Received: 5 April 2017 | Revised: 8 September 2017 | Accepted: 11 September 2017

DOI: 10.1002/env.2480

RESEARCH ARTICLE

WILEY

The continuous population approach to forest inventories and use of information in the design

A. Grafström[1] | S. Schnell[1] | S. Saarela[1] | S. P. Hubbell[2] | R. Condit[2]

[1]Department of Forest Resource Management, Swedish University of Agricultural Sciences, Umeå, Sweden

[2]Smithsonian Tropical Research Institute, Balboa, Panama

Correspondence
A. Grafström, Department of Forest Resource Management, Swedish University of Agricultural Sciences, SE-90183 Umeå, Sweden.
Email: anton.grafstrom@slu.se

An extended theoretical framework for the continuous population approach to forest inventories is derived. Here, we treat a simultaneous selection of sample points with any prescribed sampling intensity over a continuous population. Different ways to use available auxiliary information, for example, from remote sensing, by selection of approximately balanced or spatially balanced samples are considered. A large data set of spatially continuous individual tree-level data is used to demonstrate the potential of these theoretical approaches. This study shows new ways to integrate remote sensing information in designs for forest inventory applications, which can significantly reduce the variance of the Horvitz–Thompson estimator for target variables related to the auxiliary information.

KEYWORDS

balanced sampling, cube method, Horvitz–Thompson estimator, local pivotal method, Monte Carlo, representative samples

1 | INTRODUCTION

When sampling forest or other natural resources, where discrete population elements are distributed spatially over the landscape, it can often be a natural choice to apply a continuous population approach to survey sampling (Gregoire & Valentine, 2008, chapter 10). A main reason for using an areal sampling frame, which consists of infinitely many point locations, is that it may be impossible to compile a suitable list frame of units or clusters of units (Gregoire & Valentine, 2008, p. 207). The typical setup in forest inventories is that sample locations are selected by some sampling designs and that, at each location, a certain response design is applied to select a number of trees into the sample. Applying standard sampling theory, this setup leads to logical inconsistencies (Mandallaz, 1991, p. 31). Further, the same tree can be selected multiple times from different locations, and there are problems due to the spatial structure of the population because a tessellation of the study area in non-overlapping elements is impossible in many of the standard response designs (de Vries, 1986; Mandallaz, 1991).

By applying a continuous population approach, it becomes necessary that attribute densities can be observed at any point location in the study area. Therefore, the attributes of discrete population elements need to be made continuous. In forest inventories, the standard approach to derive such a surface of an attribute is the application of the inclusion zone concept. It was first presented by Grosenbaugh (1952) in connection with horizontal point sampling that includes trees into a sample proportional to their stem diameter. Grosenbaugh and Stover (1957) later described trees surrounded by imaginary rings proportional to tree diameter (point or relascope sampling) for illustration. Mandallaz (1991) was the first to present a unified approach of sampling theory for forest inventories based on infinite populations and use the term *support areas*. Later, the term *inclusion area* was introduced (Schreuder, Gregoire, & Wood, 1993, p. 114), which eventually evolved into inclusion zones. In general, the inclusion zone is a geometrical construct that is useful for deriving

inclusion probabilities of individual population units. Each population unit has its own inclusion zone with size and shape depending on the respective response design used for selecting trees into the sample at a given sample location. Inclusion zones are constructed in such a way that any point selected from within a unit's inclusion zone would select this unit into the sample. A comprehensive treatise of the inclusion zone concept for various response designs common in surveys of natural resources was given by Gregoire and Valentine (2008).

To create a continuous population from discrete population units, the units' attributes are spread over the area of the inclusion zones to create an attribute density, for example, kilogram of dry biomass per unit area. As inclusion zones do naturally overlap to some degree, attribute densities are summed up at overlapping locations to create an attribute surface. Such an alternative view of forest sampling was first described in Roesch, Green, and Scott (1993), and figures of attribute or response surfaces can be found in Fehrmann, Gregoire, and Kleinn (2012).

Having defined the attribute density, we can apply point sampling. Estimation is based on a prescribed sampling intensity function and the attribute density function. The sampling intensity function has the property that when integrated over any area, it gives the expected number of sample points that will fall into that area. In the current study, we extend the continuous population approach for forest inventories by introducing balanced and spatially balanced sampling within this framework. The auxiliary information, used to balance the samples, becomes available for use in the design by a double sampling procedure. To achieve approximately balanced samples where Horvitz–Thompson (HT) estimates of auxiliary totals are approximately equal to the corresponding population totals, we utilised the cube method (CM). The CM was introduced for finite populations by Deville and Tillé (2004). We also show how the local pivotal method (LPM), introduced by Grafström, Lundström, and Schelin (2012), can be applied for selecting samples that are well spread (spatially balanced) in the space of the auxiliary information. The LPM has been shown to be among the best designs for spreading samples in auxiliary spaces (Benedetti, Piersimoni, & Postiglione, 2015; Grafström & Ringvall, 2013; Grafström, Saarela, & Ene, 2014; Grafström & Schelin, 2014).

In Section 2, we introduce notation and the theoretical foundations for the continuous population approach to forest inventories. The setting and notation for auxiliary information are provided in Section 3. Our suggested procedures for double sampling are presented in Section 4. A simulation example with real data is provided in Section 5. This paper ends with a discussion in Section 6.

2 | CONTINUOUS POPULATION APPROACH

The finite population of interest is located on a subset $F \subset \mathbb{R}^2$ of the Euclidean plane with area $\ell(F)$. Here, the finite population $U_T = \{1, 2, \dots, N_T\}$ consists of N_T objects (trees) represented as points in the continuous population F. We wish to estimate some characteristic of the finite population, usually a population total, by the means of a probability sample. Circle plot inventories are commonly used for this purpose, and we examine some different possibilities to sample such a population under a continuous population approach with possibly variable intensity and in the presence of auxiliary information.

Let $S_T \subset U_T$ denote the random sample of identifiers. We define S_i as the number of inclusions of object i. The inclusion probability of object i is $\pi_i = \Pr(i \in S_T)$, and we let $E_i = E(S_i)$ denote the expected number of inclusions. The variable of interest, which is usually nonnegative and bounded, is denoted by y_i, and thus, for the finite population, we wish to estimate $Y = \sum_{i=1}^{N_T} y_i$. Due to the lack of a sampling frame (list) of the units in U_T, they cannot be sampled directly. Instead, we select our sample from the continuous population F.

Following Cordy (1993), a sampling design of fixed size n on F is defined by a joint distribution of n random variables. These n random variables are random locations within F, and the sample of these n locations is denoted by $S_F = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$. The marginal probability density function (pdf) of $\mathbf{x}_i$ is denoted by $f_i(\mathbf{x})$, and the joint pdf of $\mathbf{x}_i, \mathbf{x}_j, i \neq j$ is denoted by $f_{ij}(\mathbf{x}, \mathbf{x}')$. The sampling intensity is

$$\pi(\mathbf{x}) = \sum_{i=1}^{n} f_i(\mathbf{x}).$$

We require that $\pi(\mathbf{x}) > 0$ for $\mathbf{x} \in F$ (and $\pi(\cdot) = 0$ outside of F). We have $n = \int_F \pi(\mathbf{x}) \mathrm{d}\mathbf{x}$, for a design of fixed size n. The second order sampling intensity for $\mathbf{x}, \mathbf{x}' \in F$ is defined by

$$\pi(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{n} \sum_{j \neq i} f_{ij}(\mathbf{x}, \mathbf{x}').$$

Each object $i \in U_T$ has an inclusion zone $K_i \subset F$, so that object i is included if a sample point falls within K_i. In circle plot inventories, K_i is the intersection of F and a circle centred at the location of object i. Let $I_i(\mathbf{x})$ be equal to 1 if $\mathbf{x} \in K_i$ and 0 otherwise. The number of inclusions of object i can then be written as $S_i = \sum_{\mathbf{x}\in S_F} I_i(\mathbf{x})$, with expected value

$$E(S_i) = E_i = \int_{K_i} \pi(\mathbf{x})\mathrm{d}\mathbf{x} = \int_F I_i(\mathbf{x})\pi(\mathbf{x})\mathrm{d}\mathbf{x}.$$

The density function $Y(\mathbf{x})$ of the target variable on F can be defined in different ways. We give two alternatives for the density. The first one is

$$Y(\mathbf{x}) = \sum_{i\in U_T} \frac{I_i(\mathbf{x})y_i}{\ell(K_i)}, \tag{1}$$

where the sum of the y_is is divided by the area of the respective inclusion zone on the selected plot or cluster. Thus, $Y(\mathbf{x})$ is a per-area density of the target variable (e.g., volume/area). This density has been used, for example, by Mandallaz (2007, p. 56). The density is constructed so that $Y = \int_F Y(\mathbf{x})\mathrm{d}\mathbf{x}$ is identical to the corresponding finite population parameter $\sum_{i\in U_T} y_i$. This follows from

$$Y = \int_F Y(\mathbf{x})\mathrm{d}\mathbf{x} = \int_F \sum_{i\in U_T} \frac{I_i(\mathbf{x})y_i}{\ell(K_i)}\mathrm{d}\mathbf{x} = \sum_{i\in U_T} \frac{y_i}{\ell(K_i)} \int_F I_i(\mathbf{x})\mathrm{d}\mathbf{x} = \sum_{i\in U_T} y_i.$$

As the y_is are finite and all $\ell(K_i)$ are strictly positive, Fubini's theorem allows changing the order of summation and integration in the previous equation.

The second density, which to our knowledge has not previously been used, is given by

$$Y^*(\mathbf{x}) = \sum_{i\in U_T} \frac{I_i(\mathbf{x})\pi(\mathbf{x})y_i}{E_i}. \tag{2}$$

In the second density, the inclusion of object i through a point $\mathbf{x} \in K_i$ is weighted by the design likelihood of the point $\mathbf{x} \in K_i$. Also, for the second density, we have

$$Y = \int_F Y^*(\mathbf{x})\mathrm{d}\mathbf{x} = \int_F \sum_{i\in U_T} \frac{I_i(\mathbf{x})\pi(\mathbf{x})y_i}{E_i}\mathrm{d}\mathbf{x} = \sum_{i\in U_T} \frac{y_i}{E_i} \int_F I_i(\mathbf{x})\pi(\mathbf{x})\mathrm{d}\mathbf{x} = \sum_{i\in U_T} y_i.$$

In the case of a constant sampling intensity with n selected points, $\pi(\mathbf{x}) = n/\ell(F)$ and Equation 2 is equal to Equation 1 as $E_i = n/\ell(F) \cdot \ell(K_i)$. To see some difference between using the two densities, we first need to look at the continuous version of the HT estimator, which is given by (see, e.g., Cordy, 1993)

$$\hat{Y} = \sum_{\mathbf{x}\in S_F} \frac{Y(\mathbf{x})}{\pi(\mathbf{x})}. \tag{3}$$

If the second density (2) is used, then the estimator (3) is equal to the finite population estimator, as follows:

$$\hat{Y}_{U_T} = \sum_{i\in U_T} y_i \frac{S_i}{E_i}, \tag{4}$$

because

$$\sum_{\mathbf{x}\in S_F} \frac{Y^*(\mathbf{x})}{\pi(\mathbf{x})} = \sum_{\mathbf{x}\in S_F} \sum_{i\in U_T} y_i \frac{I_i(\mathbf{x})}{E_i} = \sum_{i\in U_T} y_i \frac{1}{E_i} \sum_{\mathbf{x}\in S_F} I_i(\mathbf{x}) = \sum_{i\in U_T} y_i \frac{S_i}{E_i}. \tag{5}$$

Again, we used Fubini's theorem to change the order of summation in the previous equation 5. The equality (5) does not hold in general for density (1) but only in the case of constant sampling intensity. We note that, in Equation 4, which we get by using $Y^*(\mathbf{x})$, an inclusion of object $i \in U_T$ always receives the same weight $1/E_i$.

Using density (1), we can rewrite Equation 3 as

$$\hat{Y} = \sum_{i\in U_T} y_i \sum_{\mathbf{x}\in S_F} \frac{I_i(\mathbf{x})}{\ell(K_i)\pi(\mathbf{x})}. \tag{6}$$

If we use $Y(\mathbf{x})$ in the HT estimator, then we see from Equation 6 that the weight of an inclusion of object $i \in U_T$ due to the point $\mathbf{x}$ is $(\ell(K_i)\pi(\mathbf{x}))^{-1}$. Thus, for the HT estimator with density $Y(\mathbf{x})$, it makes a difference which point $\mathbf{x}$ led to the inclusion of object $i \in U_T$. For finite population sampling, we do not know of any existing estimator that would use different weights for an inclusion of the same object depending on how it was included. In that sense, $Y^*(\mathbf{x})$ may perhaps be preferable as it always gives the same weight to an inclusion of an object. However, it is interesting that the HT estimator

based on $Y(\mathbf{x})$ uses more detailed design information than the one based on $Y^*(\mathbf{x})$, as with $Y(\mathbf{x})$, both the number of points that fall into the inclusion zones and their respective locations within the zones are used.

The variance, in Sen–Yates–Grundy form for a fixed sample size n, of the HT estimator (3) is given by

$$V_{\mathrm{SYG}}(\hat{Y}) = -\frac{1}{2}\int_F\int_F \left(\pi(\mathbf{x},\mathbf{x}') - \pi(\mathbf{x})\pi(\mathbf{x}')\right)\left(\frac{Y(\mathbf{x})}{\pi(\mathbf{x})} - \frac{Y(\mathbf{x}')}{\pi(\mathbf{x}')}\right)^2 \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}'. \tag{7}$$

The variance estimator for Equation 7 is given by

$$\hat{V}_{\mathrm{SYG}}(\hat{Y}) = -\frac{1}{2}\sum_{\mathbf{x}\in S_F}\sum_{\mathbf{x}'\neq\mathbf{x}} \frac{\pi(\mathbf{x},\mathbf{x}') - \pi(\mathbf{x})\pi(\mathbf{x}')}{\pi(\mathbf{x},\mathbf{x}')}\left(\frac{Y(\mathbf{x})}{\pi(\mathbf{x})} - \frac{Y(\mathbf{x}')}{\pi(\mathbf{x}')}\right)^2, \tag{8}$$

which is unbiased provided that $Y(\cdot)$ is bounded, $\pi(\mathbf{x},\mathbf{x}')$ is positive for all $\mathbf{x}\in F, \mathbf{x}'\in F, \mathbf{x}\neq\mathbf{x}'$, and $\int_F 1/\pi(\mathbf{x})d\mathbf{x} < \infty$ (see Cordy, 1993).

If the $\mathbf{x}$s are independent and identically distributed (i.i.d.) on F, so that $f_i(\cdot) = f_j(\cdot) = f(\cdot)$, then $\pi(\mathbf{x}) = n\cdot f(\mathbf{x})$ and $\pi(\mathbf{x},\mathbf{x}') = n(n-1)f(\mathbf{x})f(\mathbf{x}')$, $\mathbf{x}\neq\mathbf{x}'$. In this case, we call the resulting design *independent random sampling* (IRS), and algebra gives that the variance estimator (8) can be written as

$$\hat{V}_{\mathrm{IRS}}(\hat{Y}) = \frac{1}{n}\frac{1}{n-1}\sum_{\mathbf{x}\in S_F}\left(\frac{Y(\mathbf{x})}{f(\mathbf{x})} - \frac{1}{n}\sum_{\mathbf{x}'\in S_F}\frac{Y(\mathbf{x}')}{f(\mathbf{x}')}\right)^2, \tag{9}$$

or

$$\hat{V}_{\mathrm{IRS}}(\hat{Y}) = \frac{n}{n-1}\sum_{\mathbf{x}\in S_F}\left(\frac{Y(\mathbf{x})}{\pi(\mathbf{x})} - \frac{1}{n}\sum_{\mathbf{x}'\in S_F}\frac{Y(\mathbf{x}')}{\pi(\mathbf{x}')}\right)^2.$$

We recognise the previous formula (9) as the continuous version of the variance estimator we have for sampling with probabilities proportional to size and with replacement (i.e., independent). If we also have uniform intensity $f(\mathbf{x}) = 1/\ell(F)$, $\mathbf{x}\in F$, then we get the standard formula

$$\hat{V}(\hat{Y}) = \frac{(\ell(F))^2}{n}\frac{1}{n-1}\sum_{\mathbf{x}\in S_F}\left(Y(\mathbf{x}) - \bar{Y}_{S_F}\right)^2,$$

where $\bar{Y}_{S_F} = n^{-1}\sum_{\mathbf{x}\in S_F} Y(\mathbf{x})$.

3 | AUXILIARY INFORMATION

Assume that we have access to p auxiliary variables $\mathbf{Z}(\mathbf{x}) = (Z_1(\mathbf{x}), \ldots, Z_p(\mathbf{x}))^T \in \mathbb{R}^p$, which are derived the same way as $Y(\mathbf{x})$, that is, averaged on the same or a similar spatial scale over F. By using the same spatial scale, we are more likely to minimise loss of strength in the relationships between the auxiliary and target variables. In practice, it depends on the plot (or cluster) shape and which format and resolution the auxiliary information has. There are different possibilities on how to utilise $\mathbf{Z}(\mathbf{x})$ in the sampling design. We might want to sample proportionally to one of the (positive) auxiliary variables, say $Z_k(\mathbf{x})$, that is, we let $\pi(\mathbf{x}) \propto Z_k(\mathbf{x})$, hoping that it will lead to a small variation in the ratios $Y(\mathbf{x})/\pi(\mathbf{x})$, thus reducing the variance of $\hat{Y}$. Other options that we will explore include selection of so-called balanced and spatially balanced samples. To achieve this, we perform double sampling.

4 | DOUBLE SAMPLING FOR BALANCE OR SPATIAL BALANCE

First, we select a large sample S_{F1} of $N \gg n$ i.i.d. points over F, with pdf $f(\cdot)\propto\pi(\cdot)$ on F. Then, we derive the value of the auxiliary variables for each point. Because the first sample is large, the distribution of $\mathbf{x}$ in the sample will closely match the corresponding distribution given by $f(\cdot)$ on F. By the Glivenko–Cantelli theorem and its multivariate generalisations (see, e.g., DeHardt, 1971; Wolfowitz, 1954), the empirical distribution of $\mathbf{x}$ in the first sample of size N converges uniformly and almost surely to the distribution given by $f(\cdot)$ as $N\to\infty$. The empirical distribution of the sequence $\mathbf{Z}(\mathbf{x}_i)$, $i = 1, 2, \ldots, N$, of i.i.d. observations is given by

$$F_N(\mathbf{z}) = \frac{1}{N}\sum_{i=1}^{N} I\left(\mathbf{Z}(\mathbf{x}_i)\leq\mathbf{z}\right),$$

where $I(A)$ is the indicator function for the event A, and $\mathbf{Z}(\mathbf{x}) \leq \mathbf{z}$ is short for $Z_1(\mathbf{x}) \leq z_1, Z_2(\mathbf{x}) \leq z_2, \ldots, Z_p(\mathbf{x}) \leq z_p$. The distribution F_N converges uniformly and almost surely towards the distribution

$$F_{\mathbf{Z}}(\mathbf{z}) = \int_F I\left(\mathbf{Z}(\mathbf{x}) \leq \mathbf{z}\right) f(\mathbf{x}) \mathrm{d}\mathbf{x}, \tag{10}$$

as $N \to \infty$. We denote the distribution (10) as the sampling distribution of $\mathbf{Z}(\mathbf{x})$. If $f(\mathbf{x}) = 1/\ell(F)$, then the sampling distribution is equal to the continuous population distribution. The sampling intensity for the first selection here is given by $\pi_1(\mathbf{x}) = Nf(\mathbf{x}) = Nn^{-1}\pi(\mathbf{x})$. The first sample can be treated as a finite population, and for such, there are several novel sampling designs to apply. Next, we look at two such designs that can utilise auxiliary information.

4.1 | Balanced sampling

We can use the CM introduced by Deville and Tillé (2004) to select the so-called balanced samples. A sample S from a finite population of size N is balanced on an auxiliary variable $\mathbf{Q}$ if

$$\sum_{i \in S} \frac{\mathbf{Q}_i}{\pi_i} = \sum_{i=1}^{N} \mathbf{Q}_i, \tag{11}$$

where π_i, $i = 1, 2, \ldots, N$, are the inclusion probabilities. In our double sampling approach, we can treat the realised initial large sample of size N as a finite population with auxiliary variables $\mathbf{Q}_i = \mathbf{Z}(\mathbf{x}_i)/\pi_1(\mathbf{x}_i)$, $i = 1, 2, \ldots, N$. Then, a balanced second sample S with inclusion probabilities $\pi_i = n/N$ produces the final sampling intensity $\pi(\mathbf{x})$ on F and has the property

$$\sum_{i \in S} \frac{\mathbf{Z}(\mathbf{x}_i)}{\pi(\mathbf{x}_i)} = \sum_{i \in S} \frac{\mathbf{Q}_i}{\pi_i} = \sum_{i=1}^{N} \mathbf{Q}_i = \sum_{i=1}^{N} \frac{\mathbf{Z}(\mathbf{x}_i)}{\pi_1(\mathbf{x}_i)}. \tag{12}$$

The second equality in Equation 12 follows from the definition (11) of a balanced sample. This means that the HT estimators of the $\mathbf{Z}$ totals for the size n sample are equal to the HT estimators of the $\mathbf{Z}$ totals for the size N sample. Because N is supposed to be very large, we have

$$\hat{\mathbf{Z}}_1 = \sum_{\mathbf{x} \in S_{F_1}} \frac{\mathbf{Z}(\mathbf{x})}{\pi_1(\mathbf{x})} = \frac{1}{N} \sum_{\mathbf{x} \in S_{F_1}} \frac{\mathbf{Z}(\mathbf{x})}{f(\mathbf{x})} \approx \int_F \mathbf{Z}(\mathbf{x}) \mathrm{d}\mathbf{x} = \mathbf{Z},$$

as $\hat{\mathbf{Z}}_1$ is unbiased for $\mathbf{Z}$ with a variance that approaches zero as $N \to \infty$. Thus, if N is large and the size n sample is balanced for $\mathbf{Q}_i = \mathbf{Z}(\mathbf{x}_i)/\pi_1(\mathbf{x}_i)$, $i = 1, 2, \ldots, N$, then we have the desired result as follows:

$$\hat{\mathbf{Z}} = \sum_{\mathbf{x} \in S_F} \frac{\mathbf{Z}(\mathbf{x})}{\pi(\mathbf{x})} = \hat{\mathbf{Z}}_1 \approx \mathbf{Z}.$$

If $Y(\cdot)$ is linear in $\mathbf{Z}(\cdot)$, that is, $Y(\mathbf{x}) = \mathbf{Z}^T(\mathbf{x})\boldsymbol{\beta}$, then it follows that

$$\hat{Y} = \sum_{\mathbf{x} \in S_F} \frac{\mathbf{Z}^T(\mathbf{x})\boldsymbol{\beta}}{\pi(\mathbf{x})} = \hat{\mathbf{Z}}^T\boldsymbol{\beta} = \hat{\mathbf{Z}}_1^T\boldsymbol{\beta} \approx \mathbf{Z}^T\boldsymbol{\beta} = \int_F \mathbf{Z}^T(\mathbf{x})\boldsymbol{\beta} \mathrm{d}\mathbf{x} = Y.$$

This property makes balanced sampling an attractive option for linear or close to linear relationships between target variables and auxiliary variables.

An approximate variance estimator for the variance of $\hat{Y}$ under balanced sampling from a continuous population is

$$\hat{V}_{\mathrm{CM}}(\hat{Y}) = \frac{n}{n-p} \sum_{\mathbf{x} \in S_F} \left(\frac{e(\mathbf{x})}{\pi(\mathbf{x})} \right)^2, \tag{13}$$

where $e(\mathbf{x}) = Y(\mathbf{x}) - \mathbf{Z}^T(\mathbf{x})\hat{\boldsymbol{\beta}}$ and

$$\hat{\boldsymbol{\beta}} = \left[\sum_{\mathbf{x} \in S_F} \frac{\mathbf{Z}(\mathbf{x})\mathbf{Z}^T(\mathbf{x})}{\pi(\mathbf{x})^2} \right]^{-1} \sum_{\mathbf{x} \in S_F} \frac{\mathbf{Z}(\mathbf{x})Y(\mathbf{x})}{\pi(\mathbf{x})^2}.$$

A corresponding variance estimator to the previous equation 13, but for a finite population, was given by Deville and Tillé (2005). The only difference is that the generalised finite population correction term here has been removed from the sums. However, the variance estimator (13) assumes perfect balance, which will not be the case in general. Thus, the estimator (13) may have a large negative bias for small samples but may be adequate for large samples.

Deville and Tillé (2005) argued that balanced sampling can be seen as a form of calibration at the design level, and thus, use of a variance estimator based on residuals is appropriate.

To find a nearly unbiased variance estimator, we may instead use a two-phase approach in combination with simulation. The main parts of such an approach are explained here, as variance estimation is important for applications. For the first sample S_{F1}, we can (theoretically) construct the HT estimator, as follows:

$$\hat{Y}_1 = \sum_{\mathbf{x}\in S_{F1}} \frac{Y(\mathbf{x})}{\pi_1(\mathbf{x})},$$

and $\hat{Y}$ estimates $\hat{Y}_1$ without bias, that is, $E_2(\hat{Y}|S_{F1}) = \hat{Y}_1$, where $E_2(\cdot)$ denotes the expectation under the second design used to select S_F from S_{F1}. We have from the law of total variance

$$V(\hat{Y}) = E_1\left(V_2\left(\hat{Y}|S_{F1}\right)\right) + V_1\left(E_2\left(\hat{Y}|S_{F1}\right)\right) = E_1\left(V_2\left(\hat{Y}|S_{F1}\right)\right) + V_1(\hat{Y}_1),$$

where E_k, V_k denote the expectation and variance under design k, respectively. An estimator of $V_1(\hat{Y}_1)$ is

$$\hat{V}_1(\hat{Y}_1) = -\frac{1}{2}\sum_{\mathbf{x}\in S_{F1}}\sum_{\mathbf{x}'\neq\mathbf{x}} \frac{\pi_1(\mathbf{x},\mathbf{x}') - \pi_1(\mathbf{x})\pi_1(\mathbf{x}')}{\pi_1(\mathbf{x},\mathbf{x}')}\left(\frac{Y(\mathbf{x})}{\pi_1(\mathbf{x})} - \frac{Y(\mathbf{x}')}{\pi_1(\mathbf{x}')}\right)^2,$$

which is unbiased under the conditions given for Equation 8. This estimator cannot be calculated as we do not observe $Y(\mathbf{x})$ for all points in S_{F1}, but an estimator of it, based on the sample S_F, is

$$\hat{\hat{V}}_1(\hat{Y}_1) = -\frac{1}{2}\sum_{\mathbf{x}\in S_F}\sum_{\mathbf{x}'\neq\mathbf{x}} \frac{\pi_1(\mathbf{x},\mathbf{x}') - \pi_1(\mathbf{x})\pi_1(\mathbf{x}')}{\pi_1(\mathbf{x},\mathbf{x}')\pi_{2|1}(\mathbf{x},\mathbf{x}')}\left(\frac{Y(\mathbf{x})}{\pi_1(\mathbf{x})} - \frac{Y(\mathbf{x}')}{\pi_1(\mathbf{x}')}\right)^2, \tag{14}$$

where $\pi_{2|1}(\mathbf{x},\mathbf{x}')$ is the conditional joint inclusion probability of $\mathbf{x},\mathbf{x}'$ in S_F, given S_{F1}. The previous estimator (14) is conditionally unbiased and hence unbiased for $V_1(\hat{Y}_1)$ provided that all $\pi_{2|1}(\mathbf{x},\mathbf{x}')$ are strictly positive for all S_{F1}, that is, $E_2(\hat{\hat{V}}_1(\hat{Y}_1)) = \hat{V}_1(\hat{Y}_1)$ and thus $E_1E_2(\hat{\hat{V}}_1(\hat{Y}_1)) = V_1(\hat{Y}_1)$. Next, we have

$$V_2\left(\hat{Y}|S_{F1}\right) = -\frac{1}{2}\sum_{\mathbf{x}\in S_{F1}}\sum_{\mathbf{x}'\neq\mathbf{x}} \left(\pi_{2|1}(\mathbf{x},\mathbf{x}') - \pi_{2|1}(\mathbf{x})\pi_{2|1}(\mathbf{x}')\right)\left(\frac{Y(\mathbf{x})}{\pi(\mathbf{x})} - \frac{Y(\mathbf{x}')}{\pi(\mathbf{x}')}\right)^2,$$

where $\pi_{2|1}(\mathbf{x}) = n/N$ is the conditional inclusion probability of $\mathbf{x}$ in S_F given S_{F1}. The variance $V_2(\hat{Y}|S_{F1})$ is unbiasedly estimated by

$$\hat{V}_2\left(\hat{Y}|S_{F1}\right) = -\frac{1}{2}\sum_{\mathbf{x}\in S_F}\sum_{\mathbf{x}'\neq\mathbf{x}} \frac{\pi_{2|1}(\mathbf{x},\mathbf{x}') - \pi_{2|1}(\mathbf{x})\pi_{2|1}(\mathbf{x}')}{\pi_{2|1}(\mathbf{x},\mathbf{x}')}\left(\frac{Y(\mathbf{x})}{\pi(\mathbf{x})} - \frac{Y(\mathbf{x}')}{\pi(\mathbf{x}')}\right)^2, \tag{15}$$

if all $\pi_{2|1}(\mathbf{x},\mathbf{x}')$ are strictly positive for all S_{F1}. Finally, a conditional variance estimator of $V(\hat{Y})$ is given by the sum of Equations 14 and 15, that is,

$$\hat{V}(\hat{Y}) = \hat{V}_2\left(\hat{Y}|S_{F1}\right) + \hat{\hat{V}}_1(\hat{Y}_1). \tag{16}$$

It follows under the same conditions as for Equations 14 and 15 that

$$E_2\left(\hat{V}(\hat{Y})\right) = E_2\left(\hat{V}_2\left(\hat{Y}|S_{F1}\right)\right) + E_2\left(\hat{\hat{V}}_1(\hat{Y}_1)\right) = V_2\left(\hat{Y}|S_{F1}\right) + \hat{V}_1(\hat{Y}_1),$$

and

$$E_1E_2\left(\hat{V}(\hat{Y})\right) = E_1\left(V_2\left(\hat{Y}|S_{F1}\right)\right) + V_1(\hat{Y}_1) = V(\hat{Y}),$$

so Equation 16 is unbiased for $V(\hat{Y})$. Assuming that the first sample is IRS so that we can easily find $\pi_1(\mathbf{x},\mathbf{x}')$, the only unknowns needed for applying this variance estimation are $\pi_{2|1}(\mathbf{x},\mathbf{x}')$. These can be estimated through simulation, by selecting a large number (say 10,000 or more) of samples from S_{F1} using the balanced design. Estimated second-order inclusion probabilities will introduce some bias as they appear in a nonlinear way in the estimator, and additional variance will be introduced for the variance estimator, but both the bias and the additional variance can be made arbitrarily small by increasing the number of simulated samples. A specific simulation-based method for estimating second-order inclusion probabilities for balanced samples selected via the CM is derived and discussed by Breidt and Chauvet (2011), and we refer to that paper for details.

We may also argue that when $N \gg n$, then we may drop $\hat{\hat{V}}_1(\hat{Y}_1)$ and use only $\hat{V}_2(\hat{Y}|S_{F1})$ as an estimator of the variance. Typically, the variance is approximately proportional to the inverse of the sample size. If, for example, $N = 1000n$ and the

balanced design is at most 10 times more efficient than IRS at the same sample size, then we can expect the contribution from $V(\hat{Y}_1)$ to be at most 1% to the total variance.

The use of a simulation-based variance estimator in a simulation study is very computationally intensive and is not the focus of this paper. However, in application, it needs to be done only once and may be a practical solution.

A fast implementation of the CM is available in the `R` package `BalancedSampling` (Grafström & Lisic, 2016). To achieve high entropy (and thus reduce problems with small second-order inclusion probabilities) with this fast implementation, the order of the units should be randomised before applying the method.

4.2 | Spreading the sample in auxiliary space

If $Y(\mathbf{x})$ is smoother as a function of $\mathbf{Z}(\mathbf{x})$ than of $\mathbf{x}$, then it can be an efficient strategy to try to select a sample whose empirical distribution of $\mathbf{Z}(\mathbf{x})$ is close to the sampling distribution of $\mathbf{Z}(\mathbf{x})$. We say that such samples are representative for the sampling distribution of $\mathbf{Z}(\mathbf{x})$, or spatially balanced in the $\mathbf{Z}$-space. The difference between a balanced sample and a spatially balanced sample is illustrated in Figure 1.

We can use Voronoi polytopes to measure how well a sample matches the sampling distribution. The Voronoi polytope p_i for a sample point i includes all points in the population that are closer to i than to any other sample point. The measure

$$B = \frac{1}{n}\sum_{i\in s}(v_i - 1)^2,$$

where $v_i = \sum_{j\in p_i}\pi_j$ for a finite population and $v_i = \int_{p_i}\pi(\mathbf{x})\mathrm{d}\mathbf{x}$ for a continuous population, has been used as a measure of how well a sample fits the sampling distribution by, for example, Stevens and Olsen (2004) and Grafström et al. (2012). A small value of B (close to 0) indicates a good fit. How well a design succeeds in producing spatially balanced samples can be measured by the expected value of B under the design, which usually requires a simulation to estimate. Assuming that $Y(\mathbf{x})$ and $\pi(\mathbf{x})$ are well approximated by smooth functions of $\mathbf{Z}(\mathbf{x})$, spatially balanced samples produce an approximate balance for $Y(\mathbf{x})$, that is, $\hat{Y} \approx Y$. This was shown by Grafström and Lundström (2013). Compared to balanced samples that are optimal for linear relationships, spatially balanced samples can be better for nonlinear relationships.

For the second selection, we can use the LPM by Grafström et al. (2012) with equal inclusion probabilities. The LPM has been shown in some studies to be among the best designs with respect to the spatial balance in the auxiliary space. In the LPM, spatial balance (spread) is achieved by letting neighbouring units/points compete for inclusion in the sample, which leads to small joint inclusion probabilities for nearby units. The recommended distance to use in the auxiliary space, when the LPM is applied, is the Euclidean distance on standardised auxiliary variables. The standardisation is necessary to make the auxiliary variables equally important and thus create the same level of spread for each dimension.

Sampling from a finite population can be viewed as a process of thinning the population, where some population units are removed and the remaining units form the sample. In the second selection, the LPM algorithm makes sure that the thinning is done as evenly as possible in the auxiliary space by using the distance measure to avoid the selection of nearby units. The use of equal probabilities nN^{-1} in the LPM selection is important, as we then achieve the final sampling intensity $\pi(\mathbf{x})$ on F. This procedure can guarantee any prespecified inclusion intensity function while spreading the sample well

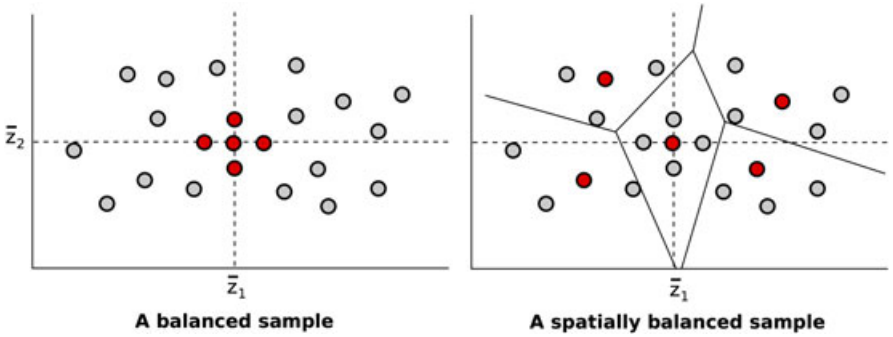

FIGURE 1 Illustration of difference between balanced and spatially balanced samples. The darker (red) dots indicate the samples. A spatially balanced sample is approximately balanced, but a balanced sample is not necessarily approximately spatially balanced. The illustration is for equal probabilities, in which case a balanced sample is mean balanced

into the auxiliary space. The empirical distribution of the auxiliary variables in the final sample will then tend to be close to the sampling distribution (10) in comparison to other sampling designs.

An approximative variance estimator for LPM was derived by Grafström and Schelin (2014), who suggested using differences between neighbours. The continuous version of the variance estimator is

$$\hat{V}_{\text{LPM}}(\hat{Y}) = \frac{1}{2} \sum_{\mathbf{x} \in S_F} \left(\frac{Y(\mathbf{x})}{\pi(\mathbf{x})} - \frac{Y(\mathbf{x}')}{\pi(\mathbf{x}')} \right)^2, \tag{17}$$

where $\mathbf{x}' \neq \mathbf{x}$ is the nearest neighbour to $\mathbf{x}$ in the sample S_F. Distance is supposed to be measured in the same space, in which the sample is spread by LPM, using the same distance function used in the LPM. Another option for variance estimation is to use a local mean variance estimator (see, e.g., Stevens & Olsen, 2003).

5 | SIMULATION EXAMPLE

To evaluate our continuous population approach for forest survey sampling by means of Monte Carlo simulations, spatially continuous tree-level data are required. This requires that, for all trees in a study area, the spatial location and other tree variables, such as diameter at breast height (DBH), tree height, and biomass, must be known. Such data sets are available either from simulations or from a global network of forest research plots distributed across the Americas, Africa, Asia, and Europe (Condit, 1998). For our study, we chose the Barro Colorado Island (BCI) forest dynamics plot in Panama (Condit, Lao, Pérez, Dolins, Foster, & Hubbell, 2012). In addition to tree-level data, some sort of auxiliary data, utilised in the sampling design, are needed. It was not possible for us to obtain auxiliary information in the form of remote sensing data. Because of that, we chose to use auxiliaries available from the tree-level data, where aboveground tree biomass (AGB) is our target variable, and tree density and tree basal area at a stem height of 1.3 m are the auxiliaries that explain AGB.

For real-world applications of the continuous population approach, only auxiliary information is required. Such information is readily available from different remote sensing platforms. In the next two sections, the BCI data set and the specific settings for our simulation are described.

5.1 | BCI forest dynamics plot

BCI is a 1,500-ha large island in the Panama Canal. It belongs together with surrounding peninsulas to the Barro Colorado Nature Monument, which has been a fully protected national biological reserve since 1923. On the island, a 50-ha forest dynamic plot is maintained by the Smithsonian Tropical Research Institute (Condit et al., 2012; Hubbell et al., 1999). To date, seven censi are available where all woody plants (except liana) down to 1 cm in stem diameter at a height of 1.3 m (DBH) have been measured. The forest type is a semi-deciduous lowland moist forest with an average canopy height of 20 to 40 m tall. The greatest part of the plot has been continuously forested for at least 1,000 years with 2 ha cleared during the 19th century but with no logging or extraction since 1923.

From the seven available censi, the latest one from 2010 was used for our simulations. Records with dead, missing, or broken trees have been removed. Diameters of trees/shrubs with multiple stems were identified by unique identifiers and aggregated to one single record preserving basal area in order to avoid duplicate coordinates. After cleaning, the data set contained 243,595 trees, corresponding to an average of 0.497 trees per square meter. The AGB of each tree was predicted using Equation 7 in Chave et al. (2014). The allometric model requires knowledge of DBH, wood specific gravity, and an environmental variable. DBH was available from the census, wood specific gravity was taken from the Global Wood Density Database (Chave et al., 2009; Zanne et al., 2009), and the environmental variable was provided by Chave et al. (2014). In case wood density was not available in the database for some species, a hierarchical approach was applied. First genera averages, the family averages, and finally the average wood specific gravity of the 10 most abundant species (based on basal area) were calculated. Depending on at which level the information was missing, corresponding values were assigned to the individual observations. Following the described procedure, individual tree AGB sums up to a total of 14,200,769 kg, yielding an average of 28.4 kg per square meter.

5.2 | Simulation

We derived the local density for the number of trees, basal area, and AGB using a 10-m-radius circular plot. For simplicity, a buffer of 10 m was added so that all trees would have an inclusion zone of equal size. The local density for the auxiliary variables are defined in our case as

$$\mathbf{Z}(\mathbf{x}) = \sum_{i \in U_T} \frac{I_i(\mathbf{x})\mathbf{z}_i}{\ell(K_i)},$$

where U_T denotes the population of 243,595 trees and $\mathbf{Z}(\mathbf{x}) = (Z_1(\mathbf{x}), Z_2(\mathbf{x}))^T$ is tree density and basal area density. Thus, $\mathbf{z}_i = (1, ba_i)^T$, where ba_i is the basal area of tree i. The derived density functions for all three variables are shown in Figure 2.

When calculating the auxiliary densities for an application, U_T would be replaced by the set of points/pixels for the auxiliary information derived, for example, from remote sensing. The target (AGB) density is defined according to Equation 1, with y_i being the AGB for tree i. In the first large sample, we chose to use $N = 10,000$ points, and in the second selection, we used $n = 40$ points, corresponding to an inventory of 40 circular field plots with 10-m radius. We utilised the CM $\mathbf{Z}(\mathbf{x})$ as balancing variables, and the equal inclusion probabilities are also used as a balancing variable in the CM to guarantee a fixed sample size. For LPM, we used the method with $\mathbf{Z}(\mathbf{x})$ as auxiliary variables and included as a separate case

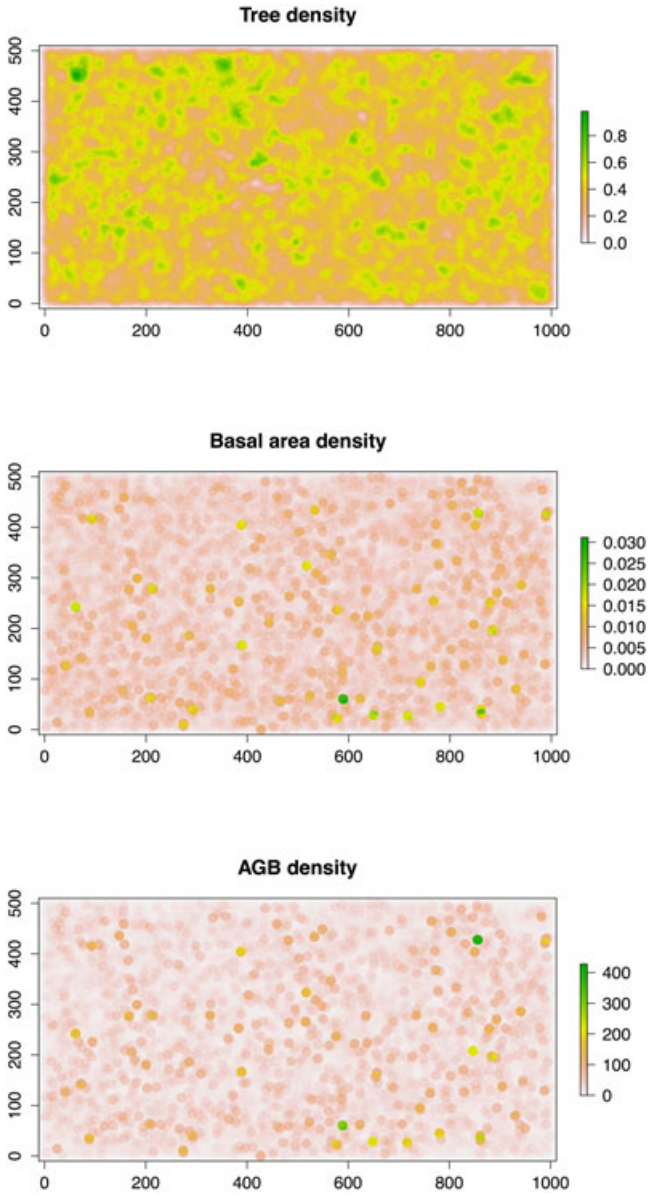

FIGURE 2 Densities for all variables on the Barro Colorado Island forest dynamics plot. AGB = aboveground tree biomass

the geographical coordinates $\mathbf{x} = (x_1, x_2)$ as auxiliary variables to make sure that the samples are well spread geographically. Further, we used a constant sampling intensity and a variable sampling intensity for all designs, which give us two main cases.

Case 1: Constant sampling intensity over the area of size $\ell(F) = 520 \times 1020 = 530{,}400$ m^2 and with sample size $n = 40$. Thus, the sampling intensity was $\pi(\mathbf{x}) = n/\ell(F) = 40/530{,}400$. Included designs are IRS, CM($\mathbf{Z}$) balanced on tree density and basal area, LPM($\mathbf{Z}$) well spread in tree density and basal area, and LPM($\mathbf{Z}, \mathbf{x}$) well spread in tree density, basal area, and geographical coordinates.

Case 2: Variable sampling intensity, proportional to basal area, over the area of size $\ell(F) = 530{,}400$ m^2 and with sample size $n = 40$. The sampling intensity was $\pi(\mathbf{x}) = n(Z_2(\mathbf{x}) + 0.001)/(Z_2 + 0.001\ell(F))$, where $Z_2 = \int_F Z_2(\mathbf{x})\mathrm{d}\mathbf{x}$ and $Z_2(\mathbf{x})$ are the basal area density. The same designs as in case 1 are used, but here, all designs are used with the variable intensity. The two different target densities $Y(\mathbf{x})$ (case 2a) and $Y^*(\mathbf{x})$ (case 2b) are utilised.

The results from the simulations are presented in Table 1. We see that, for case 1 (constant sampling intensity), we achieved design effects for the CM and the LPM in the range of 0.21 to 0.32 in comparison to IRS of the same sample size. The CM achieved slightly better results than the LPM for case 1. Adding geographical coordinates to LPM has a slight negative effect in our case. This can be explained by the rather strong auxiliaries in $\mathbf{Z}$. Conditioned on $\mathbf{Z}$, the spatial location do not add any information. However, including $\mathbf{x}$ in the space that we spread, the sample slightly reduces the spread in the $\mathbf{Z}$-subspace. Hence, we see a slight negative effect by adding spread also geographically. For case 2, where we used variable sampling intensity, we achieved design effects for the CM and the LPM in the range of 0.06 to 0.11 in comparison to IRS with constant sampling intensity and of the same sample size. In this case, the LPM performed slightly better than the CM. For IRS, the results were significantly improved by using the variable sampling intensity. The additional gain by using the LPM or the CM was smaller than that for the constant sampling intensity.

With our settings, the target density $Y(\mathbf{x})$ gave slightly better results than using $Y^*(\mathbf{x})$. However, we cannot draw any general conclusions about the two densities based only on this example, and the results may also be affected by the chosen density of the auxiliary variables.

With a constant sampling intensity, the variance estimator for the LPM performed rather poorly and had significant negative bias of up to −26%. The most likely reason for the large negative bias with a constant sampling intensity is that the ratios $Y(\mathbf{x})/\pi(\mathbf{x})$ then have a very skewed distribution in our population and that the sample size was rather small. The skewness of the ratios $Y(\mathbf{x})/\pi(\mathbf{x})$ was significantly reduced by using the variable sampling intensity, which improved the variance estimation.

TABLE 1 Simulation results for Barro Colorado Island forest dynamics plot

Design	Design effect	V	mean($\hat{V}$)	rel.bias($\hat{V}$)
Case 1: Constant sampling intensity; target density $Y(\mathbf{x}) = Y^*(\mathbf{x})$				
IRS	1	5.54×10^{12}	5.50×10^{12}	−0.7%
CM($\mathbf{Z}$)	0.21	1.18×10^{12}	4.00×10^{11}	−66.1%
LPM($\mathbf{Z}$)	0.26	1.44×10^{12}	1.06×10^{12}	−26.4%
LPM($\mathbf{Z}, \mathbf{x}$)	0.32	1.77×10^{12}	1.51×10^{12}	−14.7%
Case 2a: Variable sampling intensity; target density $Y(\mathbf{x})$				
IRS	0.13	7.46×10^{11}	7.55×10^{11}	1.2%
CM($\mathbf{Z}$)	0.07	3.96×10^{11}	2.75×10^{11}	−30.6%
LPM($\mathbf{Z}$)	0.06	3.07×10^{11}	2.90×10^{11}	−5.5%
LPM($\mathbf{Z}, \mathbf{x}$)	0.07	3.87×10^{11}	3.99×10^{11}	3.1%
Case 2b: Variable sampling intensity; target density $Y^*(\mathbf{x})$				
IRS	0.23	1.29×10^{12}	1.29×10^{12}	0.0%
CM($\mathbf{Z}$)	0.10	5.61×10^{11}	3.88×10^{11}	−30.8%
LPM($\mathbf{Z}$)	0.07	4.04×10^{11}	3.85×10^{11}	−4.7%
LPM($\mathbf{Z}, \mathbf{x}$)	0.11	5.98×10^{11}	6.04×10^{11}	1.0%

Note. A total of 10,000 samples of 40 circular plots with 10-m radius were selected with each design. The table shows design effect, which is the empirical mean square error of the design relative to independent random sampling (IRS) with constant sampling intensity. Also, the empirical variance (V), the average, and relative bias (rel.bias) of variance estimators are given. True total of aboveground tree biomass was 14,200,769 kg. CM = cube method; LPM = local pivotal method.

For the CM, we applied the approximate variance estimator (13) that assumes perfect balance and a large sample size. Obviously, the sample size $n = 40$ was not sufficient. The variance estimator had a relative bias of about −66% for equal sampling intensity and about −30% for variable sampling intensity. As we performed a simulation with a large number of samples, it was not feasible to estimate conditional second-order inclusion probabilities for each initial sample. Thus, we could not apply the unbiased estimator (16) by plugging in the estimated second-order inclusion probabilities. However, in an application with a single sample, that could be a better alternative.

6 | DISCUSSION

We presented a general statistical framework for using spatially continuous auxiliary information in balanced or spatially balanced designs for forest inventories. Our example shows that, with auxiliary information highly correlated with forest attributes, we can greatly improve the performance of the standard design-unbiased HT estimator compared to simpler designs. Even greater improvements than that demonstrated in our example can be expected for larger sample sizes. We foresee many potential applications of the sampling approaches presented here, as they may be used to select approximately balanced or spatially balanced samples from any continuous population or multivariate distribution. By using, for example, the LPM with equal probabilities on a large sample with independent observations from a multivariate distribution, we get a sample that represents the distribution much better than a sample with independent observations of the same size.

The LPM with constant sampling intensity may be preferable in multipurpose inventories because it provides a representative sample, where the sample distribution of auxiliary variables is close to the population distribution. This theoretical framework has recently been used by Grafström, Zhao, Nylander, and Petersson (2017) as a basis for the development of a new design for the national forest inventory of Sweden. The Swedish National Forest Inventory consists of two samples, a permanent and a temporary sample. Both samples are based on clusters of circular field plots. The permanent sample will continue to serve as a permanent sample (at least in the near future), but the temporary sample has been redesigned to be more representative of a set of auxiliary variables derived from remote sensing via airborne laser scanning and from a digital elevation model. The results presented by Grafström et al. (2017) indicate a significant improvement over traditional designs. The increased representativity of the selected samples induced up to 95% reduction of variance of estimators for the included auxiliary variables.

Balanced and spatially balanced samples improve the quality of the selected samples for the chosen auxiliary variables. If a target variable is independent of the auxiliary variables, then there will be no improvement for that target variable. The degree of improvement for a specific target variable will depend on the strength of the relationship between the auxiliary variables and the specific target variable. Only auxiliary variables that are believed to have significant explanatory power for at least one of the main target variables should be included in the design. Adding extra auxiliary variables, unrelated to the target variables, may reduce the spread or balance in the other included auxiliary variables, leading to a less efficient approach.

We note that further investigation is needed to assess the effect of using the different density functions for target and auxiliary variables. We would also like to stress the need for caution before applying the approximate variance estimators, as they may be severely negatively biased for small sample sizes or target variables with very skewed distributions.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the Barro Colorado Island forest dynamics research project, which was founded by S. P. Hubbell and R. B. Foster and is now managed by R. Condit, S. Lao, and R. Perez under the Center for Tropical Forest Science and the Smithsonian Tropical Research in Panama. Numerous organisations have provided funding, principally the U.S. National Science Foundation, and hundreds of field workers have contributed. The authors are also very grateful to the associate editor and two anonymous reviewers for their many useful comments.

ORCID

A. Grafström http://orcid.org/0000-0002-4345-4024
S. Saarela http://orcid.org/0000-0002-9044-7249

REFERENCES

Benedetti, R., Piersimoni, F., & Postiglione, P. (2015). *Sampling spatial units for agricultural surveys*. Berlin: Springer.

Breidt, J. F., & Chauvet, G. (2011). Improved variance estimation for balanced samples drawn via the cube method. *Journal of Statistical Planning and Inference*, *141*(1), 479–487.

Chave, J., Coomes, D., Jansen, S., Lewis, S. L., Swenson, N. G., & Zanne, A. E. (2009). Towards a worldwide wood economics spectrum. *Ecology Letters*, *12*, 351–66.

Chave, J., Rejou-Mechain, M., Burquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B., … Vieilledent, G. (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology*, *20*(10), 3177–3190.

Condit, R. (1998). *Tropical forest census plots: Methods and results from Barro Colorado Island, Panama and a comparison with other plots*. Berlin, Heidelberg, Georgetown: Springer-Verlag, R. G. Landes Company.

Condit, R., Lao, S., Pérez, R., Dolins, S. B., Foster, R. B., & Hubbell, S. P. (2012). Barro Colorado forest census plot data, 2012 Version. http://dx.doi.org/10.5479/data.bci.20130603

Cordy, C. B. (1993). An extension of the Horvitz-Thompson theorem to point sampling from a continuous population. *Statistics & Probability Letters*, *18*(5), 353–362.

DeHardt, J. (1971). Generalizations of the Glivenko-Cantelli theorem. *The Annals of Mathematical Statistics*, *42*(6), 2050–2055.

Deville, J.-C., & Tillé, Y. (2004). Efficient balanced sampling: The cube method. *Biometrika*, *91*(4), 893–912.

Deville, J.-C., & Tillé, Y. (2005). Variance approximation under balanced sampling. *Journal of Statistical Planning and Inference*, *128*(2), 569–591.

de Vries, P. G. (1986). *Sampling theory for forest inventory*. Berlin: Springer-Verlag.

Fehrmann, L., Gregoire, T. G., & Kleinn, C. (2012). Triangulation based inclusion probabilities: A design-unbiased sampling approach. *Environmental and Ecological Statistics*, *19*(1), 107–123.

Grafström, A., & Lisic, J. (2016). BalancedSampling: Balanced and spatially balanced sampling. R package version 1.5.1. Retrieved from https://CRAN.R-project.org/package=BalancedSampling

Grafström, A., & Lundström, N. L. P. (2013). Why well spread probability samples are balanced. *Open Journal of Statistics*, *3*(1), 36–41.

Grafström, A., Lundström, N. L. P., & Schelin, L. (2012). Spatially balanced sampling through the pivotal method. *Biometrics*, *68*(2), 514–20.

Grafström, A., & Ringvall, A. H. (2013). Improving forest field inventories by using remote sensing data in novel sampling designs. *Canadian Journal of Forest Research*, *43*, 1015–1022.

Grafström, A., Saarela, S., & Ene, L. T. (2014). Efficient sampling strategies for forest inventories by spreading the sample in auxiliary space. *Canadian Journal of Forest Research*, *44*, 1156–1164.

Grafström, A., & Schelin, L. (2014). How to select representative samples. *Scandinavian Journal of Statistics*, *41*, 277–290.

Grafström, A., Zhao, X., Nylander, M., & Petersson, H. (2017). A new sampling strategy for forest inventories applied to the temporary clusters of the Swedish national forest inventory. *Canadian Journal of Forest Research*, *47*, 1161–1167. https://doi.org/10.1139/cjfr-2017-0095

Gregoire, T. G., & Valentine, H. T. (2008). *Sampling strategies for natural resources and the environment*. Applied Environmental Statistics. Boca Raton: Chapman & Hall/CRC.

Grosenbaugh, L. (1952). Plotless timber estimates—New, fast, easy. *Journal of Forestry*, *50*(1), 32–37.

Grosenbaugh, L., & Stover, W. (1957). Point-sampling compared with plot-sampling in southeast Texas. *Forest Science*, *3*(1), 2–14.

Hubbell, S. P., Foster, R. B., O'Brien, S. T., Harms, K. E., Condit, R., Wechsler, B., … Loo de Lao, S. (1999). Light-gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. *Science*, *283*, 554–557.

Mandallaz, D. (1991). *A unified approach to sampling theory for forest inventory based on infinite population and superpopulation models* (PhD thesis). Zurich: Swiss Federal Institute of Technology (ETH). https://doi.org/10.3929/ethz-a-000585900

Mandallaz, D. (2007). *Sampling techniques for forest inventories*. Boca Raton: CRC Press.

Roesch Jr, F. A., Green, E. J., & Scott, C. T. (1993). An alternative view of forest sampling. *Survey Methodology*, *19*(2), 199–204.

Schreuder, H. T., Gregoire, T. G., & Wood, G. B. (1993). *Sampling methods for multiresource forest inventory*. New York: John Wiley & Sons.

Stevens, D. L., & Olsen, A. R. (2003). Variance estimation for spatially balanced samples of environmental resources. *Environmetrics*, *14*(6), 593–610.

Stevens, D. L., & Olsen, A. R. (2004). Spatially balanced sampling of natural resources. *Journal of the American Statistical Association*, *99*, 262–278.

Wolfowitz, J. (1954). Generalization of the theorem of Glivenko-Cantelli. *The Annals of Mathematical Statistics*, *25*(1), 131–138.

Zanne, A. E., Lopez-Gonzalez, G., Coomes, D. A., Ilic, J., Jansen, S., Lewis, S. L., … Chave, J. (2009). Data from: Towards a worldwide wood economics spectrum. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.234

How to cite this article: Grafström A, Schnell S, Saarela S, Hubbell SP, Condit R. The continuous population approach to forest inventories and use of information in the design. *Environmetrics*. 2017;e2480. https://doi.org/10.1002/env.2480