of the territory’s area in the setup (around 45%). Variance estimation remains approximately unbiased for the benchmark, as expected, but exhibits significant positive bias (36% – 40%) for other strategies. Overall, the proposed variance estimators generally perform well, but show substantial positive bias with the NFI design for intensification.

Strategy			Domain of interest			Territory		
S^a	S^{1b}	S_D^{2b}	$RB_{MC}(\hat{\tau}_{Dy})$ (%)	$V_{MC}(\hat{\tau}_{Dy})$ ($\times 10^8$)	$RB_{MC}\{\hat{V}(\hat{\tau}_{Dy})\}$ (%)	$RB_{MC}(\hat{\tau}_y)$ (%)	$V_{MC}(\hat{\tau}_y)$ ($\times 10^8$)	$RB_{MC}\{\hat{V}(\hat{\tau}_y)\}$ (%)
NFI	NFI		0.23	16.04	9.36	-0.33	40.53	36.17
		SRSwor	0.39	3.28	12.23	-0.38	29.80	36.64
		SYS	0.52	3.30	12.00	-0.33	29.11	39.91
		Poisson	0.45	3.32	11.43	-0.36	29.91	36.21
		NFI	0.49	2.71	37.79	-0.34	29.00	40.57
Uniform	SRSwor		0.46	17.93	0.37	-0.31	55.12	0.43
		SRSwor	0.56	3.67	0.47	-0.28	41.04	-0.15

Table 4: Monte Carlo relative bias and Monte Carlo variance for the total estimation and variances estimation in the DOI and the entire territory under 50 000 iteration.

Table 5 presents the ratio estimation results. As expected, the Horvitz-Thompson estimators are nearly unbiased across all strategies. These results show that the intensification effectively reduces variance, the NFI strategy being the most efficient one, consistent with total estimation results.

However, variance estimators for benchmark strategies are negatively biased, resulting in underestimating the variance, as noted in the Remark of [Särndal et al \(1992, p.176\)](#). For \mathcal{D} , the variance estimator is negatively biased by 16% for the NFI design without intensification. However, it becomes nearly unbiased with SRSwor, SYS, or Poisson sampling for intensification, and exhibits a slight positive bias (7%) with the NFI design for intensification. Under the benchmark strategy without intensification, the variance estimator is negatively biased, which is expected for the ratio estimator; however, the bias is large (-15%). This large bias may be due to the limited sample size of 28 units presented in [Table 3](#). With intensification, the complementary sample with a sample size of more than 100 units, the bias in the variance estimator is then significantly reduced to -3% . For \mathcal{T} , the variance estimators are slightly negatively biased (the absolute values are less than 10%), indicating good performance of the proposed variance estimators.

Strategy			Domain of interest			Territory		
S^a	S^{1b}	S_D^{2b}	$RB_{MC}(\hat{R}_D^b)$ (%)	$V_{MC}(\hat{R}_D^b)$ ($\times 10^{-3}$)	$RB_{MC}\{\hat{V}(\hat{R}_D^b)\}$ (%)	$RB_{MC}(\hat{R}^b)$ (%)	$V_{MC}(\hat{R}^b)$ ($\times 10^{-3}$)	$RB_{MC}\{\hat{V}(\hat{R}^b)\}$ (%)
NFI	NFI		-0.76	14.74	-15.83	-0.20	5.69	-6.11
		SRSwor	-0.24	3.00	-1.02	-0.16	4.05	-5.18
		SYS	-0.30	2.94	0.83	-0.18	4.01	-4.34
		Poisson	-0.29	2.97	0.52	-0.18	4.03	-4.78
	NFI		-0.22	2.81	6.86	-0.15	3.99	-3.69
Uniform	SRSwor		-0.76	14.84	-14.93	-0.31	5.75	-5.89
		SRSwor	-0.08	3.07	-3.23	-0.24	4.13	-6.05

Table 5: Monte Carlo relative bias and Monte Carlo variance for the ratio estimation and variances estimation in the DOI and the entire territory under 50 000 iterations.

6 Discussion

Other NFIs use different methods to obtain precise estimates for the DOI. For example, as noted by [Böckmann et al \(1998\)](#), some German federal states, like Lower Saxony, have established a regional Forest District Inventory (FDI), with higher sampling densities compared to the NFI, to support regional forest management. However, such FDIs are costly, and with budget and resource constraints, more cost-efficient inventory methods are needed ([von Lüpke et al, 2012](#); [von Lüpke, 2013](#); [Hill et al, 2018b](#)). FDIs are typically planned actions at the very beginning of the NFI cycle and are carried out as a part of a stratified sampling technique within a population divided into regions. The French NFI also modulates the spatial intensity of sampling spatially and temporally ([Bouriaud et al, 2023](#)). In contrast to FDIs, intensification occurs as an unplanned response to unpredictable events and unforeseen requirements in forest management and is decided to be conducted during the NFI cycle. This process typically takes place after the NFI sampling and is not a part of the initial NFI sampling strategy. In this article, therefore, we propose a solution within the NFI framework to address local needs. However, similar to FDIs, intensification must be evaluated in light of budget limitations, optimizing sample sizes to fit budget constraints.

Intensification in NFIs was previously explored by [Hansen et al \(2000\)](#), who proposed measuring the plots in previous years and incorporating them into the Annual Forest Inventory System (AFIS)

plots to improve the estimate precision. However, they identified the estimation challenges associated with unequal sampling probabilities, joint selection probabilities, and complexities of the intensified sample to be incorporated into the estimation procedure. Our study concentrated on sampling new plots, which had not been previously sampled in prior years. However, re-measuring existing plots is also a viable intensification strategy, as these plots are already identified and mapped due to the concept of permanent and semi-permanent plots in NFIs (Tomppo et al, 2010). Our formula is applicable in this context, as Hansen et al (2000)'s sampling meets the non-overlapping sample criterion. We also addressed all aforementioned issues by providing total and ratio estimators with their variance estimators under a conditional approach. By proving that all points in the continuous population have the same inclusion density in the sample, despite having different sampling intensities, the design weights for a point in that sample can be more easily calculated, conditionally on the sample size.

Blackard and Patterson (2014) presented various intensification methods that have been implemented in many states in the USA within the Forest Inventory and Analysis (FIA) framework, known as FIA plot intensification. Their intensification is based on the hexagon grid. Sub-regions are created for each FIA second-phase hexagon cell, and new plots are randomly assigned within those sub-regions. In our study, the intensified sample is a partially new sample, meaning that it was part of the NFI first-phase sampling but was not included in the NFI second-phase sampling. In contrast, the intensified plots by the FIA are totally new and have never been selected before, not even in the FIA first-phase sample. Blackard and Patterson (2014) do not disclose the estimation procedure, as if the intensified plots was not used for the estimation. Despite the differences in the way the intensified sample is made in our study as compared to the one made within FIA, the intensified plots and existing FIA base plots may be regarded as non-overlapping since they are all derived from the continuous population. Consequently, the formula in our work is applicable in this context, as long as all plots in both the FIA base sample and the intensified sample have the same inclusion density in the union sample. Furthermore, the existing FIA base plot and intensified plots may be located in the same hexagon cell, resulting in the selection of multiple points within each cell. Duong et al (2024) provided an estimation technique for this situation (i.e. multiple points sampled within a cell), which could enable the use of the intensified plots by the FIA for estimation.

As mentioned in the Introduction, models are utilized for an estimation within the DOI, with notable applications in NFIs, such as the Norwegian (Breidenbach and Astrup, 2012), Finnish (Tomppo, 2006), and Swiss (Magnussen et al, 2014) NFIs. Any estimation method, including the model-assisted and model-based ones, can benefit from intensification since it offers an increase in quantity and quality of field plot data and auxiliary information for calibration and validation (Fassnacht et al, 2024). In this article, we have presented and discussed intensification within the design-based approach; yet, an intensified sample can be useful for improving the estimation using other methods, such as the model-assisted or model-based approaches. The use of an intensified sample in the DOI to improve the prediction models can be an interesting topic in SAE. However, auxiliary data is limited in providing the information for some attributes. For instance, the DOIs could be defined based on biodiversity criteria, such as the need to preserve or protect a specific forest type. This would require extensive biodiversity data, which is difficult to obtain from RS. Therefore, increasing the availability of RS data is not helpful in this situation. The design-based method in our study is capable of addressing all situations where RS is inadaptable. A comparison

between the model approach with limited sample sizes and intensification with increased sampling units for estimating the population total and ratio presents a potential topic for future research.

This paper references the conditional approach, often used with random sample sizes (Holt and Smith, 1979; Rao, 1985; Hidioglou and Särndal, 1985; Gregoire et al, 2016). Rao (1985) introduced a design-based estimator for the small domains using post-stratification adjustment under the conditional approach; however, it exhibits conditional bias. Hidioglou and Särndal (1985) proposed a modified regression estimator (MRE), conditionally unbiased as long as the domain sample size is not smaller than its expected value. While Rao (1985) noted challenges in obtaining conditional inference for complex designs, our article demonstrates the methodology for deriving an unbiased conditional design-based estimator in all situations. Although we can derive unbiased estimators without any assumptions, variance estimation requires simplifying assumptions, which usually lead to a conservative variance estimator. Previous studies, Holt and Smith (1979) and Cochran (1953, p.136), indicated that the conditional variance of the post-stratified mean following SRS is strictly less than the unconditional variance. Gregoire et al (2016) illustrated that obtaining variance under the conditional approach is simpler to handle than the unconditional approach since the post-stratification following SRS functions similarly to stratified random sampling when conditioned on the post-strata sample sizes. Moreover, the conditional approach is fairer than the unconditional one since it makes use of the actual sample size obtained during the sampling process. Since the classification of points inside/outside of the DOI happens after the regular NFI sampling, creating two post-strata within and outside the DOI, our research is an extension of Gregoire et al (2016) to continuous populations, offering simpler and reduced variance estimation in more complex sampling designs. Moreover, combining two samples with different sampling intensities in the context of NFIs has been previously studied by Westfall et al (2018) to use the urban and NFI data for rural-urban assessments. However, they considered the situation of two independent samples, specifically urban and NFI samples, which differ from our context.

We demonstrated that the intensified sample can be utilized for the territory-level estimation; however, the variance gain depends on the DOI's proportion of the total territory and the intensified sample size. The effectiveness of the intensified sample on the territory estimation depends mostly on the proportion of the territory covered by the DOI. In our simulation, the large domain area allows us to take a sufficient sample size that can reduce variance moderately for the entire territory (Tables 4 and 5). However, in smaller domains, the intensified sample may not be sufficient to lower the variance for the territory estimation. The near unbiasedness in the variance estimation for the ratio estimator with SRS_{swr}, SYS, and Poisson sampling for intensification in Table 5 may depend on the simulation set-up and the intensified sample size, as illustrated by the simulation results. Further research of different DOI areas and different intensified samples with different sample sizes could explore how the DOI area and the intensified sample affect the entire territory estimation.

Intensification based on the first phase of an existing NFI sample is feasible through various sampling designs. Our findings show that, among the sampling designs considered in the simulation study, the choice of a sampling design has a limited impact on the estimation of the outcome that we considered in the simulation study, with sample size remaining comparable across designs. However, the choice of suitable designs for intensification depends on factors like the attributes of

interest, the type of forest, the required geographical balance, or particular constraints on the sample size. For example, the NFI design for intensification is the most effective technique for achieving spatial balance and reducing variance under the simulation set-up. On the other hand, SRS_{swr} for intensification in our simulation study may be useful since employing a fixed sample size, which could be preferred in practice to aid in decision-making concerning the constrained sample size. No design can be deemed the best, giving users the flexibility to select their preferred method. Moreover, intensification can be applied within any particular framework to obtain the NFI sample.

The sample size of the intensified sample in our simulation may not be applicable in real-world scenarios, where two main constraints affect the size: the minimum number of points needed for optimal precision and the maximum number allowed within budget limits. For the first constraint, [Cochran \(1953\)](#) and [Yamane \(1973\)](#) proposed a formula for determining sample size based on the desired degree of precision. Under the second constraint, models linking cost and sample size are needed. The sample size is then bounded and may not achieve the desired one due to significant budget constraints. In some cases, budget constraints prevent additional samples, requiring a consistent sample size across the territory. In this situation, a possibility is to decrease the number of sampling units outside the DOI to maintain the overall budget while increasing units within the DOI. However, this approach may reduce efficiency in estimating across the entire territory. In this paper, the sampling strategy to increase the DOI sample size and the estimates provided are themselves important findings of our research. Although our article does not focus on optimization, achieving optimal sample sizes within precision and budget limits can be an interesting perspective for further studies.

7 Conclusion

This paper uses intensification to increase the local sample size and produce a design-based estimation for the DOI and the entire territory. Regardless of the intensified method, we compute the design weights of all the units to combine samples under the conditional approach. Our findings show that the proposed total and ratio estimators, along with their corresponding variance estimators, are effective in estimating within the DOI and across the entire territory. Our results indicate that the sampling designs considered in the simulation study to obtain the intensified sample perform quite similarly, offering the users the flexibility in choosing a preferred design. Some methods have better control of the sample size than others; however, their expected sample sizes are approximately the same.

Declarations

- Conflict of interest/Competing interests: No competing or conflict of interest.
- Consent for publication: Authors are giving consent to publish the article in the submitted journal.
- Data availability: the dataset is available online under DOI: 10.5281/zenodo.10252806

CRedit authorship contribution statement

Trinh H.K. Duong: Writing – original draft, Visualization, Methodology, Formal analysis, Investigation, Data curation, Conceptualization, Resources, Software.

Guillaume Chauvet: Writing – review & editing, Validation, Supervision, visualization, Methodology, Formal analysis, Investigation, Conceptualization, Funding acquisition, Project administration.

Olivier Bouriaud: Writing – review & editing, Validation, Supervision, Visualization, Methodology, Formal analysis, Investigation, Data curation, Conceptualization, Funding acquisition, Project administration, Resources, Software.

APPENDIX

Appendix A Proof of formula (2)

We need the following additional assumptions:

H1: For any first-phase sample s^a , we have

$$\mathbb{P}(x \in S^{1b} | S^a = s^a, n^{1b}) = \frac{n^{1b}}{n^a}.$$

H2: S^a and n^{1b} are independent.

Assumption (H1) states that the second-phase sample, conditioned on S^a and on the sample size n^{1b} , is performed with equal inclusion probabilities. Since conditionally on S^a only, S^{1b} is selected with equal inclusion probabilities, this seems a reasonable assumption which should be respected with most second-phase sampling designs. Assumption (H2) states that n^{1b} does not depend on the first-phase sample S^a , which seems reasonable too.

Under these two assumptions, we have

$$\begin{aligned} E(\hat{\tau}_y^{1b} | S^a, n^{1b}) &= \frac{A_T}{n^{1b}} \sum_{x \in S^a} y(x) \times \underbrace{\mathbb{P}(x \in S^{1b} | S^a, n^{1b})}_{\frac{n^{1b}}{n^a} \text{ from assumption (H1)}} \\ &= \frac{A_T}{n^a} \sum_{x \in S^a} y(x). \end{aligned}$$

Then

$$\begin{aligned} E(\hat{\tau}_y^{1b} | n^{1b}) &= E\{E(\hat{\tau}_y^{1b} | S^a, n^{1b}) | n^{1b}\} \\ &= E\left\{\frac{A_T}{n^a} \sum_{x \in S^a} y(x) \middle| n^{1b}\right\} \\ &= E\left\{\frac{A_T}{n^a} \sum_{x \in S^a} y(x)\right\} \text{ from assumption (H2)} \\ &= \tau_y. \end{aligned}$$

Appendix B Proof of all the points in \mathcal{D} having the same inclusion density in S_D^b

The inclusion probability in S_D^b , conditionally on S_D^a , is denoted as

$$\pi_D^{b|a}(x) = Pr(x \in S_D^b | S_D^a) \text{ for } x \in S_D^a.$$

Due to the negative coordination between S_D^{1b} and S_D^{2b} , we have

$$\begin{aligned} \pi_D^{b|a}(x) &= Pr(x \in S_D^{1b} \cup S_D^{2b} | S_D^a) \\ &= Pr(x \in S_D^{1b} | S_D^a) + Pr(x \in S_D^{2b} | S_D^a) \\ &= \pi_D^{1b|a}(x) + \pi_D^{2b|a}(x) \end{aligned}$$

The design weight, conditionally on S_D^a , is denoted as

$$d_D^{b|a}(x) = \frac{1}{\pi_D^{b|a}(x)} \text{ for } x \in S_D^b. \quad (\text{B1})$$

The design weight in (B1) is used to obtain the conditional design-unbiased estimator for the total value of the NFI-P1 sample S_D^a in \mathcal{D} . To get the design weight for obtaining the total estimator of the DOI \mathcal{D} , we compute as

$$d_D^b(x) = d_D^a(x) \times d_D^{b|a}(x) \text{ for } x \in S_D^b.$$

The inclusion density function of all points $x \in \mathcal{D}$ is then obtained by

$$\pi_D^b(x) = \frac{1}{d_D^b(x)} \text{ for } x \in \mathcal{D}.$$

Therefore, all the points in \mathcal{D} have the same inclusion density in S_D^b .

Appendix C Monte Carlo computation

The sampling and estimation steps are repeated $M = 50\,000$ times. Given that the estimators and variance estimators are studied based on conditioning, both the Monte Carlo (MC) expectation and variance are computed conditionally. The simulations are organized into groups based on the combination of possible values of the sample sizes on which we condition. For each group $g = 1, \dots, G$, the MC expectation and variance of an estimator $\hat{\theta}$ for a parameter θ are defined as

$$E_{MC,g}(\hat{\theta}) = \frac{1}{M_g} \sum_{m=1}^{M_g} \hat{\theta}_m,$$

$$V_{MC,g}(\hat{\theta}) = \frac{1}{M_g - 1} \sum_{m=1}^{M_g} \left\{ \hat{\theta}_m - \frac{1}{M_g} \sum_{m'=1}^{M_g} \hat{\theta}_{m'} \right\}^2$$

where M_g is the number of iterations in group g and $\hat{\theta}_m$ is the value taken by the estimator for the m -th sample. The overall Conditional MC expectation and variance are computed as

$$\begin{aligned} E_{MC}(\hat{\theta}) &= \frac{\sum_{g=1}^G M_g E_{MC,g}(\hat{\theta})}{\sum_{g=1}^G M_g}, \\ V_{MC}(\hat{\theta}) &= \frac{\sum_{g=1}^G M_g V_{MC,g}(\hat{\theta})}{\sum_{g=1}^G M_g}, \end{aligned} \quad (\text{C1})$$

and the MC percent relative bias is

$$RB_{MC}(\hat{\theta}) = 100 \times \frac{E_{MC}(\hat{\theta}) - \theta}{\theta}.$$

For example, the groups g correspond to the given combination of the sample sizes n_D^a and n_D^b when computing the MC expectation and variance for the estimators $\hat{\tau}_{Dy}$ and variance estimators $\hat{V}(\hat{\tau}_{Dy})$ defined in equations (3) and (4) in \mathcal{D} . Meanwhile, the groups g will depend on the combination of n_D^a , n_D^b , and n_D^c when computing for the estimators $\hat{\tau}_y$ and variance estimators $\hat{V}(\hat{\tau}_y)$ defined in equations (6) and (7) in \mathcal{T} .

For the total estimators $\hat{\tau}_{Dy}$ and $\hat{\tau}_y$, the MC percent relative bias is computed with the true values $\theta = \tau_{Dy}$ and $\theta = \tau_y$, respectively. Meanwhile, for the variance estimators $\hat{V}(\hat{\tau}_{Dy})$ and $\hat{V}(\hat{\tau}_y)$, the MC percent relative bias is computed with the replacement of the unknown true variances $\theta = V(\hat{\tau}_{Dy})$ and $\theta = V(\hat{\tau}_y)$ by the Conditional MC variance in the formula (C1). The same approach is applied for the ratio estimators and their variance estimators.

References

- Amani M, Salehi B, Mahdavi S, et al (2019) Separability analysis of wetlands in canada using multi-source sar data. *GIScience & Remote Sensing* 56(8):1233–1260
- Astrup R, Rahlf J, Bjørkelo K, et al (2019) Forest information at multiple scales: development, evaluation and application of the norwegian forest resources map sr16. *Scandinavian Journal of Forest Research* 34(6):484–496
- Bechtold WA, Patterson PL (2005) The enhanced forest inventory and analysis program–national sampling design and estimation procedures. No. 80 in *Gen.Tech.Rep.SRS*, USDA Forest Service, Southern Research Station
- Best N, Richardson S, Clarke P, et al (2019) A comparison of model-based methods for small area estimation

- Blackard JA, Patterson PL (2014) National FIA plot intensification procedure Report, vol 329. Citeseer
- Böckmann T, Spellmann H, Hüsing F (1998) New concept and further developments in forest management in lower saxony.
- Bouriaud O, Morneau F, Bontemps JD (2023) Square-grid sampling support to reconcile systematicity and adaptivity in periodic spatial surveys of natural populations. *Journal of Vegetation Science* 34(3):e13195
- Bouriaud O, Brion P, Chauvet G, et al (2024) The weight share method in forest inventories: refining the relation between points and trees. *Canadian Journal of Forest Science* 54(10):1129–1141
- Breidenbach J, Astrup R (2012) Small area estimation of forest attributes in the norwegian national forest inventory. *European Journal of Forest Research* 131:1255–1267
- Breidenbach J, McRoberts RE, 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
- Breidenbach J, Granhus A, Hysten G, et al (2020) A century of national forest inventory in norway—informing past, present, and future decisions. *Forest ecosystems* 7:1–19
- Chauvet G, Bouriaud O, Brion P (2023) An extension of the weight share method when using a continuous sampling frame. *Survey Methodology* 49(1). Paper available at <http://www.statcan.gc.ca/pub/12-001-x/2023001/article/00011-eng.htm>
- Cochran WG (1953) *Sampling Techniques*. J. Wiley Sons, New York
- Cochran WG (1977) *Sampling Techniques*, 3rd edn. Wiley and Sons, New York
- Cordy CB (1993) An extension of the horvitz—thompson theorem to point sampling from a continuous universe. *Statistics & Probability Letters* 18(5):353–362
- Dettmann GT, Radtke PJ, Coulston JW, et al (2022) Review and synthesis of estimation strategies to meet small area needs in forest inventory. *Frontiers in Forests and Global Change* 5:813569
- Dirnberger GF, Sterba H (2014) A comparison of different methods to estimate species proportions by area in mixed stands. *Forest Systems* 23(3):534–546
- Duong TH, Bouriaud O, Chauvet G (2024) A new sampling framework for spatial surveys with application to the french national forest inventory. *Journal of Survey Statistics and Methodology* p smae045
- Duong THK, Vega C, Renaud JP, et al (2023) A large-scale artificial forest tree population to for sampling and estimation methods simulations. Zenodo DOI: 10.5281/zenodo.10252806
- Fabrizi E, Żądło T (2018) *Survey sampling and small-area estimation*

- Fassnacht FE, White JC, Wulder MA, et al (2024) Remote sensing in forestry: current challenges, considerations and directions. *Forestry: An International Journal of Forest Research* 97(1):11–37
- Fuller WA (1999) Environmental surveys over time. *Journal of Agricultural, Biological, and Environmental Statistics* pp 331–345
- Georgakis A (2019) Small area estimation in forest inventories. In: Seventh International Conference On Environmental Management, Engineering, Planning And Economics (CEMEPE 2019) And SECOTOX Conference. Mykonos island, Greece
- Goerndt ME, Monleon VJ, Temesgen H (2013) Small-area estimation of county-level forest attributes using ground data and remote sensed auxiliary information. *Forest Science* 59(5):536–548
- Grafström A, Ekström M, Jonsson BG, et al (2019) On combining independent probability samples. *Survey Methodology* 45(2):349–364
- Gregoire TG, Ringvall AH, Ståhl G, et al (2016) Conditioning post-stratified inference following two-stage, equal-probability sampling. *Environmental and ecological statistics* 23:141–154
- Gschwantner T, Gabler K, Schadauer K, et al (2010) National forest inventories reports: Austria. In: Tomppo E, Gschwantner T, Lawrence M, et al (eds) *National forest inventories: Pathways for Common Reporting*. Springer Science+Business Media, London, New York, chap 1, p 57–72
- Guldin RW (2021) A systematic review of small domain estimation research in forestry during the twenty-first century from outside the united states. *Frontiers in Forests and Global Change* 4:695929
- Hansen MH, Schreuder HT, Heinzen D (2000) Selection of plot remeasurement in an annual inventory. UNITED STATES DEPARTMENT OF AGRICULTURE FOREST SERVICE GENERAL TECHNICAL REPORT NC pp 67–78
- Hastie T, Tibshirani R, Friedman J, et al (2005) The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer* 27(2):83–85
- Hidioglou M, Särndal C (1985) An empirical study of some regression estimators for small domains. *Survey Methodology* 11(1):65–77
- Hill A (2018) Integration of small area estimation procedures in large-scale forest inventories. PhD thesis, ETH Zurich
- Hill A, Buddenbaum H, Mandallaz D (2018a) Combining canopy height and tree species map information for large-scale timber volume estimations under strong heterogeneity of auxiliary data and variable sample plot sizes. *European Journal of Forest Research* 137:489–505
- Hill A, Mandallaz D, Langshausen J (2018b) A double-sampling extension of the german national forest inventory for design-based small area estimation on forest district levels. *Remote Sensing* 10(7):1052

- Holt D, Smith TF (1979) Post stratification. *Journal of the Royal Statistical Society Series A: Statistics in Society* 142(1):33–46
- Kangas A, Maltamo M (2006) *Forest inventory: methodology and applications*, vol 10. Springer Science & Business Media
- Kermorvant C, D'amico F, Bru N, et al (2019) Spatially balanced sampling designs for environmental surveys. *Environmental monitoring and assessment* 191(8):524
- von Lüpke N (2013) Approaches for the optimisation of double sampling for stratification in repeated forest inventories
- von Lüpke N, Hansen J, Saborowski J (2012) A three-phase sampling procedure for continuous forest inventory with partial re-measurement and updating of terrestrial sample plots. *European Journal of Forest Research* 131:1979–1990
- Magnussen S (2015) Arguments for a model-dependent inference? *Forestry: An International Journal of Forest Research* 88(3):317–325
- Magnussen S, Mandallaz D, Breidenbach J, et al (2014) National forest inventories in the service of small area estimation of stem volume. *Canadian Journal of Forest Research* 44(9):1079–1090
- Mandallaz D (2007) *Sampling techniques for forest inventories*. Chapman and Hall/CRC, New York
- Mauro F, Molina I, García-Abril A, et al (2016) Remote sensing estimates and measures of uncertainty for forest variables at different aggregation levels. *Environmetrics* 27(4):225–238
- McConville KS, Moisen GG, Frescino TS (2020) A tutorial on model-assisted estimation with application to forest inventory. *Forests* 11(2). <https://doi.org/10.3390/f11020244>, URL <https://www.mdpi.com/1999-4907/11/2/244>
- McRoberts RE (2006) A model-based approach to estimating forest area. *Remote Sensing of Environment* 103(1):56–66
- McRoberts RE (2008) The national forest inventory of the united states of america. *Journal of forest and environmental science* 24(3):127–135
- McRoberts RE (2010) Probability-and model-based approaches to inference for proportion forest using satellite imagery as ancillary data. *Remote Sensing of Environment* 114(5):1017–1025
- McRoberts RE, Tomppo EO, Czaplewski RL (2015) Sampling designs for national forest assessments. *knowledge reference for national forest assessments* pp 23–40
- Neyman J (1934) On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society* 97(4):558–606

- Opsomer JD, Breidt FJ, Moisen GG, et al (2007) Model-assisted estimation of forest resources with generalized additive models. *Journal of the American Statistical Association* 102(478):400–409
- Prentius W, Zhao X, Grafström A (2021) Combining environmental area frame surveys of a finite population. *Journal of Agricultural, Biological and Environmental Statistics* 26:250–266
- Rao J (1985) Conditional inference in survey sampling. *Survey Methodology* 11(1):15–31
- Rao JN, Molina I (2015) *Small area estimation*. John Wiley & Sons
- Renaud JP, Sagar A, Barbillon P, et al (2022) Characterizing the calibration domain of remote sensing models using convex hulls. *International Journal of Applied Earth Observation and Geoinformation* 112:102939
- Sagar A, Vega C, Bouriaud O, et al (2022) Multisource forest inventories: A model-based approach using k-nn to reconcile forest attributes statistics and map products. *ISPRS Journal of Photogrammetry and Remote Sensing* 192:175–188
- Costa Saenz de San Pedro A, Satorra A, Ventura E (2003) An empirical evaluation of small area estimators. *UPF Economics and Business Working Paper* (674)
- Särndal CE, Swensson B, Wretman J (1992) *Model assisted survey sampling*. Springer-Verlag Publishing, New York
- Scott CT, Bechtold WA, Reams GA, et al (2005) Sample-based estimators used by the forest inventory and analysis national information management system. The enhanced forest inventory and analysis program—national sampling design and estimation procedures pp 43–67
- Sirin A, Medvedeva M, Maslov A, et al (2018) Assessing the land and vegetation cover of abandoned fire hazardous and rewetted peatlands: Comparing different multispectral satellite data. *Land* 7(2):71
- Ståhl G, Gobakken T, Saarela S, et al (2024) Why ecosystem characteristics predicted from remotely sensed data are unbiased and biased at the same time—and how this affects applications. *Forest Ecosystems* 11:100164
- Stevens Jr DL (1997) Variable density grid-based sampling designs for continuous spatial populations. *Environmetrics: The official journal of the International Environmetrics Society* 8(3):167–195
- Stevens Jr DL, Olsen AR (2004) Spatially balanced sampling of natural resources. *Journal of the American statistical Association* 99(465):262–278
- Tomppo E (2004) Resource assessment | forest resources. In: Burley J (ed) *Encyclopedia of Forest Sciences*. Elsevier, Oxford, p 965–973, <https://doi.org/https://doi.org/10.1016/B0-12-145160-7/00156-3>, URL <https://www.sciencedirect.com/science/article/pii/B0121451607001563>
- Tomppo E (2006) The finnish multi-source national forest inventory—small area estimation and map production. *Forest inventory—methodology and applications* Springer, Netherlands pp 195–224

- Tomppo E, Gschwantner T, Lawrence M, et al (eds) (2010) National forest inventories: Pathways for Common Reporting. Springer Science+Business Media, London, New York
- Westfall JA, Patterson PL, Edgar CB (2018) Integrating urban and national forest inventory data in support of rural–urban assessments. *Forestry: An International Journal of Forest Research* 91(5):641–649
- White JC, Tompalski P, Vastaranta MA, et al (2017) A model development and application guide for generating an enhanced forest inventory using airborne laser scanning data and an area-based approach. Natural Resources Canada
- Wolter K (2007) Introduction to variance estimation
- Yamane T (1973) *Statistics: An introductory analysis*