Spatial Statistics 35 (2020) 100404

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Design-based consistency of the Horvitz–Thompson estimator under spatial sampling with applications to environmental surveys

Lorenzo Fattorini [a], Marzia Marcheselli [a,*], Caterina Pisani [a], Luca Pratelli [b]

[a] Department of Economics and Statistics, University of Siena, Siena, Italy
[b] Naval Academy, Livorno, Italy

ARTICLE INFO

Article history:
Received 21 April 2019
Received in revised form 19 December 2019
Accepted 20 December 2019
Available online 27 December 2019

Keywords:
Continuous populations
Finite populations
Horvitz–Thompson estimator
Population totals

ABSTRACT

Spatial populations are usually located on a continuous support. They can be surfaces representing the values of the survey variable at any location, finite collections of units with the corresponding values of the survey variable, or finite collections of areal units partitioning the support, where the survey variable is the total amount of an attribute within. We derive conditions on the design sequence ensuring consistency of the Horvitz–Thompson estimator of spatial population totals, supposing minimal requirements on the survey variable. A simulation study is performed to check theoretical results. Consistency and its implications in real surveys are discussed with focus on environmental surveys.

1. Introduction

Consistency is an intuitively appealing property ensuring that the distribution of an estimator tends to be concentrated around the parameter as the sample size n increases. Its definition cannot

* Correspondence to: Department of Economics and Statistics, University of Siena, P.zza S. Francesco 8, 53100 Siena, Italy.
E-mail addresses: lorenzo.fattorini@unisi.it (L. Fattorini), marzia.marcheselli@unisi.it (M. Marcheselli), caterina.pisani@unisi.it (C. Pisani), luca_pratelli@marina.difesa.it (L. Pratelli).

https://doi.org/10.1016/j.spasta.2019.100404
2211-6753/

be immediately carried over to the finite population setting, requiring a more complex machinery (e.g. Särndal et al., 1992, section 5.3). Indeed, when a without replacement sampling scheme is adopted to select samples from a population of N units, we cannot let n approach infinity without further, artificial assumptions, because $n \leq N$ and N is fixed and finite. Therefore consistency results can be achieved supposing a sequence of increasing populations, so that both N and n can increase indefinitely. By considering an infinite sequence of units on which the survey variable Y can be recorded, a sequence, indexed by a natural number $k = 1, 2, \ldots$, of increasing, nested populations $\{\mathcal{U}_k\}$ is constructed, each of them consisting of the first N_k units. If a corresponding sequence $\{d_k\}$ of probabilistic sampling designs selecting samples of increasing size n_k is introduced, an estimator $\widehat{\theta}_k$ of the parameter θ_k is said to be design-consistent or p-consistent if the sequence of random variables $\{\widehat{\theta}_k - \theta_k\}$ converges in probability to 0. Practically, consistency implies that an estimator has little or no bias and a small standard error as long as the sample size is large.

This asymptotic framework was probably first delineated rigorously by Isaki and Fuller (1982) who, in the spirit of design-based inference, proved the consistency of the Horvitz–Thompson (HT) estimator of population means mainly on the basis of the properties of the design sequence $\{d_k\}$, requiring a minimal assumption regarding populations. Indeed, besides conditions on the design sequence, consistency only requires that the survey variable is bounded, a feature always satisfied in real surveys, where infinite values never occur. Other asymptotic results on the HT estimator available in literature (see e.g. Prášková and Sen, 2009 for a review) refer to particular class of sampling design, such as those achieved by Rosén (1997) under order sampling scheme and by Berger (1998) for asymptotically maximum entropy design.

Following the approach by Isaki and Fuller (1982), we aim to give consistency conditions for HT total estimators in spatial populations.

In design-based inference spatial populations are constituted by fixed sets of locations within a region of interest, henceforth referred to as the support, with fixed values of the survey variable attached to each location. Specifically, three types of spatial populations can be distinguished:
(i) continuous populations, constituted by a continuous set of locations with a surface giving the values of the survey variable at each location;
(ii) finite populations of units regularly or irregularly scattered over the support (e.g. the nodes of a regular network, the factories in a district, the shrubs in a natural reserve) with the value of the survey variable attached to each unit;
(iii) finite populations of areal units partitioning the support where populations of type (i) or (ii) are located (e.g. regular polygons such as satellite pixels or irregularly-shaped administrative districts) and the survey variable is the total amount of an attribute within each areal unit.
As Wang et al. (2013) point out, while the support is the region of interest for any spatial population, the nature of sampling unit varies. Indeed, the population is the continuous set of locations coinciding with the support in case (i), a discrete set of units at fixed locations on the support in case (ii) and a set of spatial subsets partitioning the support in case (iii).

Design-based inference has been widely adopted for estimating totals of spatial populations, especially in environmental studies (e.g. De Vries, 1986; Thompson, 2002; Gregoire and Valentine, 2008; De Gruijter et al., 2010). For discrete populations of type (ii) and (iii), totals are summands and estimation is performed by using the HT estimator or suitable modifications able to exploit auxiliary information such as regression or ratio estimators (e.g. Särndal et al., 1992, chapter 6). For continuous populations of type (i), totals are integrals and estimation is performed extending the HT estimator to the continuous case (e.g. Cordy, 1993). Modifications able to exploit auxiliary information are also available for the continuous case (e.g. Brus, 2000; Barabesi and Marcheselli, 2005).

Regarding the sampling schemes usually adopted in spatial surveys, an important feature is their capacity to evenly spread sample locations over the support in such a way that no one portion of the support is over- or under-represented. This property is usually referred to as spatial balance (e.g. Brown et al., 2015). Spatial balance plays an important role in spatial surveys. This is because of the two major characteristics of spatial populations, i.e. (a) the presence of the so-called spatial autocorrelation, in which the values of the survey variable in neighboring locations tend to be similar; (b) the presence of spatial strata heterogeneity where there is high variation of the survey variable between different portions of the support and low variability within. In both cases sampling schemes likely to select clustered locations should be avoided. In the first case, neighboring locations, having similar values, provide poor additional information to sample; in the second case clustering in certain parts of the support with voids in other parts provides poorly representative samples. Spatial balance, evenly spreading sample locations over the support, is able to evenly spread sample locations throughout the strata at the same time avoiding the selection of neigbouring sites. Therefore it is able to handle both the issues. Indeed, exploiting the anticipated variance criterion, Grafström and Tillé (2013) prove the optimality of spatially balanced schemes in presence of positive spatial autocorrelation, while on the basis of the same criterion, Bethel (1989) proves the optimality of stratified sampling in the case of spatial heterogeneity.

In this paper, in the spirit of Isaki and Fuller (1982), whose consistency results on the HT estimator can be applied only to spatial population of type (*ii*), we derive consistency conditions for all the three types of spatial populations under minimal, as far as possible, conditions, also proving that they hold for the most widely applied spatial sampling schemes. In some situations, the spatial population can be held fixed, with fixed total T and consistency is proven by means of the convergence in probability to T of the sequence of HT estimators $\{\widehat{T}_k\}$ arising from the sequence of designs. In other cases, a sequence of populations is presumed giving rise to the corresponding sequence of totals $\{T_k\}$ which increases indefinitely. In these cases, consistency is obtained by the convergence in probability to 1 of the sequence $\{\widehat{T_k/T_k}\}$.

Sections 2–4 are devoted to consistency conditions while in Section 5 a simulation study is presented. Practical implications of consistency in real surveys are outlined in Section 6, with focus on environmental surveys. Finally conclusions are given in Section 7. Technical details and proofs are reported in the Appendices.

Throughout the paper, λ denotes the Lebesgue measure on $\mathbb{R}^2$ and $\mathcal{A}$ represents a compact support of $\mathbb{R}^2$.

2. Consistency for continuous populations

2.1. Setting and consistency conditions

Let y be a Borelian and bounded function on the support $\mathcal{A}$ with values on $[0, L]$, where $y(p)$ is the value of the survey variable Y at the location $p \in \mathcal{A}$. We aim to estimate the population total $T = \int_{\mathcal{A}} y(p)\lambda(dp)$.

Suppose a sequence of designs $\{d_k\}$, each of them selecting an increasing number n_k of points onto $\mathcal{A}$, say $P_{k,1}, \ldots, P_{k,n_k}$. Following Cordy (1993), the designs should be such that the n_k-tuple $(P_{k,1}, \ldots, P_{k,n_k})$ is a random vector with probability density $g^{(k)}$ with respect to the product measure $\lambda^{\otimes n_k} = \lambda \times \cdots \times \lambda$ (n_k times). Let $g_i^{(k)}$ be a version of the marginal probability density of $P_{k,i}$ with respect to λ and $g_{ih}^{(k)}$ be a version of the marginal probability density of $(P_{k,i}, P_{k,h})$ with respect to $\lambda \otimes \lambda$, with $i \neq h = 1, \ldots, n_k$.

Moreover $\pi_k(p) = \sum_{i=1}^{n_k} g_i^{(k)}(p)$ is the inclusion function and $\pi_k(p, q) = \sum_{i\neq h=1}^{n_k} g_{ih}^{(k)}(p, q)$ is the pairwise inclusion function. If $\pi_k(p) > 0$ for each $p \in \mathcal{A}$, then the extension of the HT estimator to the continuous case

$$\widehat{T}_k = \sum_{i=1}^{n_k} \frac{y(P_{k,i})}{\pi_k(P_{k,i})} \tag{1}$$

is an unbiased estimator of T and, if $\int_{\mathcal{A}} \frac{1}{\pi_k(p)}\lambda(dp) < \infty$,

$$\operatorname{var}(\widehat{T}_k) = \int_{\mathcal{A}} \frac{y^2(p)}{\pi_k(p)}\lambda(dp) + \int_{\mathcal{A}^2} \left\{ \frac{\pi_k(p, q)}{\pi_k(p)\pi_k(q)} - 1 \right\} y(p)y(q)\lambda(dp)\lambda(dq). \tag{2}$$

Proposition 1. *If the design sequence is such that*

$$\lim_{k\to\infty} \sup_{p} \frac{1}{\pi_k(p)} = 0 \tag{3}$$

$$\lim_{k\to\infty} \sup_{p\neq q} \left\{ \frac{\pi_k(p,q)}{\pi_k(p)\pi_k(q)} - 1 \right\}^{+} = 0 \tag{4}$$

then $\lim_{k\to\infty} \operatorname{var}(\widehat{T}_k) = 0$ *and* $\widehat{T}_k$ *converges in probability to* T.

2.2. Some schemes ensuring consistency

The most straightforward scheme to sample spatial locations on a continuum is uniform random sampling (URS), i.e. the random and independent selection of n_k points on the support. Under URS, $\pi_k(p) = n_k/\lambda(\mathcal{A})$ for each $p \in \mathcal{A}$ and, owing to the independence of selections, $\pi_k(p,q) = \pi_k(p)\pi_k(q)$ for each $p, q \in \mathcal{A}$, from which conditions (3) and (4) are satisfied.

Despite simplicity and consistency, URS may lead to uneven surveying. Spatial balance can be achieved using quite complex schemes explicitly tailored, such as generalized random tessellation stratified sampling (Stevens and Olsen, 2004) and sampling based on space-filling Hilbert curves (Lister and Scott, 2009).

More simply, spatial balance can be obtained using tessellation stratified sampling (TSS): the support $\mathcal{A}$ is partitioned into n_k spatial subsets of equal extent and a point is randomly and independently located in each subset. Under TSS, $\pi_k(p) = n_k/\lambda(\mathcal{A})$ for each $p \in \mathcal{A}$ while, owing to the independence of selections performed within different subsets, $\pi_k(p,q) = \pi_k(p)\pi_k(q)$ for p and q belonging to two different subsets and $\pi_k(p,q) = 0$ when p and q are in the same subset. Therefore, (3) and (4) are satisfied.

Alternatively, when the support can be tessellated into n_k regular polygons of equal extent, systematic grid sampling (SGS) is widely used to achieve spatial balance. SGS consists of randomly selecting a point in one polygon and systematically repeating it in the others. However, SGS cannot be considered for consistency in the framework introduced by Cordy (1993), because while $\pi_k(p) = n_k/\lambda(\mathcal{A})$ for each $p \in \mathcal{A}$, no probability density exists for the vector $(P_{k,i}, P_{k,i+1})$.

Under URS, TSS and SGS, (1) reduces to the Monte Carlo estimator

$$\widehat{T}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \lambda(\mathcal{A}) y(P_{k,i}). \tag{5}$$

Therefore, these schemes can be viewed as Monte Carlo integration methods with URS coinciding with crude Monte Carlo integration. Consistency results on Monte Carlo integration have been already exploited. In particular, when estimating totals and functions of totals, the relative efficiency of TSS with respect to URS has been proven for finite samples, also proving that it approaches infinity as the sample size increases because TSS variances go to 0 more quickly than under URS (Barabesi and Franceschi, 2011; Barabesi et al., 2012, 2015). Consistency under SGS, which cannot be obtained in the framework by Cordy (1993), is proven in the Monte Carlo estimation framework (see Appendix C). However, even if consistent, losses of efficiency with respect of URS may occur under SGS if spatial periodicities are present in the surface.

3. Consistency for finite populations of units

3.1. Setting and consistency conditions

Let $\{\mathcal{U}_k\}$ be a nested sequence of populations of units of increasing size N_k scattered throughout $\mathcal{A}$. Moreover, let Y be a survey variable with values on $[0, L]$ and let y_j be the value of Y on unit

$j \in \mathcal{U}_k$, usually referred to as the mark of unit j. Obviously the population sequence determines a corresponding sequence of totals $\{T_k\}$. We aim to estimate the population total $T_k = \sum_{j\in\mathcal{U}_k} y_j$.

Suppose a sequence of designs $\{d_k\}$, each of them selecting a sample $\mathcal{S}_k$ from $\mathcal{U}_k$ of increasing size n_k, with first and second order inclusion probabilities $\pi_j^{(k)}$ and $\pi_{jh}^{(k)}$ for $h > j \in \mathcal{U}_k$. Then the HT estimator

$$\widehat{T}_k = \sum_{j\in\mathcal{S}_k} \frac{y_j}{\pi_j^{(k)}} \tag{6}$$

is unbiased with variance

$$\mathrm{var}(\widehat{T}_k) = \sum_{j\in\mathcal{U}_k} \Big(\frac{1-\pi_j^{(k)}}{\pi_j^{(k)}}\Big) y_j^2 + 2 \sum_{h>j\in\mathcal{U}_k} \Big(\frac{\pi_{jh}^{(k)}}{\pi_j^{(k)}\pi_h^{(k)}} - 1\Big) y_j y_h. \tag{7}$$

Proposition 2. *If the design sequence ensures*

$$\lim_{k\to\infty} \max_{h>j} \Big\{\frac{\pi_{jh}^{(k)}}{\pi_j^{(k)}\pi_h^{(k)}} - 1\Big\}^+ = 0 \tag{8}$$

and there exists $\pi_0 > 0$ such that $\min_j \pi_j^{(k)} \geq \pi_0$, then $\lim_{k\to\infty} \mathrm{var}(\widehat{T}_k/T_k) = 0$ and $\widehat{T}_k/T_k$ converges in probability to 1.

3.2. Some schemes ensuring consistency

If the population list is available, the most straightforward scheme is simple random sampling without replacement (SRSWOR). If a constant fraction $0 < \pi_0 < 1$ of units is selected from each population $\mathcal{U}_k$, then $\pi_j^{(k)} = \pi_0$ and $\pi_{jh}^{(k)} = (\pi_0^2 N_k - \pi_0)/(N_k - 1)$ for each $h > j \in \mathcal{U}_k$, from which condition (8) is satisfied.

Despite simplicity and consistency, SRSWOR may lead to uneven scattering of sampled units throughout the region. Spatial balance can be ensured by using explicitly tailored schemes, such as generalized random tessellation stratified sampling (Stevens and Olsen, 2004), the draw-by-draw sampling that excludes the selection of contiguous units (Fattorini, 2006, 2009), the local pivotal method of the first and second types (Grafström et al., 2012), the spatially correlated Poisson sampling (Grafström, 2012) and the doubly balanced spatial sampling (Grafström and Tillé, 2013). More simply, spatial balance can be achieved by partitioning the support into a fixed number of strata and then selecting the same fraction of units $0 < \pi_0 < 1$ within each stratum by SRSWOR ensuring consistency within each stratum when strata and sample sizes increase. Then consistency also holds under stratified sampling with proportional allocations (STRPA). Moreover, the relative efficiency of STRPA vs SRSWOR arises from relation (9.18) by Hedayat and Sinha (1991) for sufficiently large strata sizes.

The knowledge of the list of the population units rarely occurs in environmental surveys where, for example, units are trees or shrubs scattered over the study area and the knowledge of the list involves prohibitive efforts, especially over large areas. The unique case in which the list of units becomes available is under 3P sampling, from the acronym of *probability proportional to prediction*. This scheme is a variation of Poisson sampling and can be adopted in forest surveys when supports are of moderate sizes. Under 3P sampling, all the units are visited by a crew of experts, a prediction x_j for the value of the survey variable is given by the experts for each unit and units are independently included in the sample with probability x_j/L (Gregoire and Valentine, 2008). A lower bound l for the survey variable Y often naturally arises in most forest and environmental surveys in which units with Y values (e.g. tree height or basal area) smaller than a given threshold are not considered in the population. In this case $\pi_j^{(k)} \geq l/L$ for any $j \in \mathcal{U}_k$ and any k and condition (8) holds as $\pi_{jh}^{(k)} = \pi_j^{(k)}\pi_h^{(k)}$ owing to the independence of drawings. Therefore, under 3P sampling, $\hat{T}_k/T_k$ converges in probability to 1.

3.3. Consistency for without-list populations

When the list of units is unknown, the sampling schemes usually adopted for selecting units from the population are based on points (eventually identifying plots or transects) randomly located onto a reference area $\mathcal{B}$. In most cases $\mathcal{B}$ constitutes an enlargement of the support $\mathcal{A}$ introduced in order to avoid edge effects (see e.g. Gregoire and Valentine, 2008, section 7.5). If $\mathcal{U}$ denotes the without-list population of units, for each unit $j \in \mathcal{U}$, the scheme univocally defines the inclusion region $\mathcal{B}_j$, i.e. the subset of $\mathcal{B}$ onto which the random point p must fall to give rise to the selection of the unit. Because p is randomly selected, the first-order inclusion probability of unit j is given by $\pi_j = \lambda(\mathcal{B}_j)/\lambda(\mathcal{B})$.

Denote by $\mathcal{S}(p) \subset \mathcal{U}$ the sample of units selected by means of the random point p onto $\mathcal{B}$. If $\lambda(\mathcal{B}_j)$ can be computed for each $j \in \mathcal{S}(p)$, then the HT estimator

$$\widehat{T} = \sum_{j\in\mathcal{S}(p)} \frac{y_j}{\pi_j} = \lambda(\mathcal{B}) \sum_{j\in\mathcal{S}(p)} \frac{y_j}{\lambda(\mathcal{B}_j)} \tag{9}$$

is an unbiased estimator of the population total $T = \sum_{j\in\mathcal{U}} y_j$. Since T can be rewritten as

$$T = \int_{\mathcal{B}} g(p)\lambda(dp) \tag{10}$$

where $g(p) = \sum_{j\in\mathcal{U}} \frac{y_j}{\lambda(\mathcal{B}_j)} I_j(p)$ and $I_j(p)$ is the sample indicator function, that is equal to 1 if $j \in \mathcal{S}(p)$ and 0 otherwise (see Appendix B for the proof), (9) can be rewritten as

$$\widehat{T} = \lambda(\mathcal{B})g(p) \tag{11}$$

i.e. it can be viewed as the Monte Carlo estimator of the integral (10) at p (Gregoire and Valentine, 2008, chapter 10; Mandallaz, 2008). Practically speaking, when dealing with a without-list population, the total estimation can be rephrased by (5) as the estimation of a total over a continuum, in such a way that consistency can be achieved, as in Section 2, as the number of sample points selected onto $\mathcal{B}$ increases. Obviously, analogous considerations to those in Section 2.2 regarding URS, TSS and SGS hold.

4. Consistency for finite populations of areal units

4.1. Setting and consistency conditions

Both when a continuous population and a finite population of units are defined on $\mathcal{A}$, a sequence of partitions $\{\mathcal{P}_k\}$ of $\mathcal{A}$ can be considered, constituted by an increasing number M_k of areal units $\mathcal{A}_1^{(k)}, \ldots, \mathcal{A}_{M_k}^{(k)}$ of decreasing size, in such a way that $\lim_{k\to\infty} \max_l \lambda(\mathcal{A}_l^{(k)}) = 0$. Suppose a sequence of designs $\{d_k\}$, each of them selecting a sample $\mathcal{Q}_k$ of increasing size m_k from $\mathcal{P}_k$ in accordance with a sampling scheme that induces first and second order inclusion probabilities $\pi_l^{(k)}$ and $\pi_{lh}^{(k)}$ for $h > l \in \mathcal{P}_k$.

If a continuous population is considered and y is a Borelian function on $\mathcal{A}$, with values on $[0, L]$, let $T = \int_{\mathcal{A}} y(p)\lambda(dp)$ be the population total and let $T_l^{(k)} = \int_{\mathcal{A}_l^{(k)}} y(p)\lambda(dp)$ be the total amount of the survey variable within $\mathcal{A}_l^{(k)}$. Since $T = \sum_{l=1}^{M_k} T_l^{(k)}$ for each k, estimating T over a continuum can be approached as the total estimation in a finite population of areal units, whose list is always available. The HT estimator

$$\widehat{T}_k = \sum_{l\in\mathcal{Q}_k} \frac{T_l^{(k)}}{\pi_l^{(k)}} \tag{12}$$

is unbiased for T with variance

$$\mathrm{var}(\widehat{T}_k) = \sum_{l\in\mathcal{P}_k} \left(\frac{1-\pi_l^{(k)}}{\pi_l^{(k)}}\right) T_l^{(k)2} + 2 \sum_{h>l\in\mathcal{P}_k} \left(\frac{\pi_{lh}^{(k)}}{\pi_l^{(k)}\pi_h^{(k)}} - 1\right) T_l^{(k)} T_h^{(k)}. \tag{13}$$

Proposition 3. *For continuous populations, if the design sequence is such that*

$$\lim_{k\to\infty} \max_{l} \frac{\lambda(\mathcal{A}_l^{(k)})}{\pi_l^{(k)}} = 0 \tag{14}$$

and

$$\lim_{k\to\infty} \max_{h>l} \left(\frac{\pi_{lh}^{(k)}}{\pi_l^{(k)}\pi_h^{(k)}} - 1 \right)^{+} = 0 \tag{15}$$

then $\lim_{k\to\infty} \operatorname{var}(\widehat{T}_k) = 0$ *and* $\widehat{T}_k$ *converges in probability to* T.

A sufficient condition for (14) to hold is $\min_l \pi_l^{(k)} \geq \pi_0 > 0$ for any k.

Alternatively, if we have a finite population of units scattered throughout $\mathcal{A}$, consistency cannot be proven taking the population fixed. In this case, it is necessary to consider a sequence of nested populations $\mathcal{U}_k$ of increasing size N_k. Let Y be a survey variable with values on $[0, L]$ and y_j be the value of Y on unit $j \in \mathcal{U}_k$, so that the total for kth population is $T_k = \sum_{j\in\mathcal{U}_k} y_j$. Once again estimation in finite populations of units can be switched into estimation in finite populations of areal units whose lists are always available.

Let $T_l^{(k)} = \sum_{j\in\mathcal{U}_l^{(k)}} y_j$ be the total amount of the survey variable within the areal unit $\mathcal{A}_l^{(k)}$, where $\mathcal{U}_l^{(k)}$ denotes the sub-population of the $N_l^{(k)}$ units located within $\mathcal{A}_l^{(k)}$. As $T_k = \sum_{l\in\mathcal{P}_k} T_l^{(k)}$, the HT estimator $\widehat{T}_k$ given by (12) is unbiased with variance (13).

Proposition 4. *For finite populations of units, if (15) holds and*

$$\lim_{k\to\infty} \max_{l} \frac{T_l^{(k)}}{\pi_l^{(k)} T_k} = 0 \tag{16}$$

then $\lim_k \operatorname{var}(\widehat{T}_k/T_k) = 0$ *and* $\widehat{T}_k/T_k$ *converges in probability to 1.*

There is a perfect analogy with the continuous case. In the case of finite populations of units, it suffices to replace $\lambda(\mathcal{A}_l^{(k)})$ with $T_l^{(k)}/T_k$. However, while in the continuous case only partitions are involved, in the discrete case it is necessary to presume a sort of evenness in the enlargements of the nested populations, in such a way that population units do not aggregate into some areal units.

4.2. Some schemes ensuring consistency

The most straightforward scheme to select areal units is SRSWOR. If a constant fraction $0 < \pi_0 < 1$ of areal units is selected from the partition $\mathcal{P}_k$, then $\pi_l^{(k)} = \pi_0$ and $\pi_{lh}^{(k)} = (\pi_0^2 M_k - \pi_0)/(M_k - 1)$ for each $h > l \in \mathcal{P}_k$. Hence conditions (14) and (15) are satisfied. In the discrete case, condition (16) requires that the amounts of the survey variable within areal units divided by their total approach 0.

SRSWOR may lead to uneven scattering of sampled areas. Spatial balance can be achieved using the quite complex schemes referred in Section 3.2 to sample units scattered over a continuum, assigning to each areal unit a location (e.g. its barycenter). More simply, spatial balance can be achieved using the one per stratum sampling (OPSS) (Breidt, 1995), i.e. partitioning $\mathcal{P}_k$ into m_k strata of approximately M_k/m_k contiguous units and randomly selecting one unit within each stratum. Under OPSS, $\pi_l^{(k)} = \pi_0$ for each $l \in \mathcal{P}_k$, $\pi_{lh}^{(k)} = \pi_0^2$ if the areal units l and h belong to different strata while $\pi_{lh}^{(k)} = 0$ if they belong to the same stratum, so that conditions (14) and (15) are satisfied. Once again, in the discrete case, convergence to 0 of the amounts of the survey variable within areal units divided by their total is required. Also for areal units, the relative efficiency of OPSS vs SRSWOR arises, *mutatis mutandis*, from relation (9.18) by Hedayat and Sinha (1991) for sufficiently large strata sizes.

When $\mathcal{A}$ is regular and can be partitioned into M_k regular polygons of equal extent, systematic sampling (SYS) can be performed, i.e. partitioning $\mathcal{P}_k$ into m_k equally-shaped strata of M_k/m_k

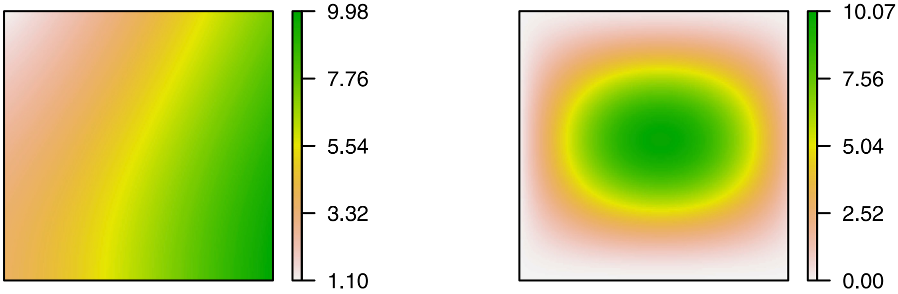

Fig. 1. Maps of the two artificial surfaces.

contiguous polygons, randomly selecting one unit in a stratum and repeating it in the remaining strata. However, under SYS $\pi_l^{(k)} = \pi_0$ for each $l \in \mathcal{P}_k$, while $\pi_{lh}^{(k)} = \pi_0$ if polygons l and h have the same position in the strata and $\pi_{lh}^{(k)} = 0$ otherwise. Therefore, (15) is not satisfied. Notwithstanding this, we have proven the consistency of $\widehat{T}_k$ under SYS in the continuous case (see Appendix D). On the other hand, in the case of finite populations of units, consistency under SYS cannot be proven. Irrespective of consistency, SYS performance greatly depends on the population characteristics (e.g. Cochran, 1977, section 8) and large losses of precision with respect to SRSWOR may occur when spatial periodicities are present.

5. Simulation study

In order to generate continuous populations, finite populations of points and populations of areal units, two artificial surfaces on the unit square, referred to as surface 1 and surface 2 were considered. At any point $p = (p_1, p_2)$ of the unit square the two surfaces were defined respectively by

$$y(p) = C_1(\sin^2 p_1 + \cos^2 p_2 + p_1), \quad y(p) = C_2(\sin 3p_1 \sin^2 3p_2)^2,$$

where the constants C_1, C_2 ensured a maximum value of 10 in both cases (see Fig. 1).

The two surfaces were chosen to represent the major characteristics of spatial populations, as mentioned in the Introduction. Both of them are continuous, hence the values of the survey variable in neighboring locations tend to be similar, thus entailing a spatial autocorrelation in the resulting populations. Both of them varied relevantly throughout the unit square: surface 1 showed an increasing trend moving from the left part to the right part of the square, while the second showed an increasing trend toward the center of the square, thus entailing a spatial stratification with different values of the survey variable in different parts of the square.

5.1. Continuous populations

Surface 1 and 2 gave rise to population totals of 5.54 and 3.50, respectively. In both cases, $n = 25, 100, 400$ points were selected on the unit square in accordance with URS, TSS and SGS. In particular TSS and SGS were performed by partitioning the unit square into 5×5, 10×10 and 20×20 grids of equal-sized quadrats and randomly or systematically selecting a point per quadrat. For any combination of surface, sampling scheme and sample size n, $R = 10\,000$ samples were independently selected and the Monte Carlo estimator (5) was computed.

5.2. Finite populations of points

Three nested populations of $N = 1000, 5000$ and $10\,000$ points were located in the unit square in accordance with four spatial patterns referred to as regular, random, trended and clustered.

For the regular pattern, the nested populations were constructed generating the first 1000 points completely at random on the unit square but discarding those having distances smaller than $1/\sqrt{1000}$ to those previously generated, then adding further 4000 points completely at random on the unit square but discarding those having distances smaller than $1/\sqrt{5000}$ to those previously generated, and finally adding further 5000 points completely at random discarding those having distances smaller than $1/\sqrt{10000}$.

For the random pattern, the nested populations were constructed by simply generating 10 000 points completely at random in the unit square and then assigning the first 1000 to the first population, the first 5000 to the second population and all of them to the third.

For the trended pattern, the nested populations were constructed generating 10 000 points of coordinates $(1-u_1^2, 1-u_2^2)$ where u_1, u_2 were independent random numbers uniformly distributed on (0, 1). The first 1000 points were assigned to the first population, the first 5000 to the second population and all of them to the third.

Finally, for the clustered pattern, the nested populations were constructed generating 10 cluster centers completely at random on the unit square and then assigning 100 points to each cluster generated from a spherical normal distribution centered on the cluster center with variances 0.025, then adding further 400 points to each cluster from the same normal distribution and finally adding further 500 points to each cluster from the same distribution. Points falling outside the unit square were discarded and newly generated. For each pattern and population, marks were obtained by considering the values of surface 1 and 2 at the generated points. Totals obviously changed with surfaces, spatial patterns and population sizes and were reported in Table 2.

For any population arising from the combination of surface, spatial pattern and population size N, $R = 10\,000$ samples of size $n = N/10$ were independently selected by means of SRSWOR and STRPA and the HT estimator (6) was computed. STRPA was performed by previously partitioning the unit square into 16 spatial strata of equal size and then selecting 10% of points within each stratum by means of SRSWOR.

Subsequently, for any surface and any spatial pattern, the populations of size $N = 1000$ were sampled as if they were without list, by means of $n = 25, 100, 400$ plots of radius $\rho = 0.013$. Plot centers were selected by means of URS, TSS and SGS on the unit square enlarged by a buffer of width ρ to avoid edge effects. Plot radius was selected to ensure that the surveyed area was about 20% of the unit square area when 400 plots were considered. TSS and SGS were performed as described in Section 5.1. For each surface, each spatial pattern, each sampling scheme and each n, $R = 10\,000$ samples were independently selected and the Monte Carlo estimator (5) was computed.

5.3. *Finite populations of areal units*

Simulation for finite populations of areal units was performed both when a continuous population and a finite population of units on the unit square were considered. In both cases, three populations of $M = 400, 1600, 6400$ areal units were constructed by partitioning the unit square into grids of 20×20, 40×40, 80×80 quadrats, respectively.

In the continuous framework, population values were the integrals of the surface onto the quadrats. The population totals were 5.54 and 3.50 when surface 1 and 2 were considered, respectively. For each surface and each population size M, $R = 10\,000$ samples of size $m = M/10$ were independently selected by means of SRSWOR, OPSS and SYS and the HT estimator (12) was computed. OPSS and SYS were performed by partitioning the grids into blocks of 2×5 contiguous quadrats and then randomly or systematically selecting one quadrat per block.

Alternatively, the three populations of $M = 400, 1600, 6400$ quadrats were superimposed to the nested point populations of sizes $N = 1000, 5000, 10\,000$ generated in Section 5.2, in such a way that both N and M increased. In this case, population values were the sums of the marks attached to each point in the quadrats (population totals were the same reported in Table 2). Analogously to the continuous case, for each population of spatial units, $R = 10\,000$ samples of size $m = M/10$ were independently selected by means of SRSWOR, OPSS and SYS and the HT estimator (12) was computed.

Table 1
Relative efficiencies for any combination of sample size n, surface and sampling scheme. For any surface, the benchmark is the empirical relative standard error for $n = 25$ under URS.

n	Surface 1			Surface 2		
	URS	TSS	SGS	URS	TSS	SGS
25	1.00	5.06	1.04	1.00	2.52	8.89
100	1.98	19.97	2.08	2.00	10.07	23.35
400	3.97	82.11	4.15	3.97	40.02	50.94

Table 2
Relative efficiencies for any combination of population size N (giving rise to total T), spatial pattern, surface and sampling scheme with sample size $n = N/10$. For any surface and any spatial pattern, the benchmark is the empirical relative standard error achieved for $N = 1000$ under SRSWOR.

	N	T	Surface1		Surface 2	
			SRSWOR	STRPA	SRSWOR	STRPA
Regular	1000	5423.36	1.00	4.09	1.00	2.04
	5000	27 542.04	2.24	8.90	2.23	4.63
	10 000	55 326.07	3.18	12.71	3.17	6.51
Random	1000	5446.18	1.00	4.14	1.00	2.04
	5000	27 454.91	2.25	9.13	2.24	4.62
	10 000	55 023.73	3.18	12.71	3.15	6.62
Trended	1000	6234.43	1.00	4.00	1.00	1.99
	5000	30 979.18	2.23	8.74	2.24	4.38
	10 000	61 882.13	3.13	12.30	3.16	6.28
Clustered	1000	6020.92	1.00	8.49	1.00	2.44
	5000	30 127.31	2.25	18.56	2.26	5.46
	10 000	60 244.34	3.16	27.00	3.21	7.72

5.4. Performance indicators

The Monte Carlo distributions of the HT estimates were adopted to empirically determine the sampling variance and its behavior as the sample size, and eventually the population size, increases. As index of precision, the empirical relative standard error

$$\frac{1}{T}\sqrt{\frac{1}{R}\sum_{r=1}^{R}(\widehat{T}_r - T)^2}$$

was computed, where $\widehat{T}_r$ was the estimate achieved in the rth sample and T was the total under estimation.

Because precision greatly depends on population characteristics, in order to facilitate comparisons we took the strategies adopting the simplest sampling design, i.e. URS for continuous populations and SRSWOR for finite populations, and/or the smallest sampling effort as benchmarks. Then, we checked relative efficiency by computing the ratios of the empirical relative standard error of the benchmark design to that of the other designs. Relative efficiencies were expected to increase as sample sizes increased and as effective sampling schemes were adopted.

5.5. Results

Simulation results, reported in Tables 1–5, fully confirmed the consistency results of the previous sections. In the case of continuous populations, the benchmark was achieved with $n = 25$ points located by means of URS and the relative efficiencies for any combination of sample size n, surface and sampling scheme are reported in Table 1. Under this scheme, precision increased at a rate proportional to $\sqrt{n}$, as from the theory of Monte Carlo integration discussed in Section 2.2, while

Table 3
Relative efficiencies for without-list populations of $N = 1000$ points for any combination of sample size n, spatial pattern, surface and sampling scheme. For any surface, the benchmark is the empirical relative standard error achieved for $n = 25$ under URS.

	n	Surface 1			Surface 2		
		URS	TSS	SGS	URS	TSS	SGS
Regular	25	1.00	1.05	0.95	1.00	1.15	1.18
	100	1.99	2.12	2.18	2.00	2.36	2.73
	400	3.98	4.40	5.07	4.05	4.96	6.65
Random	25	1.00	1.02	0.95	1.00	1.10	1.13
	100	1.99	2.08	2.15	2.01	2.23	2.88
	400	3.97	4.41	5.97	4.09	4.76	5.83
Trended	25	1.00	1.16	0.77	1.00	1.04	1.02
	100	1.99	2.57	1.44	2.00	2.16	2.01
	400	3.91	6.03	2.66	3.98	4.65	4.22
Clustered	25	1.00	1.08	1.47	1.00	1.06	1.24
	100	1.99	2.42	3.64	1.96	2.42	3.35
	400	3.96	6.59	11.85	3.91	6.35	11.43

Table 4
Relative efficiencies for populations of quadrats for any combination of population size M, surface and sampling scheme. For any surface the benchmark is the empirical relative standard error achieved for $M = 20 \times 20$ under SRSWOR.

	Surface 1			Surface 2		
	SRSWOR	OPSS	SYS	SRSWOR	OPSS	SYS
20×20	1.00	7.68	1.19	1.00	2.46	17.72
40×40	2.02	31.17	2.57	1.98	9.68	40.58
80×80	4.04	112.20	5.15	3.97	38.12	83.87

higher precision and greater increase rates were achieved under TSS. On the other hand, even if results on SGS confirmed consistency, as proven in Appendix C, relative efficiencies with respect to the benchmark were small for the first surface, while they were the greatest with the second surface, resulting even better than TSS. That confirmed the strong dependence of the performance of SGS on the surface under study.

When populations of points were considered, for all the spatial patterns the benchmark was achieved with populations of $N = 1000$ points and SRSWOR and results are given in Table 2. Notwithstanding sampling fractions remained the same, precision increased as outlined in Section 3.2. The rate of increase was approximately proportional to $\sqrt{n}$ for both the surfaces, all the spatial patterns and both sampling schemes. For surface 1, the relative efficiency of STRPA with respect to SRSWOR was about 8 in the case of clustered pattern and about 4 for the other patterns. In the case of surface 2 improvements provided by stratification were less marked, relative efficiencies being about 2 for all patterns.

In the case of without-list finite populations of $N = 1000$ points sampled by means of circular plots of radius $\rho = 0.013$, for all the spatial patterns the benchmark performance was that achieved with $n = 25$ plots under URS. Results, reported in Table 3, were similar to those achieved for continuous populations even if the efficiencies of TSS and SGS with respect to URS were considerably less marked. Under URS, precision increased at a rate proportional to $\sqrt{n}$ for all the spatial patterns, while a higher rate was invariably achieved under TSS. SGS showed the greatest efficiencies and rate of increase with regular, random and clustered patterns, but its performance was the worst with trended pattern, confirming its dependence on the population characteristics. When finite populations of $M = 20 \times 20$, 40×40, 80×80 quadrats with population values being the totals of the two surfaces within quadrats were considered, for both surfaces the benchmark was achieved with the population of $M = 20 \times 20$ quadrats sampled by means of SRSWOR and results are reported in Table 4. Under this scheme, precision increased at a rate proportional to $\sqrt{n}$, while higher precision with greater increase rates was achieved under TSS. Once again, simulation results confirmed

Table 5
Relative efficiencies for populations of quadrats for any combination of point population size N, quadrat population size M, spatial pattern, surface and sampling scheme. For any combination of spatial pattern and surface, the benchmark is the empirical relative standard error achieved for $N = 1000$ and $M = 20 \times 20$ under SRSWOR.

	N	M	Surface 1			Surface 2		
			SRSWOR	OPS	SYS	SRSWOR	OPSS	SYS
Regular	1000	20 × 20	1.00	1.25	0.84	1.00	1.44	2.47
		40 × 40	1.02	1.06	1.13	1.30	1.49	1.50
		80 × 80	0.91	0.91	0.78	1.32	1.32	1.82
	5000	20 × 20	1.36	2.45	1.05	1.15	2.41	5.51
		40 × 40	2.01	2.57	1.69	2.03	3.52	4.39
		80 × 80	2.29	2.41	2.89	2.87	3.48	4.30
	10 000	20 × 20	1.47	3.61	0.98	1.19	2.69	9.22
		40 × 40	2.47	3.98	1.79	2.21	5.21	7.46
		80 × 80	3.18	3.72	2.68	3.59	5.28	6.59
Random	1000	20 × 20	1.00	1.13	1.24	1.00	1.29	2.09
		40 × 40	1.04	1.06	1.13	1.22	1.31	0.99
		80 × 80	1.05	1.06	0.88	1.31	1.34	1.45
	5000	20 × 20	1.58	2.35	1.69	1.22	2.16	4.37
		40 × 40	1.99	2.25	2.32	1.98	2.80	2.84
		80 × 80	2.24	2.38	2.89	2.65	3.07	2.69
	10 000	20 × 20	1.71	3.07	1.41	1.27	2.53	4.47
		40 × 40	2.52	3.25	2.37	2.18	3.82	5.41
		80 × 80	3.08	3.33	4.43	3.30	4.19	4.53
Trended	1000	20 × 20	1.00	1.18	0.49	1.00	1.17	1.22
		40 × 40	1.55	1.94	0.67	1.21	1.28	1.60
		80 × 80	2.15	2.49	0.93	1.25	1.29	1.03
	5000	20 × 20	1.17	1.55	0.55	1.40	2.28	2.12
		40 × 40	1.96	2.75	0.77	2.15	2.71	2.37
		80 × 80	3.20	4.23	1.04	2.60	2.88	2.11
	10 000	20 × 20	1.16	1.58	0.57	1.48	2.84	2.73
		40 × 40	2.06	2.95	0.78	2.46	3.89	3.09
		80 × 80	3.51	4.95	1.08	3.48	4.03	2.54
Clustered	1000	20 × 20	1.00	1.07	1.58	1.00	1.12	3.27
		40 × 40	1.68	2.05	1.83	1.68	2.13	2.41
		80 × 80	2.82	3.87	3.10	2.73	3.90	5.07
	5000	20 × 20	0.99	1.06	1.80	1.01	1.12	3.63
		40 × 40	1.71	2.13	2.39	1.70	2.19	2.98
		80 × 80	3.22	5.60	5.55	3.19	5.85	11.62
	10 000	20 × 20	0.98	1.04	1.78	1.01	1.14	4.49
		40 × 40	1.71	2.14	2.50	1.74	2.22	3.12
		80 × 80	3.29	5.99	6.32	3.32	6.47	16.31

consistency under SYS, as proven in Appendix D, which however provided small efficiencies with respect to the benchmark scheme for the first surface, while they were the greatest with the second surface.

Finally, finite populations of $M = 20 \times 20$, 40×40, 80×80 quadrats with population values being the totals of marks within quadrats were considered and, in particular, point populations of size $N = 1000, 5000, 10\,000$ were generated according to the four spatial patterns. Areal unit populations were sampled by means of $m = M/10$ quadrats selected using SRSWOR, OPSS and SYS. For both surfaces and each spatial pattern, the benchmark was achieved with the population of $M = 20 \times 20$ quadrats and $N = 1000$ points sampled by means of SRSWOR. As argued in Section 4, results in Table 5 showed that consistency under SRSWOR and OPSS required that both M and N jointly increased. The results also confirmed a possible lack of consistency under SYS, which usually outperformed SRSWOR and OPSS with clustered and random patterns, but had performance even worse than SRSWOR with trended pattern and surface 1. Under SRSWOR precision tended to increase more slowly than to a rate proportional to $\sqrt{n}$, while better precision with greater increase rates were achieved under OPSS, that was invariably more efficient with respect to SRSWOR.

6. Consistency and real surveys

Design-based inference is of large use in environmental surveys, especially in large-scale forest surveys such as national forest inventories (e.g. Tomppo et al., 2010). Usually the main targets are to estimate land use, especially forest cover that can be expressed as integrals of dichotomous variables, together with totals of some attributes regarding finite populations of objects within some land cover classes (e.g. total volume of trees within forested lands). Therefore populations of types (*i*) and (*ii*) are both involved and the goal is to estimate their totals by the same survey. That is done by randomly locating points onto the study region in accordance with a sampling scheme, recording the land cover class at each point and selecting samples of objects within plots of pre-fixed size centered at the selected points (Fattorini, 2014).

Consistency of the resulting estimators holds under the more widely adopted sampling schemes, such as TSS and SGS (Sections 2.2 and 3.3). Therefore, as the number of points increases, i.e. when the subset grain is sufficiently small with respect to the size of the study area, thus providing a sufficiently large number of sample points, the probability distribution of the estimators can be considered concentrated around the true parameters. These considerations support the results of some surveys such as the Italian National Forest Inventory where about 300 000 points, one per square kilometer, were selected on the Italian territory (Fattorini et al., 2006) and the IUTI survey, a land use survey promoted and carried out in 2008 in Italy, where 1 200 000 points, one each 250 square meters, were selected (Pagliarella et al., 2016).

In most large-scale forest surveys, a second-phase of sampling is performed because it may be demanding to visit and perform estimation for all the selected points (e.g. Marchetti et al., 2018). In this case, estimator (5) is only virtual and, in a second-phase, a sub-sample of points is selected using a finite population sampling scheme. Fattorini et al. (2017) give sufficient conditions for the second-phase designs to ensure consistency of the two-phase estimators when TSS is performed in the first phase.

As pointed out by Opsomer et al. (2007), in the last years there is a "tremendous" opportunity of exploit auxiliary data derived from remote sensing sources such as photo-interpreted land-cover class, elevation, slope and Lidar metrics, in order to improve the accuracy of the two-phase estimators by means of calibration strategies performed introducing assisting super-population models. Because the resulting model-assisted estimators, such as regression and ratio estimators, can be invariably expressed as smooth functions of HT estimators of totals (e.g. Särndal et al., 1992, chapter 6), if consistency holds for the HT estimators it also holds for the model-assisted counterparts.

6.1. A small-scale case study

Environmental surveys at large scale (e.g. countries and parks) are usually performed by environmental agencies with great availability of staffs and resources. That allows for sustaining the sampling of an adequate number of locations per unit area that, joined with the large extents of the areas under study, would provide very large sample sizes and consistent estimation. However, consistency should be pursued also in small-scale study, usually performed by few researchers and volunteers, by adopting suitable sampling strategies, possibly exploiting auxiliary information and taking the sampling effort as large as possible. In spring 2016 we were involved in a sample survey performed in a small stand constituted by a rectangular area of size 70×140 m^2 within a seminatural mixed oak-hornbeam flood plain forest in the province of Mantova (North Italy). Owing to the moderate size of the study area, 3P sampling was adopted to select trees, as customary in most forest surveys over small stands. A crew of experts visited the population $\mathcal{U}$ of the $N = 510$ trees lying in the stand with bole diameter at breast height greater than 7 cm, giving a prediction x_j for the height (m) of each tree $j \in \mathcal{U}$. Then, 3P sampling was performed by independently selecting each tree with probability $\pi_j = x_j/L$, where L was set equal to 151.373 m, in such a way to ensure $\pi_j \leq 1$ for each $j \in \mathcal{U}$ and such that the sum of probabilities provided an expected sample size of 50 trees, with an expected sampling fraction of about 10%. The selected sample $\mathcal{S}$ was constituted by $n = 54$ trees whose heights y_j were measured for each $j \in \mathcal{S}$. The use of 3P sampling and of the HT estimator was judged an effective strategy to estimate the average height of trees. Indeed, 3P sampling provided inclusion probabilities proportional to the predicted height. Therefore, if predictions were good, 3P sampling was also likely to provide inclusion probability approximately proportional to the true heights, that, if possible, would reduce sampling variance to 0 (e.g. Särndal et al., 1992, section 3.5). Predictions performed by the crew turned out to be very good as they explained the 99.8% of the height variability in the sample. In addition, stated the relevant sampling fraction of about 10%, the strategy was judged likely to provide an estimate near to the true average value. That was confirmed by the sample results. The HT estimate of the average height (HT estimate of total divided by 510) was of 16.03 m with an estimated sampling variance (e.g. Särndal et al., 1992, equation 3.5.6) of 4.11 m^2, providing a quite satisfactory estimate of the relative standard error of 12.6%. The quality of the predictions allowed to perform a presumably reliable assessment of both spatial autocorrelation and spatial stratified heterogeneity of the survey variable. Using the predictions, spatial autocorrelation was quantified by means of the Moran Index (Moran, 1950) in which only the nearest tree was considered as neighbor. Moreover, spatial stratified heterogeneity was assessed by partitioning the study area into four strata of size 35×70 m^2 and computing the q-index (Wang et al., 2016). The values of the two indexes were equal to 0.21 and 0.02, respectively, highlighting that both spatial autocorrelation and stratified heterogeneity were extremely weak. Thus, the use of 3P sampling which does not take into account the spatial pattern of the trees seems to be suitable.

7. Conclusions

Consistency of the design-based estimators of totals of spatial populations is pursued under minimal assumptions regarding the population characteristics, by focusing on the design sequences. Even if there is no sequence in real surveys, however the presumed sequence is inspired by the sampling scheme actually adopted to select the sample. Therefore consistency can be considered real, not modeled or assumed.

For continuous populations, it suffices to hold the support and the surface as fixed and simply considering a design sequence selecting an increasing number of sample points in the support. A slightly more complex machinery is necessary for finite populations of units scattered onto a support, where the Isaki and Fuller (1982) asymptotic scenario is exploited, taking the support fixed and considering a sequence of nested populations increasing within. However, when the scattered units have no list, as frequently happens in environmental surveys, and hence it is necessary to sample them by points, eventually identifying plots or transects, the population can be held fixed and consistency can be achieved from the scheme adopted to locate an increasing number of points on the support. Also in the case of populations of areal units, the support as well as the surface are held fixed in the continuous case and consistency is achieved from the scheme adopted to select areal units from a sequence of partitions constituted by an increasing number of areas of decreasing extents. Results are less general in the finite case, where a sequence of nested populations must be introduced and conditions on their enlargement are necessary.

Consistency is essential also for estimating more complex parameters: if consistent estimators $\widehat{T}_1, \ldots, \widehat{T}_K$ are available for the totals $T_1, \ldots, T_K$ of K attributes, then for a wide class of functions f, the plug-in estimator $f(\widehat{T}_1, \ldots, \widehat{T}_K)$ is consistent for $f(T_1, \ldots, T_K)$ (e.g. Särndal et al., 1992, remark 5.3.1).

For the three types of spatial populations, Table 6 shows the list of sampling designs that meet sufficient conditions for consistency, as proven in Sections 2.2, 3.2 and 4.2. Besides the URS and SRSWOR schemes, that constitute theoretical benchmarks, consistency holds for most schemes widely adopted in environmental surveys. All these schemes naturally achieve spatial balance by means of stratified or systematic selection of sample locations. Unfortunately, owing to their complexity in second-order inclusion probabilities, we cannot prove consistency for other complex spatially balanced schemes recently appeared in literature, that we have already mentioned in Section 3.2. However, owing to their effectiveness in providing spatial balance and their empirical performance (Fattorini et al., 2015), consistency presumably holds also for these schemes. On the other hand we have not considered consistency for schemes providing unbalanced selections, with

Table 6
List of some widely adopted spatial sampling schemes satisfying sufficient conditions for consistency.

Population type	Consistent schemes
Continuous populations or finite populations of points without list	URS TSS SGS
Finite populations of points with list	SRSWOR STRPA 3P sampling
Finite populations of areal units in the continuous case	SRSWOR OPSS SYS
Finite populations of areal units in the discrete case	SRSWOR OPSS

clustering of sample locations in some parts of the support and voids in other parts. They are rarely adopted in spatial surveys. Their use has been recently discouraged in forest studies (e.g. Pagliarella et al., 2018).

Regarding the relative performance of the schemes of Table 6, the simulation study has shown that the standard errors of the benchmark schemes (URS and SRSWOR) usually decrease with the square root of the sample size, while faster rates of decreases are invariably obtained adopting stratified schemes (TSS and OPSS). Therefore stratified allocations invariably provide efficiency with respect to random allocations, that tends to increase with the sample size. On the other hand the relative efficiencies of systematic allocations heavily depend on the characteristics of the spatial populations. These results suggest the use of stratified schemes as the most cautelative solution less linked to the unknown population features. Finally, stated this dependence on the population characteristics, it is extremely difficult to determine if the sample size adopted in the survey is sufficiently large to meet consistency, i.e. to provide an estimator with a small standard error. In this framework, the estimation of the sampling variance seems to be the sole way for a fair evaluation of the actual precision. However, as pointed out by Grafström (2012), variance estimation is a challenging issue in spatial sampling owing to the presence of some second-order inclusion probabilities equal to 0 in most spatially balanced schemes. Therefore, the construction of reliable variance estimators is a future necessary step.

Appendix A. Proofs of propositions

A.1. Proof of Proposition 1

Because the surface $y(p)$ is bounded by L, from (2)

$$\begin{aligned}
\mathrm{var}(\widehat{T}_k) &= \int_{\mathcal{A}} \frac{y^2(p)}{\pi_k(p)}\lambda(dp) + \int_{\mathcal{A}^2}\left\{\frac{\pi_k(p,q)}{\pi_k(p)\pi_k(q)} - 1\right\}y(p)y(q)\lambda(dp)\lambda(dq) \\
&\le L^2\int_{\mathcal{A}} \frac{1}{\pi_k(p)}\lambda(dp) + L^2\int_{\mathcal{A}^2}\left\{\frac{\pi_k(p,q)}{\pi_k(p)\pi_k(q)} - 1\right\}^{+}\lambda(dp)\lambda(dq) \\
&\le L^2 \sup_p \frac{1}{\pi_k(p)}\int_{\mathcal{A}}\lambda(dp) + L^2\sup_{p\neq q}\left\{\frac{\pi_k(p,q)}{\pi_k(p)\pi_k(q)} - 1\right\}^{+}\int_{\mathcal{A}^2}\lambda(dp)\lambda(dq) \\
&= L^2\lambda(\mathcal{A})\left[\sup_p\frac{1}{\pi_k(p)} + \lambda(\mathcal{A})\sup_{p\neq q}\left\{\frac{\pi_k(p,q)}{\pi_k(p)\pi_k(q)} - 1\right\}^{+}\right].
\end{aligned}$$

Therefore, under (3) and (4) $\lim_{k\to\infty}\ \mathrm{var}(\widehat{T}_k) = 0$ and $\widehat{T}_k$ converges in probability to T. □

A.2. Proof of Proposition 2

Since $|y_j| \leq L$, from (7)

$$\begin{aligned}
\mathrm{var}\Big(\frac{\widehat{T}_k}{T_k}\Big) &= \frac{1}{T_k^2}\Big\{\sum_{j\in\mathcal{U}_k}\Big(\frac{1}{\pi_j^{(k)}}-1\Big)y_j^2 + 2\sum_{h>j\in\mathcal{U}_k}\Big(\frac{\pi_{jh}^{(k)}}{\pi_j^{(k)}\pi_h^{(k)}}-1\Big)y_jy_h\Big\} \\
&\leq \frac{L}{T_k^2}\sum_{j\in\mathcal{U}_k}\Big(\frac{1}{\pi_j^{(k)}}-1\Big)y_j + \frac{2}{T_k^2}\sum_{h>j\in\mathcal{U}_k}\Big(\frac{\pi_{jh}^{(k)}}{\pi_j^{(k)}\pi_h^{(k)}}-1\Big)^+ y_jy_h \\
&\leq \frac{L}{T_k^2}\max_j\Big(\frac{1}{\pi_j^{(k)}}-1\Big)\sum_{j\in\mathcal{U}_k}y_j + \frac{2}{T_k^2}\max_{h>j}\Big(\frac{\pi_{jh}^{(k)}}{\pi_j^{(k)}\pi_h^{(k)}}-1\Big)^+\sum_{h>j\in\mathcal{U}_k}y_jy_h \\
&\leq \frac{L}{T_k}\max_j\Big(\frac{1}{\pi_j^{(k)}}-1\Big) + \max_{h>j}\Big(\frac{\pi_{jh}^{(k)}}{\pi_j^{(k)}\pi_h^{(k)}}-1\Big)^+.
\end{aligned}$$

Therefore, under condition (8) and if there exists $\pi_0 > 0$ such that

$$\min_j \pi_j^{(k)} \geq \pi_0,$$

then $\lim_{k\to\infty} \mathrm{var}(\widehat{T}_k/T_k) = 0$ and because $E(\widehat{T}_k/T_k) = 1$, $\widehat{T}_k/T_k$ converges in probability to 1. □

A.3. Proof of Proposition 3

In the case of continuous populations, $T_l^{(k)} = \int_{\mathcal{A}_l^{(k)}} y(p)\lambda(dp) \leq L\,\lambda(\mathcal{A}_l^{(k)})$, therefore the first term in (13) verifies

$$\begin{aligned}
\sum_{l\in\mathcal{P}_k}\frac{1-\pi_l^{(k)}}{\pi_l^{(k)}}T_l^{(k)2} &\leq L\sum_{l\in\mathcal{P}_k}\frac{1-\pi_l^{(k)}}{\pi_l^{(k)}}\lambda(\mathcal{A}_l^{(k)})T_l^{(k)} \\
&\leq L\sum_{l\in\mathcal{P}_k}\frac{\lambda(\mathcal{A}_l^{(k)})}{\pi_l^{(k)}}T_l^{(k)} \\
&\leq L^2\max_l\frac{\lambda(\mathcal{A}_l^{(k)})}{\pi_l^{(k)}}\sum_{l\in\mathcal{P}_k}\lambda(\mathcal{A}_l^{(k)}) = L^2\lambda(\mathcal{A})\max_l\frac{\lambda(\mathcal{A}_l^{(k)})}{\pi_l^{(k)}}
\end{aligned}$$

and the second term satisfies

$$\begin{aligned}
2\sum_{h>l\in\mathcal{P}_k}\Big(\frac{\pi_{lh}^{(k)}}{\pi_l^{(k)}\pi_h^{(k)}}-1\Big)T_l^{(k)}T_h^{(k)} &\leq 2L^2\sum_{h>l\in\mathcal{P}_k}\Big(\frac{\pi_{lh}^{(k)}}{\pi_l^{(k)}\pi_h^{(k)}}-1\Big)^+\lambda(\mathcal{A}_l^{(k)})\lambda(\mathcal{A}_h^{(k)}) \\
&\leq L^2\max_{h>l}\Big(\frac{\pi_{lh}^{(k)}}{\pi_l^{(k)}\pi_h^{(k)}}-1\Big)^+\sum_{h\neq l\in\mathcal{P}_k}\lambda(\mathcal{A}_l^{(k)})\lambda(\mathcal{A}_h^{(k)}) \\
&\leq L^2\lambda^2(\mathcal{A})\max_{h>l}\Big(\frac{\pi_{lh}^{(k)}}{\pi_l^{(k)}\pi_h^{(k)}}-1\Big)^+.
\end{aligned}$$

The two inequalities jointly imply

$$\mathrm{var}(\widehat{T}_k) \leq L^2\lambda(\mathcal{A})\max_l\frac{\lambda(\mathcal{A}_l^{(k)})}{\pi_l^{(k)}} + L^2\lambda^2(\mathcal{A})\max_{h>l}\Big(\frac{\pi_{lh}^{(k)}}{\pi_l^{(k)}\pi_h^{(k)}}-1\Big)^+$$

then, under (14) and (15), $\lim_{k\to\infty} \mathrm{var}(\widehat{T}_k) = 0$ and $\widehat{T}_k$ converges in probability to T. □

A.4. Proof of Proposition 4

In the case of finite population of units, from (13)

$$\begin{aligned}
\mathrm{var}(\frac{\widehat{T}_k}{T_k}) &= \frac{1}{T_k^2}\Big\{\sum_{l\in\mathcal{P}_k}\Big(\frac{1}{\pi_l^{(k)}}-1\Big)T_l^{(k)2}+2\sum_{h>l\in\mathcal{P}_k}\Big(\frac{\pi_{lh}^{(k)}}{\pi_l^{(k)}\pi_h^{(k)}}-1\Big)T_l^{(k)}T_h^{(k)}\Big\}\\
&\le \frac{1}{T_k}\sum_{l\in\mathcal{P}_k}\frac{T_l^{(k)}}{\pi_l^{(k)}T_k}T_l^{(k)}+\frac{2}{T_k^2}\sum_{h>l\in\mathcal{P}_k}\Big(\frac{\pi_{lh}^{(k)}}{\pi_l^{(k)}\pi_h^{(k)}}-1\Big)^+T_l^{(k)}T_h^{(k)}\\
&\le \frac{1}{T_k}\max_l\frac{T_l^{(k)}}{\pi_l^{(k)}T_k}\sum_{l\in\mathcal{P}_k}T_l^{(k)}+\frac{2}{T_k^2}\max_{h>l}\Big(\frac{\pi_{lh}^{(k)}}{\pi_l^{(k)}\pi_h^{(k)}}-1\Big)^+\sum_{h>l\in\mathcal{P}_k}T_l^{(k)}T_h^{(k)}\\
&\le \max_l\frac{T_l^{(k)}}{\pi_l^{(k)}T_k}+\max_{h>l}\Big(\frac{\pi_{lh}^{(k)}}{\pi_l^{(k)}\pi_h^{(k)}}-1\Big)^+.
\end{aligned}$$

Under condition (15) and (16) $\lim_{k\to\infty}\ \mathrm{var}(\widehat{T}_k/T_k)=0$ and because $E(\widehat{T}_k/T_k)=1$, $\widehat{T}_k/T_k$ converges in probability to 1. □

Appendix B. Relationships (10) and (11)

As to (10), because $I_j(p)$ is equal to 1 if $p\in\mathcal{B}_j$ and equal to 0 otherwise, from the definition of $g(p)$, it follows that

$$\begin{aligned}
\int_{\mathcal{B}} g(p)\lambda(dp) &= \sum_{j\in\mathcal{U}}\int_{\mathcal{B}}\frac{y_j}{\lambda(\mathcal{B}_j)}I_j(p)\lambda(dp)\\
&= \sum_{j\in\mathcal{U}}\frac{y_j}{\lambda(\mathcal{B}_j)}\int_{\mathcal{B}}I_j(p)\lambda(dp)=\sum_{j\in\mathcal{U}}y_j=T.
\end{aligned}$$

As to (11), from (9) it follows that

$$\widehat{T}=\lambda(\mathcal{B})\sum_{j\in\mathcal{S}(p)}\frac{y_j}{\lambda(\mathcal{B}_j)}=\lambda(\mathcal{B})\sum_{j\in\mathcal{U}}\frac{y_j}{\lambda(\mathcal{B}_j)}I_j(p)=\lambda(\mathcal{B})g(p).\quad\square$$

Appendix C. Consistency of systematic grid sampling

Let $\mathcal{A}_1^{(k)},\ldots,\mathcal{A}_{n_k}^{(k)}$ be the n_k regular polygons of size $\lambda(\mathcal{A})/n_k$ partitioning $\mathcal{A}$. Moreover, let $P_{k,1}$ be a random point in $\mathcal{A}_1^{(k)}$ and $P_{k,2},\ldots,P_{k,n_k}$ the systematically repeated points in $\mathcal{A}_2^{(k)},\ldots,\mathcal{A}_{n_k}^{(k)}$ in such a way that

$$\widehat{T}_k=\frac{\lambda(\mathcal{A})}{n_k}\sum_{i=1}^{n_k}y(P_{k,i})$$

is the Monte Carlo estimator (5). For an arbitrary $\epsilon>0$, let g be a continuous function such that $\int_{\mathcal{A}}|y(p)-g(p)|\ \lambda(dp)<\epsilon$ and let

$$\widehat{T}_k'=\frac{\lambda(\mathcal{A})}{n_k}\sum_{i=1}^{n_k}g(P_{k,i})$$

be the Monte Carlo estimator of $T'=\int_{\mathcal{A}}g(p)\ \lambda(dp)$. Since

$$E[|\widehat{T}_k-T|]\le E[|\widehat{T}_k-\widehat{T}_k'|]+E[|\widehat{T}_k'-T'|]+|T'-T|$$

and the following inequalities hold

$$|T' - T| < \epsilon$$

$$E[|\widehat{T}_k - \widehat{T}'_k|] \leq \frac{\lambda(\mathcal{A})}{n_k} \sum_{i=1}^{n_k} E[|y(P_{k,i}) - g(P_{k,i})|] = \int_{\mathcal{A}} |y(p) - g(p)| \, \lambda(dp) < \epsilon,$$

for the arbitrariness of ϵ, it suffices to prove the consistency of $\widehat{T}'_k$. To this aim, thanks to uniform continuity of g on $\mathcal{A}$, for any $\epsilon > 0$ there exists an integer k_0 such that

$$\Big|\frac{\lambda(\mathcal{A})}{n_k} g(P_{k,i}) - \int_{\mathcal{A}_i^{(k)}} g(p)\,\lambda(dp)\Big| \leq \int_{\mathcal{A}_i^{(k)}} |g(P_{k,i}) - g(p)| \, \lambda(dp) < \frac{\lambda(\mathcal{A})}{n_k}\epsilon$$

for any $k > k_0$ and for any $i = 1, \ldots, n_k$. Then, when $k > k_0$,

$$|\widehat{T}'_k - T'| \leq \sum_{i=1}^{n_k} \Big|\frac{\lambda(\mathcal{A})}{n_k} g(P_{k,i}) - \int_{\mathcal{A}_i^{(k)}} g(p)\,\lambda(dp)\Big| < \lambda(\mathcal{A})\epsilon$$

and the consistency is proven.

Appendix D. Consistency of systematic sampling of areal units in the continuous case

Since in each stratum there are $M_k/m_k = m_0$ regular polygons of size $\lambda(\mathcal{A})/M_k$, denote by $\mathcal{A}_{1,l}^{(k)}, \ldots, \mathcal{A}_{m_0,l}^{(k)}$ the polygons in the stratum l and by $T_{1,l}^{(k)}, \ldots, T_{m_0,l}^{(k)}$ the corresponding totals within polygons. Under systematic sampling, the Horvitz–Thompson estimator (12) can be rewritten as

$$\widehat{T}_k = m_0 \sum_{h=1}^{m_0} s_h^{(k)} I_{h,k}$$

where $s_h^{(k)} = \sum_{l=1}^{m_k} T_{h,l}^{(k)} = \int_{\mathcal{A}_h^{(k)}} y(p)\,\lambda(dp)$, while $I_{h,k}$ is equal to 1 if $\mathcal{A}_{h,1}^{(k)}$ is sampled and equal to 0 otherwise and $\mathcal{A}_h^{(k)} = \cup_{l=1}^{m_k} \mathcal{A}_{h,l}^{(k)}$ is the union of the polygons constituting the systematic sample $(h = 1, \ldots, m_0)$. Moreover, let g be a continuous function on $\mathcal{A}$ such that

$$\int_{\mathcal{A}} |y(p) - g(p)| \, \lambda(dp) < \epsilon$$

for a fixed $\epsilon > 0$. If $T' = \int_{\mathcal{A}} g(p)\lambda(dp)$, $\widehat{T}'_k = m_0 \sum_{h=1}^{m_0} s_h'^{(k)} I_{h,k}$, with $s_h'^{(k)} = \sum_{l=1}^{m_k} T_{h,l}'^{(k)} = \int_{\mathcal{A}_h^{(k)}} g(p)\,\lambda(dp)$ and $T_{1,l}'^{(k)}, \ldots, T_{m_0,l}'^{(k)}$ represent the totals within polygons in stratum l, it holds

$$E[|\widehat{T}_k - T|] \leq E[|\widehat{T}_k - \widehat{T}'_k|] + E[|\widehat{T}'_k - T'|] + [|T' - T|].$$

Since $[|T' - T|] < \epsilon$ and

$$|\widehat{T}_k - \widehat{T}'_k| = |m_0 \sum_{h=1}^{m_0} s_h^{(k)} I_{h,k} - m_0 \sum_{h=1}^{m_0} s_h'^{(k)} I_{h,k}| = |m_0 \sum_{h=1}^{m_0} \int_{\mathcal{A}_h^{(k)}} (y(p) - g(p)) \, \lambda(dp) I_{h,k}|$$
$$\leq m_0 \sum_{h=1}^{m_0} \int_{\mathcal{A}_h^{(k)}} |y(p) - g(p)| \, \lambda(dp) I_{h,k} \leq m_0 \epsilon,$$

for the arbitrariness of ϵ, it suffices to prove consistency of $\widehat{T}'_k$. Thanks to uniform continuity of g on $\mathcal{A}$, fixed $\epsilon > 0$, there exists an integer k_0 such that for any $k > k_0$ and any $h = 1, \ldots, m_0$ it holds

$$\max_{l=1,\ldots,m_k} |T_{1,l}'^{(k)} - T_{h,l}'^{(k)}| < \frac{\lambda(\mathcal{A})}{M_k}\epsilon$$

in such a way that

$$|s_1'^{(k)} - s_h'^{(k)}| \leq \sum_{l=1}^{m_k} |T_{1,l}'^{(k)} - T_{h,l}'^{(k)}| \leq \frac{\lambda(\mathcal{A})}{m_0}\epsilon$$

and hence

$$|m_0 s_1^{\prime(k)} - T'| = |\sum_{h=1}^{m_0}(s_1^{\prime(k)} - s_h^{\prime(k)})| \leq \sum_{h=1}^{m_0} |s_1^{\prime(k)} - s_h^{\prime(k)}| \leq \lambda(\mathcal{A})\epsilon.$$

Then, it follows that

$$|m_0 s_1^{\prime(k)} - \widehat{T}_k'| = m_0|\sum_{h=1}^{m_0}(s_1^{\prime(k)} - s_h^{\prime(k)})I_{h,k}| \leq m_0 \sum_{h=1}^{m_0} |s_1^{\prime(k)} - s_h^{\prime(k)}|I_{h,k} \leq \lambda(\mathcal{A})\epsilon.$$

Finally it follows

$$|\widehat{T}_k' - T'| = |\widehat{T}_k' - m_0 s_1^{\prime(k)} + m_0 s_1^{\prime(k)} - T'| \leq |\widehat{T}_k' - m_0 s_1^{\prime(k)}| + |m_0 s_1^{\prime(k)} - T'| \leq 2\lambda(\mathcal{A})\epsilon$$

and the consistency of $\widehat{T}_k'$ is ensured from the arbitrariness of ϵ. □

References

Barabesi, L., Fattorini, L., Marcheselli, M., Pisani, C., Pratelli, L., 2015. The estimation of diversity indexes by using stratified allocations of plots, points or transects. Environmetrics 26, 202–215.
Barabesi, L., Franceschi, S., 2011. Sampling properties of spatial total estimators under tessellation stratified designs. Environmetrics 22, 271–278.
Barabesi, L., Franceschi, S., Marcheselli, M., 2012. Properties of design-based estimation under stratified spatial sampling. Ann. Appl. Stat. 6, 210–228.
Barabesi, L., Marcheselli, M., 2005. Monte Carlo integration strategies for design-based regression estimators of spatial mean. Environmetrics 16, 803–817.
Berger, Y.G., 1998. Rate of convergence to normal distribution for the Horvitz-Thompson estimator. J. Stat. Plan. Inference 67, 209–226.
Bethel, J., 1989. Minimum variance estimation in stratified sampling. J. Am. Stat. Assoc. 84, 260–265.
Breidt, F.J., 1995. Markov chain designs for one-per-stratum sampling. Surv. Methodol. 21, 63–70.
Brown, J.A., Robertson, B.L., Mc Donald, T., 2015. Spatially balanced sampling: applications to environmental surveys. Procedia Environ. Sci. 27, 6–9.
Brus, D.J., 2000. Using regression models in design-based estimation of spatial mean of soil properties. Eur. J. Soil Sci. 51, 159–172.
Cochran, W.G., 1977. Sampling Techniques. Wiley, New York.
Cordy, C.B., 1993. An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe. Stat. Probab. Lett. 18, 353–362.
De Gruijter, J.J., Brus, D.J., Bierkens, M.F.P., Knotters, M., 2010. Sampling for Natural Resource Monitoring. Springer, Berlin.
De Vries, P.G., 1986. Sampling Theory for Forest Inventories. Springer-Verlag, Berlin.
Fattorini, L., 2006. Applying the Horvitz–Thompson criterion in complex designs: a computer-intensive perspective for estimating inclusion probabilities. Biometrika 93, 269–278.
Fattorini, L., 2009. An adaptive algorithm for estimating inclusion probabilities and performing the Horvitz–Thompson criterion in complex designs. Comput. Statist. 24, 623–639.
Fattorini, L., 2014. Design-based methodological advances to support national forest inventories: a review of recent proposals. iForest 8, 6–11.
Fattorini, L., Corona, P., Chirici, G., Pagliarella, M.C., 2015. Design-based strategies for sampling spatial units from regular grids with applications to forest surveys, land use and land cover estimation. Environmetrics 26, 216–228.
Fattorini, L., Marcheselli, M., Pisani, C., 2006. A three-phase sampling strategy for large-scale multiresource forest inventories. J. Agric. Biol. Environ. Stat. 11, 1–21.
Fattorini, L., Marcheselli, M., Pisani, C., Pratelli, L., 2017. Design-based asymptotics for two-phase sampling strategies in environmental surveys. Biometrika 104, 195–205.
Grafström, A., 2012. Spatial correlated Poisson sampling. J. Stat. Plan. Inference 142, 139–147.
Grafström, A., Lundström, N.L.P., Schelin, L., 2012. Spatially balanced sampling through the pivotal method. Biometrics 68, 514–520.
Grafström, A., Tillé, Y., 2013. Doubly balanced spatial sampling with spreading and restitution of auxiliary totals. Environmetrics 24, 120–131.
Gregoire, T.G., Valentine, H.T., 2008. Sampling Strategies for Natural Resources and the Environment. Chapman & Hall/CRC, Boca Raton.
Hedayat, A.S., Sinha, B.K., 1991. Design and Inference in Finite Population Sampling. Wiley, New York.
Isaki, C.T., Fuller, W.A., 1982. Survey design under the regression superpopulation model. J. Amer. Statist. Assoc. 77, 89–96.
Lister, A.J., Scott, C.T., 2009. Use of space-filling curves to select sample locations in natural resource monitoring studies. Environ. Monit. Assess. 149, 71–80.
Mandallaz, D., 2008. Sampling Techniques for Forest Inventories. Chapman & Hall, Boca Raton.

Marchetti, M., Garfi, V., Pisani, C., Franceschi, S., Marcheselli, M., Corona, P., Puletti, N., Vizzarri, M., Di Cristoforo, M., Ottaviani, M., Fattorini, L., 2018. Inference on forest attributes and ecological diversity of tree-outside-forest by a two-phase inventory. Ann. Forest Sci. 75 (37).
Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. Biometrika 37, 17–23.
Opsomer, J.D., Breidt, F.G., Moisen, G.G., Kauermann, G., 2007. Model-assisted estimation of forest resources with generalized additive models. J. Amer. Statist. Assoc. 102, 400–416.
Pagliarella, M.C., Corona, P., Fattorini, L., 2018. Spatially-balanced sampling vs unbalanced stratified sampling for assessing forest change: evidences in favor of spatial balance. Environ. Ecol. Stat. 25, 111–123.
Pagliarella, M.C., Sallustio, L., Capobianco, G., Conte, E., Corona, P., Fattorini, L., Marchetti, M., 2016. From one- to two-phase sampling to reduce costs of remote sensing-based estimation of land-cover and land-use proportions and their changes. Remote Sens. Environ. 184, 410–417.
Prášková, Z., Sen, P.K., 2009. Asymptotics in finite population sampling. In: Pfeffermann, D., Rao, C.R. (Eds.), Sample Surveys: Design, Methods and Applications. Elsevier.
Rosén, B., 1997. Asymptotic theory for order sampling. J. Statist. Plann. Inference 62, 135–158.
Särndal, C.E., Swensson, B., Wretman, J., 1992. Model Assisted Survey Sampling. Springer, New York.
Stevens, D.J., Olsen, A.R., 2004. Spatially balanced sampling of natural resources. J. Amer. Statist. Assoc. 99, 262–278.
Thompson, S.K., 2002. Sampling, second ed. Wiley, New York.
Tomppo, L.M., Gschwantner, T., Laurence, M., McRoberts, R.E., 2010. National Forest Inventories: Pathways for Common Reporting. Springer, Heidelberg.
Wang, J.F., Jiang, C., Hu, M., Cao, Z., Guo, Y., Li, L., Liu, T., Meng, B., 2013. Design based spatial sampling: Theory and implementation. Environ. Model. Softw. 40, 280–288.
Wang, J.F., Zhangb, T.L., Fu, B.J., 2016. A measure of spatial stratified heterogeneity. Ecol. Indic. 67, 250–256.