Electronic Journal of Statistics
Vol. 13 (2019) 2121–2149
ISSN: 1935-7524
https://doi.org/10.1214/19-EJS1572

Design-based mapping for finite populations of marked points

Lorenzo Fattorini[1,*], **Marzia Marcheselli**[1,**], **Caterina Pisani**[1,†] **and Luca Pratelli**[2,‡]

[1]*Department of Economics and Statistics, University of Siena, Piazza S. Francesco, 8 53100 Siena, Italy*
e-mail: [*]lorenzo.fattorini@unisi.it; [**]marzia.marcheselli@unisi.it *e-mail:* [†]caterina.pisani@unisi.it

[2]*Naval Academy, viale Italia 72, 57100 Livorno, Italy*
e-mail: [‡]luca_pratelli@marina.difesa.it

Abstract: The estimation of marks for a finite population of points scattered onto a study region is considered when a sample of these points is selected by a probabilistic sampling scheme. At each point, the mark is estimated by means of an inverse distance weighting interpolator. The design-based asymptotic properties of the resulting maps are derived when the study area remains fixed, a sequence of nested populations with increasing size is considered and samples of increasing size are selected. Conditions ensuring design-based asymptotic unbiasedness and consistency are given. They essentially require that marks are the values of a pointwise or uniformly continuous deterministic function, the enlargement of the populations is rather regular and the sequence of sampling designs ensures an asymptotic spatial balance. A computationally simple mean squared error estimator is proposed. A simulation study is performed to assess the theoretical results on artificial populations. Finally, an application for mapping the values of the height of trees in a forest stand located in North Italy is reported.

MSC 2010 subject classifications: Primary 62G05; secondary 62G20.
Keywords and phrases: Marked populations, mark estimation, sampling, inverse distance weighting interpolator, design consistency, Monte Carlo study.

Received June 2018.

Contents

1. Introduction

Finite populations of marked points describe the positions and the characteristics of a finite collection of units spread in a study region. As typical examples, units/points are towns in a geographic region and marks are their population sizes ([3]); in forest studies, units/points are trees in a forest stand and marks are tree heights or stem diameters (e.g. [9]; [21]) or tree crown dimensions delineated from high spatial resolution remotely sensed imagery (e.g. [4]); in natural hazard studies, points are earthquake epicenters and marks are their corresponding magnitudes (e.g. [14]).

In most cases, finite populations of marked points are analysed in a model-based framework, supposing that they are outcomes of marked point processes in the plane. Model construction for marked point processes is a quite complex task involving a complex sequence of assumptions. In the simplest case, when marks are supposed to be independently and identically distributed and independent of point locations, the model is referred to as independently marked (or randomly labelled) point process. This model consists of two independent random components: a point process and a sequence of i.i.d. marks. In more realistic cases, marks are supposed to be generated by a random field, i.e. they may be dependent. These models were introduced by [18] and subsequently adopted by [20] to analyse observed gauge measurements for rainfalls. They are referred to as geostatistical (or external) marking or random field models. However, assuming independence between marks and locations may be unrealistic, especially in presence of interactions between points, as obvious in the biological context of competition. More realistically, marks and points should be supposed to be dependent giving rise to models referred to as non-geostatistical marking. Models assuming a close relationship between point density and marks are the density-dependent marked Cox processes ([19]).

We do not want to address the huge literature regarding marked point processes and model-based inference on finite populations of marked points (see [16] for a comprehensive review on this topic). Rather, we attempt to reconstruct finite populations of marked points in a design-based framework, i.e. treating points and marks as fixed characteristics and deriving the properties of the interpolated marks on the basis of the sampling scheme adopted to select a sample of points on which marks are recorded. As pointed out by [22], the main attraction of the design-based approach is that "Design-based inference is objective, nobody can challenge that the sample was really selected according to the given sampling design. The probability distribution associated with the design is real, not modelled or assumed".

However, when any assumption about the marked point process generating the population and marks is avoided and points and marks are considered as fixed attributes, then uncertainty only stems from the sampling scheme and making maps becomes challenging. Indeed, when estimating the mark of a single point, either the point is sampled and there is no need for estimation, or the point is unsampled so that we have no information about it to perform estimation. In these cases the use of an assisting model to estimate unsampled values is unavoidable.

Recently, [2] (see also [27]) exploit the very simple Tobler's first law of geography as assisting model, i.e. units that are close in space tend to be more similar than units that are far apart ([26]). Accordingly, the authors adopt the so-called inverse distance weighting (IDW) interpolator in which the marks at the not sampled points are estimated by a weighted sum of the sampled marks with weights inversely decreasing with distances to the point under estimation.

However, in a design-based framework, the IDW estimator - as any other possible model-assisted estimator - is biased and there is no way to achieve general conclusions about its properties. Actually, from a design-based point of view, the sole way to render statistically sound the IDW estimation is to prove some sort of design-based asymptotic unbiasedness and consistency. The main purpose of this paper is to determine the conditions ensuring unbiasedness and consistency of the IDW interpolator as the study area remains fixed and population and sample sizes approach infinity.

The paper is organized as follows. Section 2 provides the statement of the problem. In Section 3 the asymptotic framework is delineated and conditions ensuring asymptotic unbiasedness and consistency are derived. In Section 4 some sampling schemes are considered, some of them of wide applications in forest and environmental surveys, under which unbiasedness and consistency are achieved. A computationally simple, asymptotically conservative estimator of the mean square error of the IDW interpolator is proposed in Section 5. The theoretical results are assessed by a simulation study performed on artificial populations of marked points in Section 6. An application of the method on a population of trees located in North Italy is reported in Section 7. Concluding remarks are given in Section 8. Proofs of the main results of Sections 3, 4 and 5 are reported in the Appendix.

2. Statement of the problem

Consider a study region $\mathcal{A}$ which is supposed to be a connected and compact set of $\mathbb{R}^2$. Moreover, suppose a population of N points $p_1, \dots, p_N$ scattered onto $\mathcal{A}$. In the following, with a slight abuse of notation, $\mathcal{U}$ will denote both the set of the N points and the set of indexes $\{1, \dots, N\}$ identifying them. For each pair of points $h > j \in \mathcal{U}$ denote by $d_{jh} = \|p_j - p_h\|$, where $\|\|$ denotes a norm in $\mathbb{R}^2$. Moreover, for each point j, let y_j be the amount of the survey variable Y, usually referred to as the mark of unit j.

We are interested in reconstructing the marked population, i.e. in estimating the value y_j for each $j \in \mathcal{U}$ by using the values recorded in a sample $\mathcal{S}$ of n points selected from $\mathcal{U}$ by means of a sampling scheme inducing invariably positive first-order inclusion probabilities π_j and second-order inclusion probabilities π_{jh} ($h > j \in \mathcal{U}$).

Exploiting the law of geography by [26] in a model-assisted framework, [2] propose to estimate the y_js by means of IDW interpolator

$$\hat{y}_j = Z_j y_j + (1 - Z_j) \sum_{i \in \mathcal{U}} w_{ij} y_i \tag{2.1}$$

where Z_j is the random variable equal to 1 if $j \in \mathcal{S}$ and 0 otherwise,

$$w_{ij} = \frac{Z_i \phi(d_{ij})}{\sum_{l \in \mathcal{U}} Z_l \phi(d_{lj})}$$

is the weight attached to the mark of unit i to estimate the mark of unit j and $\phi : [0, \infty) \to \mathbb{R}^+$ is a not increasing function on $]0, \infty)$, with $\phi(0) = 0$ and $\lim_{d \to 0^+} \phi(d) = \infty$. From (2.1) it follows that $\hat{y}_j$ is equal to the true value y_j when $j \in \mathcal{S}$. As in any design-based approach to inference, the sole random variables involved in (2.1) are the Z_js that describe the sampling outcome.

The design-based expectation and variance of (2.1) (see [2]) are

$$E(\hat{y}_j) = \pi_j y_j + \sum_{i \in \mathcal{U}} E\{(1 - Z_j) w_{ij}\} y_i$$

$$V(\hat{y}_j) = \pi_j y_j^2 + \sum_{h,i \in \mathcal{U}} E\{(1 - Z_j) w_{ij} w_{hj}\} y_i y_h - \{E(\hat{y}_j)\}^2.$$

Because no theoretical result about bias and precison is available, investigations are needed to derive conditions under which design-based asymptotic unbiasedness and consistency hold.

3. Asymptotic results

As is customary in the finite population asymptotic framework (e.g. [22], Section 5.3) let $\mathcal{V} = \{p_1, p_2, \dots\}$ be an infinite sequence of points onto $\mathcal{A}$ and $y(\mathcal{V}) = \{y_1, y_2, \dots\}$ be the corresponding sequence of marks. A sequence $\{\mathcal{U}_k\}$

of populations is considered where $\mathcal{U}_1$ consists of the first N_1 points from $\mathcal{V}$, $\mathcal{U}_2$ consists of the first N_2 points from $\mathcal{V}$ with $N_2 > N_1$, and so on, in such a way that $\{\mathcal{U}_k\}$ turns out to be a sequence of nested populations of increasing sizes. Finally suppose a sequence of designs to select a sample $\mathcal{S}_k$ of size n_k from $\mathcal{U}_k$. For each k and for each pair $h > j \in \mathcal{U}_k$ denote by $\pi_j^{(k)}$ and $\pi_{jh}^{(k)}$ the first- and second-order inclusion probabilities induced by the kth design and denote by $Z_j^{(k)}$ the indicator variable which equals 1 if unit j is selected from $\mathcal{U}_k$ and 0 otherwise.

In this framework, for any natural number j let $k(j)$ be the population index such that $j \in \mathcal{U}_k$ for any $k \geq k(j)$. For any population $\mathcal{U}_k$, with $k \geq k(j)$, consider the IDW interpolator of y_j. From equation (2.1) it follows that

$$\hat{y}_j^{(k)} = Z_j^{(k)} y_j + (1 - Z_j^{(k)}) \sum_{i \in \mathcal{U}_k} w_{ij}^{(k)} y_i$$

where

$$w_{ij}^{(k)} = \frac{Z_i^{(k)} \phi(d_{ij})}{\sum_{l \in \mathcal{U}_k} Z_l^{(k)} \phi(d_{lj})}.$$

The goal is to determine the asymptotic design-based behaviour of the IDW interpolator as $k \to \infty$, i.e. as the population of points onto $\mathcal{A}$ becomes larger and larger.

The IDW interpolator is defined to be *pointwise design consistent* at p_j if $|\hat{y}_j^{(k)} - y_j|$ converges in probability to 0, i.e.

$$p \lim_{k \to \infty} |\hat{y}_j^{(k)} - y_j| = 0. \tag{3.1}$$

Moreover, the IDW interpolator is defined to be *uniformly design consistent* with respect to $\{\mathcal{U}_k\}$ if $\sup_{j \in \mathcal{U}_k} |\hat{y}_j^{(k)} - y_j|$ converges in probability to 0, that is

$$p \lim_{k \to \infty} \sup_{j \in \mathcal{U}_k} |\hat{y}_j^{(k)} - y_j| = 0. \tag{3.2}$$

3.1. Some notations

For any $\delta > 0$ denote by

$$\mathcal{B}_j(\delta) = \{p : \ ||p - p_j|| \leq \delta\}$$

the closed disc of radius δ centered at p_j, henceforth referred to as the δ-disc of point j and, when $k \geq k(j)$, denote by

$$B_j^{(k)}(\delta) = \{i : \ i \in \mathcal{U}_k, \, p_i \in \mathcal{B}_j(\delta)\}$$

the set of points in $\mathcal{U}_k$ that are in the δ-disc of point j, henceforth referred to as the δ-neighbors of point j in $\mathcal{U}_k$. Moreover, denote by

$$\Delta_j^{(k)}(\delta) = \sup_{i \in B_j^{(k)}(\delta)} |y_i - y_j|$$

the maximum difference between the mark of point j and those of its δ-neighbors in $\mathcal{U}_k$ and by

$$Z_j^{(k)}(\delta) = \sum_{i \in B_j^{(k)}(\delta)} Z_i^{(k)}$$

the number of sampled points among the δ-neighbors of point j.

It should be noticed that

$$\sup_{k \geq k(j)} \Delta_j^{(k)}(\delta) = \sup_{p_i \in \mathcal{B}_j(\delta)} |y_i - y_j|$$

in such a way that y is continuous at p_j if and only if

$$\lim_{\delta \to 0^+} \sup_{k \geq k(j)} \Delta_j^{(k)}(\delta) = 0. \tag{3.3}$$

Moreover, denoting by

$$\Delta(\delta) = \sup_k \max_{j \in \mathcal{U}_k} \Delta_j^{(k)}(\delta)$$

then y is continuous on $\mathcal{V}$ if and only if

$$\lim_{\delta \to 0^+} \Delta(\delta) = 0. \tag{3.4}$$

3.2. The basic assumptions

In order to achieve the asymptotic design-based properties of (2.1) we consider that the function y, defined on $\mathcal{V}$ by $y_j = y(p_j)$ for any natural number j, is valued in $[0, L]$. The assumption seems reasonable because, in real world, the values of a survey variable are usually non negative, and even if high, unlikely attain infinity. However, even if y_j takes negative values, assuming $|y_j| \leq L/2$ all the following results still hold. Thus, in the following, it will not be restrictive to suppose y_j valued in $[0, L]$. Moreover, this assumption entails that $\widehat{y}_j^{(k)}$ is bounded by L, in such a way that pointwise or uniform design consistency of the IDW interpolator also entails pointwise or uniform design asymptotic unbiasedness.

For achieving design consistency of (2.1) it is necessary to assume the existence of a real number $v > 0$ such that for each $k = 1, 2, \ldots$

$$\mathcal{V} \subset \bigcup_{j \in \mathcal{U}_k} \mathcal{B}_j(v N_k^{-\frac{1}{2}}). \tag{3.5}$$

The assumption is crucial and entails a sort of regularity in the enlargements of the populations of the sequence, ensuring that for each population $\mathcal{U}_k$ the N_k locations are evenly spread throughout $\mathcal{V}$, in such a way that for an adequate radius the discs centered at these points completely cover $\mathcal{V}$. In more practical words, assumptions (3.5) ensures the absence of isolated points that would be likely to be too far from the sampled points.

If assumption (3.5) holds, then it is reasonable to further require that for a p_j and any arbitrary $\varepsilon > 0$ there exist an integer k_0 and a real number $v > 0$ such that

$$\Pr\{Z_j^{(k)}(vN_k^{-\frac{1}{2}}) = 0\} < \varepsilon \qquad \forall\ k > k_0. \tag{3.6}$$

In other words, assumption (3.6) ensures that for a sufficiently large k, the $vN_k^{-\frac{1}{2}}$-disc of p_j definitively contains sampled points with high probability. A condition more restrictive than (3.6) is obtained if for any $\varepsilon > 0$ there exists an integer k_0 such that

$$\Pr\{\min_{j\in\mathcal{U}_k} Z_j^{(k)}(vN_k^{-\frac{1}{2}}) = 0\} < \varepsilon \qquad \forall\ k > k_0. \tag{3.7}$$

Both conditions concern the design sequence, ensuring a spatial balance asymptotically achieved by the sampling scheme, i.e. for a sufficiently large k the scheme is able to evenly spread sampled units in such a way that any unit is likely to have neighboring sampled units.

Finally, regarding the distance function ϕ, we assume that

$$\lim_{d\to 0^+} d^2\phi(d) = \infty. \tag{3.8}$$

If $\phi(d) = d^{-\beta}$, then (3.8) is verified for any $\beta > 2$ and if $\phi(d) = g(d)d^{-2}$ then (3.8) holds when g is a positive function with $\lim_{d\to 0^+} g(d) = \infty$. It should be noticed that actually (3.8) does not constitute an assumption because it can be always ensured by the user.

3.3. Design consistency results

The following result is essential for subsequently proving both pointwise and uniform design consistency (see Section A.1 for the proof).

Lemma. *For any $\alpha, \delta, v > 0$ and for any $j \in \mathcal{U}_k$, with $k \geq k(j)$ it holds that*

$$\begin{aligned} E\{(\widehat{y}_j^{(k)} - y_j)^2\} \leq 2\Delta_j^{(k)}(\delta)^2 + 2L^2\Big[&\Pr\{Z_j^{(k)}(vN_k^{-\alpha}) = 0\} \\ &+ \frac{\phi^2(\delta)n_k^2}{\phi^2(vN_k^{-\alpha})}\Pr\{Z_j^{(k)}(vN_k^{-\alpha}) > 0\}\Big] \end{aligned} \tag{3.9}$$

and

$$\begin{aligned} E\{\max_{j\in\mathcal{U}_k}(\hat{y}_j^{(k)} - y_j)^2\} \leq 2\max_{j\in\mathcal{U}_k}\Delta_j^{(k)}(\delta)^2 + 2L^2\Big[&\Pr\{\min_{j\in\mathcal{U}_k} Z_j^{(k)}(vN_k^{-\alpha}) = 0\} \\ &+ \frac{\phi^2(\delta)n_k^2}{\phi^2(vN_k^{-\alpha})}\Pr\{\min_{j\in\mathcal{U}_k} Z_j^{(k)}(vN_k^{-\alpha}) > 0\}\Big]. \end{aligned} \tag{3.10}$$

Exploiting the previous Lemma, the further result establishes sufficient conditions for pointwise and uniform design consistency (see Section A.2 for the proof).

Result 1. *If y is continuous at p_j, conditions (3.6) and (3.8) ensure pointwise design consistency at p_j. If y is continuous on $\mathcal{V}$, conditions (3.7) and (3.8) ensure uniform design consistency with respect to $\{\mathcal{U}_k\}$.*

Further results on the consistency convergence rate of the IDW interpolator can be achieved if further assumptions are introduced about y and about the balancing features of the sampling scheme.

In particular, suppose that y is a Lipschitz function on $\mathcal{V}$, i.e. there exists a constant C such that

$$|y_i - y_j| \leq C||p_i - p_j||, \qquad \forall \ p_i, p_j \in \mathcal{V} \tag{3.11}$$

and also suppose that for an integer j there exists a real number $c_j > 0$ such that, for any $k > k(j)$,

$$\Pr\{Z_j^{(k)}(\upsilon N_k^{-1/2}) = 0\} \leq c_j \upsilon^{-2} \qquad \text{when} \ \ 0 < \upsilon \leq N_k^{1/2}\text{diam}(\mathcal{A}) \tag{3.12}$$

where diam$(\mathcal{A})$ is the diameter of $\mathcal{A}$. Condition (3.11) entails a limitation in how fast y can change on $\mathcal{V}$. Regarding condition (3.12) it is at once apparent that it implies condition (3.6). Therefore condition (3.12) concerns the design sequence, requiring an enhanced balancing power of the sampling scheme to spread sampled points onto $\mathcal{V}$, in such a way that the probability of having no point sampled within the disc centred at p_j decreases at least proportionally to the disc size. Under these two conditions the following result holds (see Section A.3 for the proof).

Result 2. *If conditions (3.11) and (3.12) hold, and if $\phi(d) = d^{-\beta}$ with $\beta > 2$, then*

$$E\{(\hat{y}_j^{(k)} - y_j)^2\} \leq H n_k^{\frac{2-\beta}{2\beta+1}} \qquad \forall \ k > k(j)$$

where H is a constant depending on β, C, c_j and L.

From Result 2, under conditions (3.11) and (3.12), design consistency is ensured with a $O(n_k^{\frac{2-\beta}{2\beta+1}})$ convergence rate, so that the use of large β values in the distance function seems advisable.

4. Asymptotic results under some spatial schemes

When sampling populations of points scattered onto a study region, a wide variety of sampling schemes is available. The most straightforward scheme is simple random sampling without replacement (SRSWOR). Despite its simplicity, SRSWOR may lead to uneven surveying of the study region.

In order to avoid this drawback, the achievement of spatially balanced samples, i.e. samples in which the selected points are well spread throughout the population, has been the main target for a long time. Spatial balance can be straightforwardly performed by partitioning the study area into strata and then selecting points within strata by SRSWOR, where the number of points selected within each stratum is proportional to the stratum size (proportional allocation).

A more effective spatial balance can be obtained by means of quite complex schemes such as the generalized random tessellation stratified sampling by [24], the draw-by-draw sampling that excludes the selection of contiguous units proposed by Fattorini in [5] and [6], the local pivotal method of the first and second type by [11], the spatially correlated Poisson sampling by [10] and the doubly balanced spatial sampling by [12]. The last one seems the most effective, joining the spatial balance provided by the local pivotal methods with further spatial properties. For example, if points coordinates are used as two balancing variables, the scheme not only provides samples of points well spread onto the region, but samples and the population have approximately the same barycenters. In this sense, the sampling is doubly spatially balanced.

It should be noticed that the use of these schemes, from the most straightforward SRSWOR to the more complex spatially balanced ones, necessitates the list of points in the population together with their locations. Therefore, these schemes cannot be adopted in forest and environmental surveys where we have to sample populations of units such as trees or shrubs scattered over large areas because list and locations of these units are prohibitive to achieve in terms of time and resources. In most cases, natural populations are sampled without knowing the list, by means of plots or transects located on the study area in accordance with probabilistic designs (e.g. [13], Chapters 7, 8, 9). Obviously, in these cases, in absence of list and locations, even the population mapping is precluded.

On the other hand, these problems rarely arise in economic and social studies regarding towns or firms, whose list and locations are readily available from administrative sources. Probably, the unique relevant case in which the mapping of natural populations becomes possible is under a sampling scheme referred to as 3P sampling from the acronym of probability proportional to prediction. The scheme is a variation of Poisson sampling and has been widely used mainly in North America to estimate wood volume available for sale or other forest attributes such as tree height and basal area from small forest stands, usually of no more than $10-15$ hectares in size. The properties and applications of 3P sampling have been studied extensively in major forest measurement texts, and it has been proven to be very effective (e.g. [1], [15], [23], [28]).

Under 3P sampling, all the units of the population are visited by a crew of experts (and hence mapped), a prediction x_j for the value of the survey variable is given by the experts for each unit j of the population and units are independently included in the sample with probabilities $\pi_j = x_j/M$ where M must be large enough to ensure that $\pi_j \leq 1$ for each j ([13]). Since when 3P sampling is implemented all the units locations are recorded, marks can be estimated for the unsampled units.

Unfortunately, condition (3.7), necessary for uniform design consistency, cannot be proven for most schemes. Even condition (3.6), necessary for pointwise design consistency, may be difficult to prove. Therefore, for proving pointwise consistency at least, it is useful to give sufficient conditions for (3.6) in terms of first- and second-order inclusion probabilities, as customary in some consistency theorems of finite population sampling.

To this aim we not only suppose condition (3.5) for $\mathcal{V}$, but we further require that it is *regular*, i.e. for any $p_j \in \mathcal{V}$ and for any natural number m there exist a real number $u > 0$ and an integer k_0 such that

$$\mathrm{Card}\{B_j^{(k)}(uN_k^{-1/2}) \cap \mathcal{U}_k\} > m \qquad \forall\ k > k_0. \tag{4.1}$$

Condition (4.1) ensures that for a sufficiently large k, any population point has many neighboring points around. The following result states sufficient conditions for (3.6) (see Section A.4 for the proof).

Result 3. *If $\mathcal{V}$ is regular and if*

$$\inf_k \min_{j \in \mathcal{U}_k} \pi_j^{(k)} > 0, \tag{4.2}$$

and

$$\lim_{k\to\infty} \max_{h>j\in\mathcal{U}_k} \left(\frac{\pi_{jh}^{(k)}}{\pi_j^{(k)}\pi_h^{(k)}} - 1\right)^+ = 0 \tag{4.3}$$

then condition (3.6) holds.

Result 3 establishes conditions for (3.6) quite akin to those required in [17] for the design consistency of the Horvitz-Thompson estimator of population totals, i.e. that the design sequence ensures first-order inclusion probabilities invariably greater than a given threshold and that the second-order inclusion probabilities approach the product of the corresponding first-order ones as k increases. Then, Result 3 can be exploited to readily prove pointwise consistency under SRSWOR, stratified spatial sampling with proportional allocation and 3P sampling (see Sections A.5 and A.6 for the proofs).

Finally, regarding the above-mentioned complex schemes proposed to generate spatially balanced samples, owing to their complexity, the second-order inclusion probabilities do not admit closed expressions. Therefore condition (4.3) cannot be adopted to prove their pointwise design consistency. However, their effectiveness in providing spatial balance (see e.g. [25]) and their superiority over SRSWOR empirically evidenced by several simulation studies may induce to presume that pointwise design consistency should also hold for these schemes.

If a further assumption about the spatial pattern of points is introduced, then it is also possible to give conditions ensuring the convergence rate established from Result 2 in terms of first and second-order inclusion probabilities. That can be done by reinforcing the regularity assumption (4.1) supposing that for an integer j there exists a real number $c_j > 0$ such that

$$\mathrm{Card}\{B_j^{(k)}(\upsilon N_k^{-1/2}) \cap \mathcal{U}_k\} \geq c_j \upsilon^2 \qquad \forall\ k > k(j),\ 0 < \upsilon \leq N_k^{1/2}\mathrm{diam}(\mathcal{A}). \tag{4.4}$$

Obviously condition (4.4) implies (4.1), requiring a tendency of the spatial pattern to aggregate points rapidly around p_j, in such a way that the number of neighbors in a disc tend to increase more quickly than the size of the disc. It is worth noting that this tendency occurs under clustered spatial patterns, as is well evidenced, even if in a model-based framework, by the Ripley K function (see e.g. [3], Chapter 8).

The following result states sufficient conditions involving inclusion probabilities for ensuring the convergence rate established by Result 2 (see Section A.7 for the proof).

Result 4. *If conditions (4.2), (4.3) and (4.4) hold and if*

$$\max_{h>i\in\mathcal{U}_k}\left(\frac{\pi_{ih}^{(k)}}{\pi_i^{(k)}\pi_h^{(k)}}-1\right)^+=0 \tag{4.5}$$

then (3.12) holds and, consequently if y is a Lipschitz function on $\mathcal{V}$, Result 2 follows.

Practically speaking, the ability of the design sequence to ensure that the probability of having no point sampled near a point decreases at least proportionally with the area around the point, as established by condition (3.12), is ensured by condition (4.4) joined with some quite obvious characteristics of first- and second-order inclusion probabilities, thus ensuring a convergence rate of the order established by Result 2. For example, if (4.4) holds, this convergence rate holds under stratified spatial sampling with proportional allocation (STRSPA), whose inclusion probabilities satisfy conditions (4.2) and (4.5), as argued in Section A.5.

From these considerations it is apparent that clustered spatial patterns should represent more favorable situations for spatial interpolation than regular or random patterns as well as spatial stratifications are always suitable as a straightforward way to achieve balanced samples.

5. Mean squared error estimation

In order to provide an estimator for the mean squared error (MSE) of (2.1), it should be noted that in practical situations, especially for large areas, the estimation has to be performed for thousands of points. Accordingly any MSE estimator should not be computationally demanding. In this sense, time-consuming resampling procedures such as bootstrap or jackknife should be avoided.

Owing to Tobler's law (which has motivated the estimation), the marks of the sampled points nearest to point j are likely to be a good (known) proxy for y_j. Therefore, a simple estimator for $\mathrm{MSE}(\hat{y}_j)$ is given by

$$\hat{V}_j=(\hat{y}_j-y_{near(j)})^2 \tag{5.1}$$

where $near(j)$ is the label of the unit such that $near(j)\in\mathcal{S}$, $near(j)\neq j$ and

$$\left\|p_j-p_{near(j)}\right\|=\min_{\substack{i\in\mathcal{S}\\ i\neq j}}\|p_j-p_i\|\,.$$

In order to determine the design-based asymptotic properties of (5.1), we still adopt the asymptotic framework of Section 3. Accordingly, for any fixed $j\in\mathcal{U}_k$ and any population $\mathcal{U}_k$ with $k\geq k(j)$ we refer to $\mathrm{MSE}(\hat{y}_j^{(k)})$ and $\hat{V}_j^{(k)}=(\hat{y}_j^{(k)}-$

$\tilde{y}_j^{(k)})^2$ where $\tilde{y}_j^{(k)} = y_{near(k,j)}$ and $near(k,j)$ is the label of the sampled point nearest to point j in the k population of the sequence, i.e. $near(k,j) \in \mathcal{S}_k$, $near(k,j) \neq j$ and

$$\left\| p_j - p_{near(k,j)} \right\| = \min_{\substack{i \in \mathcal{S}_k \\ i \neq j}} \| p_j - p_i \| \,.$$

Under condition (3.6), it is at once apparent that $(1 - Z_j^{(k)}) \| p_j - p_{near(k,j)} \|$ converges in probability to 0. Thus, if p_j is a continuity point of y, it follows that

$$\lim_{k \mapsto \infty} \mathrm{E}\{(y_j - \tilde{y}_j^{(k)})^2\} = 0. \tag{5.2}$$

Moreover, the following result holds (see Section A.8 for the proof).

Result 5. *For any p_j and any k it holds*

$$\mathrm{MSE}(\hat{y}_j^{(k)}) \leq \mathrm{E}(\hat{V}_j^{(k)}) + 3L[\mathrm{E}\{(y_j - \tilde{y}_j^{(k)})^2\}]^{1/2}. \tag{5.3}$$

Result 5 and condition (5.2) jointly prove that $\hat{V}_j^{(k)}$ can be considered an asymptotically conservative estimator of $\mathrm{MSE}(\hat{y}_j^{(k)})$ at any continuity point of y.

6. Simulation studies

6.1. Artificial populations

Two functions, referred to as F1 and F2, were considered on the unit square for generating marks. In particular, for any point $p = (p_1, p_2)$ of the unit square

F1: $y(p) = C_1(\sin 3p_1 \sin^2 3p_2)^2$

F2: $y(p) = \begin{cases} C_2 p_1 p_2 & \min(p_1, p_2) < 0.5 \\ C_2(1 + p_1 p_2) & otherwise \end{cases}$

where the constants C_1, C_2 ensured a maximum mark of 10 in both cases. For any function, three nested populations of $N = 1000, 5000$ and 10000 points were located in the unit square in accordance with four spatial patterns referred to as regular (RE), random (RA), trended (TR) and clustered (CL). All the populations were generated in such a way to ensure a sort of regularity when increasing from 1000 to 10000 points in such a way to approach condition (3.5).

The nested RE populations were constructed by generating the first 1000 points completely at random but discarding those having distances smaller than $1000^{-1/2}$ to those previously generated, then adding further 4000 points completely at random but discarding those having distances smaller than $5000^{-1/2}$ to those previously generated, and finally adding further 5000 points completely at random but discarding those having distances smaller than $10000^{-1/2}$.

The nested RA populations were constructed by simply generating 10000 points completely at random and then assigning the first 1000 to the first population, the first 5000 to the second population and all of them to the third.

The nested TR populations were constructed generating 10000 pairs of random numbers (u_1, u_2) uniformly distributed on (0,1), performing the transformation $(1-u_1^2, 1-u_2^2)$ and then assigning the first 1000 to the first population, the first 5000 to the second population and all of them to the third.

The nested CL populations were constructed generating 10 cluster centres completely at random and then assigning 100 points to each cluster generated from a spherical normal distribution centred on the cluster centre 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.

6.2. Sampling and estimation

For any population arising from the combination of function, spatial pattern and size N, $R = 10000$ samples of size $n = N/10$ were independently selected by means of SRSWOR, STRSPA and doubly balanced spatial sampling (DBALSS). STRSPA 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. DBALSS was performed by selecting units with equal first order inclusion probabilities 0.1 in accordance with the algorithm by [12], balancing the samples with respect to the spatial coordinates. Once the samples were selected, (2.1) was adopted to estimate the marks for all the points in the population by using $d^{-\beta}$ as distance function, with $\beta = 2, 3, 4$. Once the mark estimates were achieved, their MSEs were estimated by means of equation (5.1).

6.3. Performance indicators

At the end of the R simulation runs, the absolute bias (AB)

$$AB_j = \Big| \frac{1}{R} \sum_{i=1}^{R} \hat{y}_{ji} - y_j \Big| \qquad j = 1, \ldots, N$$

the root mean squared error (RMSE)

$$RMSE_j = \Big[\frac{1}{R} \sum_{i=1}^{R} (\hat{y}_{ji} - y_j)^2 \Big]^{1/2}, \qquad j = 1, \ldots, N$$

and the absolute bias of the root MSE estimator (ABRMSEE)

$$ABRMSEE_j = \Big| \frac{1}{R} \sum_{i=1}^{R} \hat{V}_{ji}^{1/2} - RMSE_j \Big|, \qquad j = 1, \ldots, N$$

were computed from the Monte Carlo distributions of the estimates $\hat{y}_{ji}$ and of the RMSE estimates $\hat{V}_{ji}^{1/2}$ $(j = 1, \ldots, N; i = 1, \ldots, R)$. Tables 1-8 report the minima, maxima and averages of ABs, RMSEs and ABRMSEEs for any

Table 1

Minima, maxima and means of ABs, RMSEs and ABRMSEEs achieved with a sampling fraction of 10% for populations arising from F1 and RE, for any combination of population size, sampling scheme and distance function

$\phi(d)$	N	Scheme	AB			RMSE			ABRMSEE		
			Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
d^{-2}	1000	SRSWOR	0.00	2.19	0.93	0.29	2.40	1.13	0.00	1.26	0.28
		STRSPA	0.00	2.10	0.89	0.21	2.30	1.05	0.00	1.22	0.26
		DBALSS	0.00	2.04	0.81	0.14	2.17	0.92	0.00	1.05	0.21
	5000	SRSWOR	0.00	1.87	0.68	0.12	2.02	0.78	0.00	0.71	0.12
		STRSPA	0.00	1.87	0.68	0.09	2.01	0.77	0.00	0.71	0.12
		DBALSS	0.00	1.79	0.62	0.05	1.89	0.67	0.00	0.55	0.09
	10000	SRSWOR	0.00	1.71	0.61	0.08	1.87	0.69	0.00	0.54	0.09
		STRSPA	0.00	1.70	0.61	0.06	1.87	0.68	0.00	0.53	0.09
		DBALSS	0.00	1.64	0.56	0.03	1.74	0.61	0.00	0.45	0.07
d^{-3}	1000	SRSWOR	0.00	1.59	0.45	0.36	1.82	0.73	0.00	1.26	0.26
		STRSPA	0.00	1.55	0.41	0.27	1.77	0.66	0.00	1.22	0.23
		DBALSS	0.00	1.38	0.32	0.17	1.55	0.49	0.00	1.03	0.15
	5000	SRSWOR	0.00	1.00	0.19	0.14	1.21	0.31	0.00	0.72	0.09
		STRSPA	0.00	1.00	0.18	0.13	1.22	0.30	0.00	0.72	0.08
		DBALSS	0.00	0.85	0.13	0.07	0.95	0.20	0.00	0.54	0.05
	10000	SRSWOR	0.00	0.83	0.13	0.07	0.98	0.21	0.00	0.54	0.05
		STRSPA	0.00	0.82	0.13	0.07	0.97	0.21	0.00	0.53	0.05
		DBALSS	0.00	0.69	0.09	0.02	0.78	0.14	0.00	0.42	0.04
d^{-4}	1000	SRSWOR	0.00	1.34	0.28	0.28	1.57	0.62	0.05	1.22	0.30
		STRSPA	0.00	1.30	0.25	0.18	1.52	0.57	0.01	1.18	0.26
		DBALSS	0.00	1.13	0.18	0.07	1.29	0.42	0.00	0.99	0.17
	5000	SRSWOR	0.00	0.75	0.08	0.05	0.90	0.23	0.01	0.68	0.09
		STRSPA	0.00	0.73	0.07	0.05	0.91	0.23	0.00	0.68	0.09
		DBALSS	0.00	0.60	0.05	0.01	0.69	0.16	0.00	0.52	0.05
	10000	SRSWOR	0.00	0.55	0.04	0.02	0.69	0.15	0.00	0.51	0.05
		STRSPA	0.00	0.55	0.04	0.01	0.69	0.15	0.00	0.50	0.05
		DBALSS	0.00	0.41	0.03	0.00	0.52	0.11	0.00	0.38	0.03

Table 2

Minima, maxima and means of ABs, RMSEs and ABRMSEEs achieved with a sampling fraction of 10% for populations arising from F1 and RA, for any combination of population size, sampling scheme and distance function

$\phi(d)$	N	Scheme	AB			RMSE			ABRMSEE		
			Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
d^{-2}	1000	SRSWOR	0.00	2.60	0.88	0.27	2.84	1.09	0.00	1.64	0.29
		STRSPA	0.00	2.66	0.84	0.19	2.88	1.01	0.00	1.62	0.27
		DBALSS	0.00	2.42	0.75	0.13	2.59	0.87	0.00	1.59	0.22
	5000	SRSWOR	0.00	2.08	0.65	0.12	2.22	0.76	0.00	0.88	0.13
		STRSPA	0.00	2.05	0.64	0.09	2.18	0.74	0.00	0.88	0.13
		DBALSS	0.00	2.01	0.57	0.04	2.13	0.64	0.00	0.75	0.10
	10000	SRSWOR	0.00	1.90	0.58	0.08	2.03	0.67	0.00	0.59	0.10
		STRSPA	0.00	1.90	0.58	0.06	2.02	0.66	0.00	0.57	0.10
		DBALSS	0.00	1.85	0.52	0.03	1.95	0.58	0.00	0.54	0.08
d^{-3}	1000	SRSWOR	0.00	1.75	0.44	0.31	2.04	0.72	0.00	1.59	0.27
		STRSPA	0.00	1.70	0.41	0.23	1.99	0.65	0.00	1.57	0.24
		DBALSS	0.00	1.57	0.31	0.09	1.85	0.48	0.00	1.53	0.17
	5000	SRSWOR	0.00	1.33	0.18	0.11	1.47	0.30	0.00	0.89	0.09
		STRSPA	0.00	1.34	0.17	0.10	1.48	0.29	0.00	0.89	0.08
		DBALSS	0.00	1.22	0.12	0.04	1.33	0.19	0.00	0.75	0.06
	10000	SRSWOR	0.00	0.92	0.12	0.07	1.13	0.21	0.00	0.61	0.06
		STRSPA	0.00	0.92	0.12	0.07	1.11	0.20	0.00	0.59	0.06
		DBALSS	0.00	0.80	0.08	0.03	0.89	0.13	0.00	0.49	0.04
d^{-4}	1000	SRSWOR	0.00	1.57	0.28	0.18	1.85	0.62	0.02	1.51	0.31
		STRSPA	0.00	1.53	0.26	0.13	1.80	0.56	0.00	1.49	0.27
		DBALSS	0.00	1.41	0.19	0.04	1.66	0.40	0.00	1.42	0.18
	5000	SRSWOR	0.00	1.05	0.08	0.05	1.19	0.23	0.00	0.88	0.09
		STRSPA	0.00	1.06	0.08	0.04	1.20	0.22	0.00	0.88	0.09
		DBALSS	0.00	0.92	0.06	0.01	1.03	0.16	0.00	0.74	0.06
	10000	SRSWOR	0.00	0.63	0.05	0.02	0.78	0.15	0.00	0.56	0.06
		STRSPA	0.00	0.63	0.05	0.02	0.77	0.15	0.00	0.55	0.06
		DBALSS	0.00	0.52	0.04	0.00	0.58	0.11	0.00	0.45	0.04

TABLE 3

Minima, maxima and means of ABs, RMSEs and ABRMSEEs achieved with a sampling fraction of 10% for populations arising from F1 and TR, for any combination of population size, sampling scheme and distance function

$\phi(d)$	N	Scheme	AB			RMSE			ABRMSEE		
			Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
d^{-2}	1000	SRSWOR	0.00	3.48	0.69	0.06	3.82	0.85	0.00	1.64	0.25
		STRSPA	0.00	3.32	0.66	0.06	3.66	0.79	0.00	1.68	0.23
		DBALSS	0.00	3.20	0.59	0.01	3.51	0.69	0.00	1.65	0.19
	5000	SRSWOR	0.00	3.09	0.51	0.02	3.31	0.60	0.00	1.02	0.11
		STRSPA	0.00	3.06	0.51	0.02	3.27	0.59	0.00	0.99	0.11
		DBALSS	0.00	2.98	0.45	0.01	3.16	0.51	0.00	0.83	0.09
	10000	SRSWOR	0.00	2.65	0.46	0.01	2.84	0.53	0.00	0.72	0.08
		STRSPA	0.00	2.63	0.46	0.01	2.81	0.52	0.00	0.72	0.08
		DBALSS	0.00	2.54	0.41	0.00	2.70	0.46	0.00	0.64	0.06
d^{-3}	1000	SRSWOR	0.00	2.33	0.39	0.02	2.57	0.61	0.00	1.90	0.25
		STRSPA	0.00	2.22	0.35	0.02	2.49	0.55	0.00	1.90	0.21
		DBALSS	0.00	2.23	0.29	0.00	2.46	0.42	0.00	1.88	0.16
	5000	SRSWOR	0.00	1.50	0.17	0.00	1.71	0.27	0.00	1.09	0.09
		STRSPA	0.00	1.43	0.16	0.00	1.65	0.26	0.00	1.04	0.08
		DBALSS	0.00	1.29	0.12	0.00	1.46	0.18	0.00	0.84	0.06
	10000	SRSWOR	0.00	1.25	0.11	0.00	1.42	0.19	0.00	0.77	0.05
		STRSPA	0.00	1.23	0.11	0.00	1.39	0.18	0.00	0.77	0.05
		DBALSS	0.00	1.10	0.08	0.00	1.22	0.12	0.00	0.62	0.04
d^{-4}	1000	SRSWOR	0.00	2.26	0.28	0.01	2.57	0.54	0.00	2.05	0.29
		STRSPA	0.00	2.13	0.24	0.01	2.46	0.48	0.01	2.01	0.24
		DBALSS	0.00	2.13	0.19	0.00	2.42	0.36	0.00	2.01	0.17
	5000	SRSWOR	0.00	1.21	0.08	0.00	1.43	0.21	0.00	1.07	0.09
		STRSPA	0.00	1.15	0.08	0.00	1.37	0.20	0.00	1.02	0.08
		DBALSS	0.00	0.98	0.06	0.00	1.14	0.14	0.00	0.81	0.05
	10000	SRSWOR	0.00	0.86	0.05	0.00	1.01	0.13	0.00	0.74	0.05
		STRSPA	0.00	0.84	0.05	0.00	0.98	0.13	0.00	0.74	0.05
		DBALSS	0.00	0.72	0.04	0.00	0.80	0.09	0.00	0.57	0.03

TABLE 4

Minima, maxima and means of ABs, RMSEs and ABRMSEEs achieved with a sampling fraction of 10% for populations arising from F1 and CL, for any combination of population size, sampling scheme and distance function

$\phi(d)$	N	Scheme	AB			RMSE			ABRMSEE		
			Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
d^{-2}	1000	SRSWOR	0.00	2.38	0.23	0.04	2.58	0.31	0.00	1.21	0.12
		STRSPA	0.00	2.28	0.22	0.03	2.44	0.29	0.00	1.20	0.11
		DBALSS	0.00	2.25	0.19	0.02	2.38	0.23	0.00	1.08	0.09
	5000	SRSWOR	0.00	2.12	0.16	0.01	2.25	0.19	0.00	0.83	0.05
		STRSPA	0.00	2.12	0.16	0.01	2.23	0.18	0.00	0.84	0.05
		DBALSS	0.00	2.09	0.14	0.01	2.20	0.16	0.00	0.86	0.04
	10000	SRSWOR	0.00	2.16	0.14	0.01	2.29	0.17	0.00	0.78	0.04
		STRSPA	0.00	2.16	0.14	0.01	2.28	0.16	0.00	0.78	0.04
		DBALSS	0.00	2.14	0.12	0.00	2.27	0.14	0.00	0.78	0.03
d^{-3}	1000	SRSWOR	0.00	1.56	0.15	0.04	1.69	0.24	0.00	1.22	0.14
		STRSPA	0.00	1.57	0.15	0.03	1.69	0.23	0.00	1.21	0.13
		DBALSS	0.00	1.50	0.12	0.02	1.61	0.17	0.00	1.09	0.09
	5000	SRSWOR	0.00	1.42	0.08	0.01	1.51	0.12	0.00	0.88	0.05
		STRSPA	0.00	1.42	0.08	0.01	1.51	0.12	0.00	0.88	0.05
		DBALSS	0.00	1.36	0.06	0.01	1.45	0.09	0.00	0.90	0.04
	10000	SRSWOR	0.00	1.37	0.06	0.01	1.46	0.09	0.00	0.80	0.03
		STRSPA	0.00	1.38	0.06	0.01	1.46	0.09	0.00	0.81	0.03
		DBALSS	0.00	1.35	0.05	0.00	1.43	0.07	0.00	0.79	0.02
d^{-4}	1000	SRSWOR	0.00	1.43	0.13	0.04	1.57	0.23	0.02	1.24	0.15
		STRSPA	0.00	1.44	0.13	0.03	1.57	0.22	0.01	1.23	0.14
		DBALSS	0.00	1.35	0.10	0.02	1.47	0.17	0.00	1.10	0.10
	5000	SRSWOR	0.00	1.22	0.06	0.01	1.32	0.10	0.00	0.92	0.06
		STRSPA	0.00	1.22	0.06	0.01	1.31	0.10	0.00	0.92	0.06
		DBALSS	0.00	1.14	0.04	0.01	1.23	0.08	0.00	0.92	0.04
	10000	SRSWOR	0.00	1.12	0.04	0.01	1.21	0.07	0.00	0.83	0.04
		STRSPA	0.00	1.12	0.04	0.01	1.21	0.07	0.00	0.83	0.04
		DBALSS	0.00	1.07	0.03	0.00	1.15	0.05	0.00	0.80	0.02

TABLE 5

Minima, maxima and means of ABs, RMSEs and ABRMSEEs achieved with a sampling fraction of 10% for populations arising from F2 and RE, for any combination of population size, sampling scheme and distance function

$\phi(d)$	N	Scheme	AB			RMSE			ABRMSEE		
			Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
d^{-2}	1000	SRSWOR	0.24	2.96	0.76	0.28	3.29	0.88	0.00	1.92	0.16
		STRSPA	0.22	2.95	0.72	0.25	3.27	0.82	0.00	1.86	0.12
		DBALSS	0.19	2.92	0.67	0.22	3.27	0.75	0.00	2.02	0.11
	5000	SRSWOR	0.16	3.20	0.56	0.19	3.47	0.63	0.01	2.09	0.08
		STRSPA	0.16	3.22	0.55	0.18	3.48	0.61	0.00	2.08	0.07
		DBALSS	0.14	3.17	0.51	0.15	3.45	0.55	0.00	2.11	0.05
	10000	SRSWOR	0.15	3.13	0.50	0.18	3.35	0.56	0.00	1.76	0.06
		STRSPA	0.15	3.12	0.50	0.17	3.35	0.55	0.00	1.79	0.06
		DBALSS	0.12	3.07	0.46	0.14	3.30	0.50	0.00	1.73	0.04
d^{-3}	1000	SRSWOR	0.04	2.87	0.37	0.07	3.38	0.53	0.00	2.41	0.21
		STRSPA	0.03	2.81	0.34	0.05	3.38	0.46	0.00	2.37	0.16
		DBALSS	0.02	2.84	0.29	0.03	3.41	0.39	0.00	2.59	0.15
	5000	SRSWOR	0.01	3.13	0.18	0.02	3.58	0.25	0.00	2.66	0.09
		STRSPA	0.01	3.15	0.18	0.02	3.58	0.24	0.00	2.64	0.08
		DBALSS	0.00	3.07	0.14	0.01	3.55	0.18	0.00	2.69	0.06
	10000	SRSWOR	0.01	2.94	0.14	0.02	3.37	0.19	0.00	2.29	0.06
		STRSPA	0.01	2.95	0.14	0.01	3.38	0.18	0.00	2.31	0.06
		DBALSS	0.00	2.84	0.11	0.01	3.30	0.14	0.00	2.28	0.04
d^{-4}	1000	SRSWOR	0.00	2.87	0.26	0.03	3.49	0.43	0.00	2.74	0.24
		STRSPA	0.00	2.90	0.23	0.02	3.49	0.37	0.00	2.72	0.20
		DBALSS	0.00	2.87	0.20	0.01	3.53	0.31	0.00	2.92	0.18
	5000	SRSWOR	0.00	3.11	0.11	0.01	3.67	0.19	0.00	2.99	0.10
		STRSPA	0.00	3.13	0.10	0.01	3.67	0.18	0.00	2.97	0.09
		DBALSS	0.00	3.04	0.08	0.00	3.63	0.14	0.00	3.00	0.07
	10000	SRSWOR	0.00	2.90	0.08	0.00	3.46	0.13	0.00	2.66	0.07
		STRSPA	0.00	2.91	0.07	0.00	3.47	0.13	0.00	2.67	0.07
		DBALSS	0.00	2.83	0.06	0.00	3.40	0.10	0.00	2.64	0.05

TABLE 6

Minima, maxima and means of ABs, RMSEs and ABRMSEEs achieved with a sampling fraction of 10% for populations arising from F2 and RA, for any combination of population size, sampling scheme and distance function

$\phi(d)$	N	Scheme	AB			RMSE			ABRMSEE		
			Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
d^{-2}	1000	SRSWOR	0.17	3.04	0.67	0.23	3.27	0.82	0.00	1.52	0.16
		STRSPA	0.16	3.02	0.64	0.21	3.22	0.74	0.00	1.38	0.12
		DBALSS	0.10	3.05	0.59	0.14	3.26	0.67	0.00	1.74	0.11
	5000	SRSWOR	0.11	3.08	0.53	0.16	3.36	0.60	0.00	1.85	0.09
		STRSPA	0.11	3.10	0.52	0.15	3.36	0.59	0.00	1.82	0.08
		DBALSS	0.07	3.08	0.47	0.09	3.34	0.52	0.00	1.87	0.06
	10000	SRSWOR	0.11	3.26	0.47	0.15	3.49	0.54	0.00	1.91	0.07
		STRSPA	0.11	3.28	0.47	0.14	3.50	0.53	0.00	1.94	0.07
		DBALSS	0.08	3.27	0.43	0.10	3.48	0.47	0.00	1.92	0.05
d^{-3}	1000	SRSWOR	0.02	2.92	0.34	0.05	3.28	0.48	0.00	2.20	0.20
		STRSPA	0.02	2.88	0.30	0.04	3.21	0.41	0.00	2.08	0.15
		DBALSS	0.00	2.97	0.25	0.02	3.32	0.35	0.00	2.52	0.14
	5000	SRSWOR	0.00	2.89	0.18	0.02	3.36	0.25	0.00	2.34	0.09
		STRSPA	0.00	2.90	0.17	0.02	3.35	0.24	0.00	2.31	0.08
		DBALSS	0.00	3.02	0.13	0.01	3.46	0.18	0.00	2.56	0.06
	10000	SRSWOR	0.00	3.12	0.13	0.01	3.50	0.18	0.00	2.39	0.06
		STRSPA	0.00	3.16	0.13	0.01	3.53	0.17	0.00	2.40	0.06
		DBALSS	0.00	3.20	0.10	0.00	3.62	0.13	0.00	2.77	0.04
d^{-4}	1000	SRSWOR	0.00	2.87	0.23	0.02	3.37	0.40	0.00	2.61	0.23
		STRSPA	0.00	2.82	0.21	0.02	3.35	0.34	0.00	2.49	0.18
		DBALSS	0.00	3.00	0.18	0.01	3.47	0.29	0.00	2.84	0.16
	5000	SRSWOR	0.00	3.01	0.11	0.01	3.52	0.18	0.00	2.70	0.10
		STRSPA	0.00	3.02	0.10	0.01	3.51	0.18	0.00	2.68	0.09
		DBALSS	0.00	3.22	0.08	0.00	3.71	0.14	0.00	3.04	0.07
	10000	SRSWOR	0.00	3.19	0.07	0.00	3.69	0.12	0.00	2.96	0.07
		STRSPA	0.00	3.17	0.07	0.00	3.67	0.12	0.00	2.91	0.06
		DBALSS	0.00	3.40	0.05	0.00	3.89	0.09	0.00	3.35	0.05

TABLE 7
Minima, maxima and means of ABs, RMSEs and ABRMSEEs achieved with a sampling fraction of 10% for populations arising from F2 and TR, for any combination of population size, sampling scheme and distance function

$\phi(d)$	N	Scheme	AB			RMSE			ABRMSEE		
			Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
d^{-2}	1000	SRSWOR	0.05	2.75	0.72	0.10	2.96	0.87	0.00	1.58	0.17
		STRSPA	0.05	2.64	0.67	0.09	2.84	0.78	0.00	1.42	0.13
		DBALSS	0.02	2.72	0.61	0.03	2.91	0.70	0.00	1.71	0.11
	5000	SRSWOR	0.02	2.63	0.54	0.04	2.86	0.62	0.00	1.36	0.09
		STRSPA	0.02	2.63	0.53	0.04	2.86	0.60	0.00	1.34	0.08
		DBALSS	0.01	2.70	0.48	0.01	2.92	0.53	0.00	1.63	0.06
	10000	SRSWOR	0.02	2.78	0.49	0.02	2.99	0.56	0.00	1.35	0.07
		STRSPA	0.02	2.78	0.49	0.02	2.99	0.55	0.00	1.31	0.07
		DBALSS	0.01	2.83	0.44	0.01	3.03	0.49	0.00	1.59	0.05
d^{-3}	1000	SRSWOR	0.00	2.68	0.29	0.04	3.21	0.41	0.00	2.18	0.15
		STRSPA	0.00	2.66	0.33	0.04	3.15	0.49	0.00	2.17	0.20
		DBALSS	0.00	2.72	0.25	0.01	3.36	0.34	0.00	2.59	0.13
	5000	SRSWOR	0.00	2.91	0.16	0.01	3.27	0.23	0.00	2.33	0.08
		STRSPA	0.00	2.93	0.17	0.01	3.29	0.24	0.00	2.36	0.09
		DBALSS	0.00	2.96	0.13	0.00	3.32	0.17	0.00	2.62	0.06
	10000	SRSWOR	0.00	2.88	0.13	0.00	3.30	0.18	0.00	2.20	0.06
		STRSPA	0.00	2.87	0.13	0.00	3.29	0.18	0.00	2.20	0.06
		DBALSS	0.00	3.10	0.10	0.00	3.46	0.13	0.00	2.41	0.05
d^{-4}	1000	SRSWOR	0.00	2.71	0.22	0.04	3.38	0.40	0.00	2.61	0.22
		STRSPA	0.00	2.80	0.19	0.04	3.43	0.33	0.00	2.65	0.18
		DBALSS	0.00	2.83	0.16	0.01	3.57	0.27	0.00	3.06	0.15
		SRSWOR	0.00	2.95	0.10	0.01	3.46	0.18	0.00	2.79	0.10
		STRSPA	0.00	2.94	0.10	0.01	3.45	0.17	0.00	2.75	0.09
		DBALSS	0.00	3.00	0.08	0.00	3.53	0.13	0.00	2.98	0.07
	10000	SRSWOR	0.00	2.92	0.07	0.01	3.49	0.13	0.00	2.74	0.07
		STRSPA	0.00	2.95	0.07	0.01	3.50	0.13	0.00	2.75	0.07
		DBALSS	0.00	3.27	0.06	0.00	3.72	0.10	0.00	2.97	0.05

TABLE 8
Minima, maxima and means of ABs, RMSEs and ABRMSEEs achieved with a sampling fraction of 10% for populations arising from F2 and CL, for any combination of population size, sampling scheme and distance function

$\phi(d)$	N	Scheme	AB			RMSE			ABRMSEE		
			Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
d^{-2}	1000	SRSWOR	0.00	2.72	0.22	0.03	3.09	0.28	0.00	1.85	0.09
		STRSPA	0.00	2.77	0.20	0.02	3.09	0.24	0.00	1.83	0.07
		DBALSS	0.00	2.73	0.18	0.01	3.01	0.21	0.00	1.87	0.05
	5000	SRSWOR	0.00	2.63	0.15	0.01	2.88	0.18	0.00	1.26	0.03
		STRSPA	0.00	2.63	0.15	0.01	2.87	0.18	0.00	1.24	0.03
		DBALSS	0.00	2.67	0.13	0.00	2.88	0.15	0.00	1.25	0.02
	10000	SRSWOR	0.00	2.71	0.14	0.00	2.94	0.16	0.00	1.30	0.02
		STRSPA	0.00	2.69	0.14	0.00	2.92	0.15	0.00	1.27	0.02
		DBALSS	0.00	2.78	0.12	0.00	2.99	0.13	0.00	1.33	0.02
d^{-3}	1000	SRSWOR	0.00	2.75	0.13	0.01	3.24	0.19	0.00	2.30	0.10
		STRSPA	0.00	2.79	0.12	0.01	3.24	0.17	0.00	2.30	0.08
		DBALSS	0.00	2.73	0.10	0.01	3.15	0.14	0.00	2.38	0.06
	5000	SRSWOR	0.00	2.75	0.07	0.00	3.14	0.10	0.00	1.97	0.04
		STRSPA	0.00	2.75	0.07	0.00	3.13	0.09	0.00	1.95	0.04
		DBALSS	0.00	2.79	0.05	0.00	3.15	0.07	0.00	2.01	0.03
	10000	SRSWOR	0.00	2.89	0.05	0.00	3.27	0.07	0.00	2.11	0.03
		STRSPA	0.00	2.85	0.05	0.00	3.24	0.07	0.00	2.05	0.03
		DBALSS	0.00	3.09	0.04	0.00	3.41	0.05	0.00	2.23	0.02
d^{-4}	1000	SRSWOR	0.00	2.39	0.11	0.01	3.33	0.18	0.00	2.54	0.12
		STRSPA	0.00	2.77	0.10	0.01	3.33	0.16	0.00	2.56	0.10
		DBALSS	0.00	2.71	0.08	0.01	3.24	0.13	0.00	2.65	0.08
	5000	SRSWOR	0.00	2.77	0.05	0.00	3.27	0.08	0.00	2.40	0.04
		STRSPA	0.00	2.77	0.05	0.03	3.26	0.08	0.00	2.38	0.04
		DBALSS	0.00	2.88	0.04	0.00	3.28	0.06	0.00	2.46	0.03
	10000	SRSWOR	0.00	2.97	0.04	0.00	3.46	0.06	0.00	2.58	0.03
		STRSPA	0.00	2.93	0.04	0.00	3.42	0.06	0.00	2.53	0.03
		DBALSS	0.00	3.23	0.03	0.00	3.63	0.04	0.00	2.73	0.02

combination of mark function, sampling scheme, distance function, population size and spatial pattern.

6.4. Results

The simulation results confirmed the theoretical findings of Sections 3-5. Because F1 was continuous and points were generated to approach condition (4.1), for $\beta > 2$ all the sampling schemes ensured pointwise design consistency. Indeed, as N and n increased, a sharp decrease of the minima, maxima and averages of ABs and RMSEs occurred for $\beta > 2$. The decreases were less marked for $\beta = 2$, when consistency was not ensured by the distance function. As previously argued, the best performance was achieved with CL and TR patterns, where the clumping of points was more marked than in the case of RA and RE patterns.

Regarding the sampling schemes, spatial balance ensured the best performance, stated that DBALLS performed better than STRSPA that in turn performed better than SRSWOR (see Tables 1-4). The function F2 showed a discontinuity at the edge of the upper right quadrant of the unit square. Hence, pointwise design consistency was ensured only away from the discontinuity lines for $\beta > 2$ and for all the sampling schemes. Actually, sharp decreases occurred only for the minima and averages of ABs and RMSEs for $\beta > 2$, while maxima remained about constant as the sample size increased. Besides this fact, the results are quite similar to those achieved for F1. The decreases (where they occurred) were less marked for $\beta = 2$, the best performance was achieved in presence of CL and TR patterns while RE and RA patterns provided smaller precision and DBALLS performed best while SRSWOR performed worst (see Tables 5-8).

Regarding the RMSE estimation, the bias of (5.1) reduced as the sample size increased, confirming the theoretical results of Section 5. Regarding the β choice, the performance in terms of ABs and RMSEs tended to improve as β increased reaching its best for $\beta = 4$. On the other hand, the bias of the RMSE estimator tended to increase when passing from $\beta = 3$ to $\beta = 4$ especially for moderate population and sample sizes. Even if the issue needs for more theoretical and empirical investigations, a β choice between 3 and 4 seems to be a suitable compromise value.

7. Case study

A sample survey was performed to provide the map of tree heights in a seminatural mixed oak-hornbeam flood plain forest in the province of Mantova (North Italy). The forest stand was a rectangular area of size 70×140 m^2. Owing to the moderate size of the forest, 3P sampling was adopted to select trees, as customary in most forest surveys over small stands, thus allowing for the subsequent mapping of the tree heights. Because the purpose of the study was simply to produce the height map without giving explanation about the process that dislocated trees and generated heights, the complex task of modelling the marked point process and estimating its parameters was avoided and we opted to produce the maps in a design-based framework adopting the IDW interpolation.

As pointed out in the Introduction, the merit was that the resulting precision - being based on 3P sampling actually used to select trees - was real, not assumed or modelled. The survey proceeded in accordance with the following steps. In spring 2016 a crew of experts visited the population $\mathcal{U}$ of the $N = 510$ trees lying in the stand, recording their locations and giving a prediction x_j for the height of each tree $j \in \mathcal{U}$ (see Figure 1a). Then, 3P sampling was performed by independently selecting each tree with probability $\pi_j = x_j/M$ where M was an upper bound for heights that 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. The selected sample $\mathcal{S}$ was constituted by $n = 54$ trees whose heights y_j were measured for each $j \in \mathcal{S}$.

Subsequently, in order to perform the IDW interpolation, we realized that 3P sampling not only provided a scheme suitable to work with forest stands of moderate sizes exploiting predictions at design level, but predictions themselves should provide good proxies for the actual trees heights that could be exploited at estimation level. Indeed, for the sampled trees both predictions x_j and actual heights were known, in such a way that also the prediction errors $e_j = y_j - x_j$ were known for each $j \in \mathcal{S}$.

Therefore, in the spirit of the difference estimation criterion adopted to exploit proxies for improving the Horvitz-Thompson estimator of population totals (e.g. [22] Section 6.3), the prediction errors rather than the heights were interpolated by IDW, in such a way that the interpolated heights were given by the predicted heights plus the interpolated errors, i.e.

$$\widehat{y}_j = x_j + \widehat{e}_j, \quad j \in \mathcal{U}. \tag{7.1}$$

It should be noticed that (7.1) kept on being a genuine interpolator for y_j because when $j \in \mathcal{S}$, then $\widehat{e}_j = e_j$, in such a way that $\widehat{y}_j = x_j + \widehat{e}_j = y_j$. Moreover, the procedure was likely to better meet design consistency requirements, because the assumption of smoothness was even more realistic when considering prediction errors than heights. Indeed, jumps and irregularities in the tree heights throughout the stand were presumably absorbed by similar jumps and irregularities in the corresponding predictions.

Consequently, if predictions were good, prediction errors turned out to be more smoothed than heights. Practically speaking the interpolation proceeded in accordance with the following steps. Once the prediction errors e_js were computed from sample data for each $j \in \mathcal{S}$, the interpolated errors $\widehat{e}_j$ were computed for each $j \in \mathcal{U}$ by means of equation (2.1), where the e_js are used instead of the y_js and using the distance function $\phi(d) = d^{-3}$. Once the errors were interpolated, the interpolated heights were achieved by means of equation (7.1). Figures 1b and 1c show the maps of the interpolated errors and heights, respectively. Moreover, regarding the estimation of the precision, the uncertainty in the interpolated heights only stemmed from the uncertainty in the interpolated errors, the predictions being constant. Therefore, mean squared errors were estimated by means of equation (5.1), where once again the e_js are used instead of the y_js. Figure 1d shows the map of the estimated root mean squared errors.

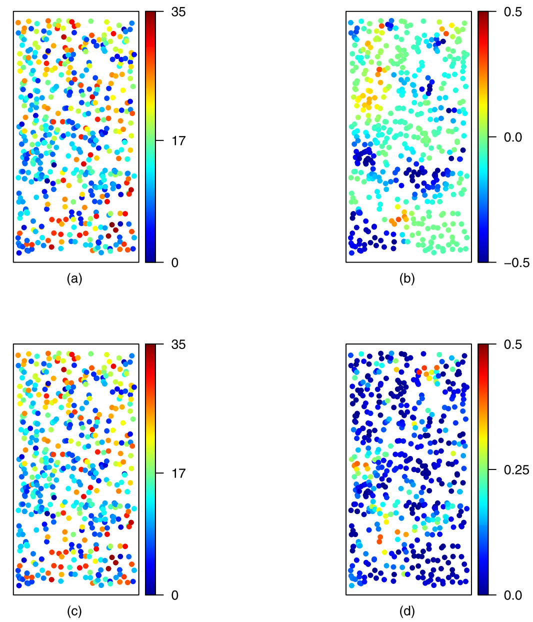

Fig 1. *Maps of the predicted heights (a), interpolated prediction errors (b), interpolated heights (c) and estimated root means squared errors (d).*

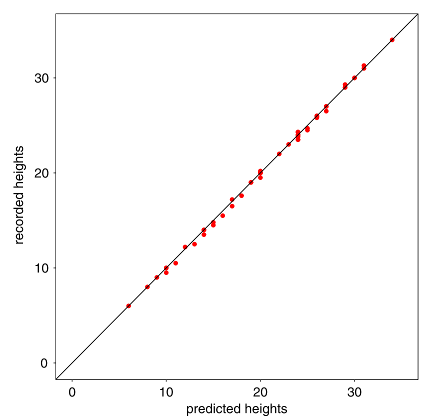

Fig 2. *Plot of the predicted heights vs the actual heights for the sampled trees.*

Some considerations are due. First of all, it should be noticed that the prediction errors can take negative values but they are obviously bounded by M, in such a way that all the consistency results hold. Moreover, at least for the sampled trees, the predicted heights turned out to be good proxies of the actual eights, as showed in Figure 2. In particular, predictions explained the 0.998 of the height variability in the sample. In turn, the goodness of predictions should entail smoothness of the prediction errors that are the quantities to be interpolated. Finally, the spatial pattern of trees throughout the stand did not evidence isolated individuals whose prediction would be problematic. Therefore, from the theoretical findings, but also at the light of the simulation results that showed the effectiveness of the IDW interpolation for populations increasing regularly with smoothed mark functions (i.e. the F1 case), the map of tree heights reported in Figure 1c is likely to be a reliable picture of the actual map.

8. Conclusions

Accurate and updated wall-to wall maps depicting the spatial pattern of ecological and economic attributes throughout the study area represents a crucial information for evaluations, decision making and planning. Traditionally maps, as well as most of the issues of spatial statistics, are approached in a model-based framework (e.g. [3]). Recently we have addressed map estimation in a complete design-based framework simply adopting the IDW interpolator and deriving its properties from the sampling scheme. We first approached this issue for finite population of spatial units, when the survey variable is the amount of an attribute within units ([7]). Subsequently, we considered map estimation for continuous populations when the survey variable is, at least in principle, defined at each point of the continuum representing the study area ([8]). Finally, in this paper we have faced map estimation for finite populations of marked points.

Maps of marked point populations are frequently possible in economic and social studies, where locations of units such as towns or firms are available from administrative sources or official data. On the other hand, because the list and locations of units in natural populations, such as communities of animals and plants, are prohibitive to achieve, maps rarely arise from ecological and environmental surveys. The sole occasions in which locations of natural point populations are available occur under 3P sampling, because the scheme entails visiting and hence recording locations for all the population units.

Our design-based approach to spatial mapping avoids the massive modelling involved in model-based approaches, i.e. the use of spatial models on lattices required for finite populations of spatial units (e.g. [3], Chapter 6), the use of second-order stationary spatial processes required for continuous populations (e.g. [3], Chapter 3) and the marked point processes in the plane required for finite populations of marked points (e.g. [3], Chapter 8). Design-based asymptotic unbiasedness and consistency of the resulting maps are achieved exploiting different asymptotic scenarios at the cost of supposing i) some forms of smoothness of the survey variables throughout the study area; ii) some sort of regularities that are necessary in the case of finite populations such as regularities in the shape of spatial units or regularities in the enlargements of the point populations; iii) asymptotically balanced spatial sampling schemes; iv) the use of distance functions sharing some mathematical properties. It is worth noting that iii) is satisfied by most of the more common sampling schemes adopted in spatial surveys and iv) does not constitute an assumption because it can be readily ensured by the user.

Finally, a quite surprising result should be emphasized. While for estimating totals or averages of marked point populations, aggregated or trended spatial patterns heavily deteriorate the design-based precision of the estimators with respect to randomly or regular patterns (e.g. [13]), the opposite occurs in map estimation. In this case both theoretical considerations and empirical studies suggest that aggregated and trended patterns provide the best results. Even if unexpected, this result has a clear explanation. Because aggregated and trended patterns tend to clump points more rapidly than regular and random patterns, distances between points tend to be smaller in the former cases, so that IDW interpolation tends to exploit nearer points, that, in case of smoothed mark functions, means exploiting more similar marks, thus providing a better fitting of the Tobler's law which assists inference.

Acknowledgements

The authors thank Piermaria Corona, Director of CREA Research Centre for Forestry and Wood, Arezzo (Italy) who kindly supplied the data for the case study.

Appendix A: proofs of the main results

A.1. *Proof of Lemma*

From the definition of IDW interpolator of y_j, for any $k \geq k(j)$ it follows that

$$
\begin{aligned}
|\hat{y}_j^{(k)} - y_j| &= |(1 - Z_j^{(k)}) \frac{\sum_{i \in \mathcal{U}_k} (y_i - y_j) Z_i^{(k)} \phi(d_{ij})}{\sum_{i \in \mathcal{U}_k} Z_i^{(k)} \phi(d_{ij})}| \\
&\leq (1 - Z_j^{(k)}) \frac{\sum_{i \in \mathcal{U}_k} |y_i - y_j| Z_i^{(k)} \phi(d_{ij})}{\sum_{i \in \mathcal{U}_k} Z_i^{(k)} \phi(d_{ij})} \\
&\leq \frac{\sum_{i \in \mathcal{U}_k} |y_i - y_j| Z_i^{(k)} \phi(d_{ij})}{\sum_{i \in \mathcal{U}_k} Z_i^{(k)} \phi(d_{ij})}.
\end{aligned}
$$

For any $\delta > 0$, $|y_i - y_j| \leq \Delta_j^{(k)}(\delta)$ when $i \in B_j^{(k)}(\delta)$, $\phi(d_{ij}) \leq \phi(\delta)$ when $i \notin B_j^{(k)}(\delta)$ and $|y_i - y_j| \leq L$ for any $i \in \mathcal{U}_k$. Therefore, denoted by $T_j^k = I\{Z_j^{(k)}(vN_k^{-\alpha}) = 0\}$, for any α, $v > 0$, it follows that

$$
\begin{aligned}
|\hat{y}_j^{(k)} - y_j| &\leq \Delta_j^{(k)}(\delta) M_{k,\delta} + L(1 - M_{k,\delta}) \qquad \text{(A.1)} \\
&\leq \Delta_j^{(k)}(\delta) + L(1 - M_{k,\delta})(1 - T_j^k) + L(1 - M_{k,\delta}) T_j^k \\
&\leq \Delta_j^{(k)}(\delta) + L\phi(\delta) \frac{\sum_{i \notin B_j^{(k)}(\delta)} Z_i^{(k)}}{\sum_{i \in \mathcal{U}_k} Z_i^{(k)} \phi(d_{ij})} (1 - T_j^k) + L T_j^k
\end{aligned}
$$

$$\leq \Delta_j^{(k)}(\delta) + \frac{L\phi(\delta)n_k}{\sum_{i\in\mathcal{U}_k} Z_i^{(k)}\phi(d_{ij})}(1-T_j^k) + LT_j^k$$

where $M_{k,\delta} = \frac{\sum_{i\in B_j^{(k)}(\delta)} Z_i^{(k)}\phi(d_{ij})}{\sum_{i\in\mathcal{U}_k} Z_i^{(k)}\phi(d_{ij})}$. Inequality (A.1) is obvious when $Z_j^{(k)} = 1$, because in this case $|\hat{y}_j^{(k)} - y_j| = 0$. On the other hand, when $Z_j^{(k)} = 0$ and $I\{Z_j^{(k)}(vN_k^{-\alpha}) > 0\}$, there is at least one $i \in \mathcal{U}_k$, $i \neq j$ such that $\phi(d_{ij}) \geq \phi(vN_k^{-\alpha})$. Therefore

$$\sum_{i\in\mathcal{U}_k} Z_i^{(k)}\phi(d_{ij}) \geq \phi(vN_k^{-\alpha}) \tag{A.2}$$

in such a way that inequality (A.1) reduces to

$$\begin{aligned}|\hat{y}_j^{(k)} - y_j| &\leq \Delta_j^{(k)}(\delta) + \frac{L\phi(\delta)n_k}{\phi(vN_k^{-\alpha})}(1-T_j^k) + LT_j^k \qquad \text{(A.3)}\\ &= \Delta_j^{(k)}(\delta) + L\beta_k\end{aligned}$$

where

$$\beta_k = \frac{\phi(\delta)n_k}{\phi(vN_k^{-\alpha})}(1-T_j^k) + T_j^k.$$

Accordingly, from (A.3)

$$\begin{aligned}(\hat{y}_j^{(k)} - y_j)^2 &\leq 2\{\Delta_j^{(k)}(\delta)\}^2 + 2L^2\beta_k^2\\ &= 2\{\Delta_j^{(k)}(\delta)\}^2 + 2L^2\Big[\frac{\phi^2(\delta)n_k^2}{\phi^2(vN_k^{-\alpha})}(1-T_j^k) + T_j^k\Big].\end{aligned}$$

Therefore, taking the expectation of both sides of the previous inequality it follows that

$$\begin{aligned}E\{(\hat{y}_j^{(k)} - y_j)^2\} &\leq 2\{\Delta_j^{(k)}(\delta)\}^2 + 2L^2\left[\frac{\phi^2(\delta)\,n_k^2}{\phi^2(vN_k^{-\alpha})}(1-E\{T_j^k\}) + E\{T_j^k\}\right]\\ &\leq 2\{\Delta_j^{(k)}(\delta)\}^2 + 2L^2\left[\frac{\phi^2(\delta)\,n_k^2}{\phi^2(vN_k^{-\alpha})} + E\{T_j^k\}\right]\end{aligned}$$

that proves the first part of Lemma.

Moreover, denoted by $\widetilde{T}_j^k = I\{\min_{j\in\mathcal{U}_k} Z_j^{(k)}\left(vN_k^{-\alpha}\right) = 0\}$, arguing as in the first part of the proof that leads to (A.1), it follows that

$$|\hat{y}_j^{(k)} - y_j| \leq \Delta_j^{(k)}(\delta) + \frac{L\phi(\delta)\,n_k}{\sum_{i\in\mathcal{U}_k} Z_i^{(k)}\phi(d_{ij})}(1-\widetilde{T}_j^k) + L\widetilde{T}_j^k.$$

Then, owing to (A.2)

$$\max_{j\in\mathcal{U}_k}|\hat{y}_j^{(k)} - y_j| \leq \max_{j\in\mathcal{U}_k}\Delta_j^{(k)}(\delta) + \frac{L\phi(\delta)n_k}{\sum_{i\in\mathcal{U}_k} Z_i^{(k)}\phi(d_{ij})}(1-\widetilde{T}_j^k) + L\widetilde{T}_j^k.$$

Therefore, the second part of Lemma follows arguing as in the final part of the proof regarding the first part.

A.2. Proof of Result 1

Let $\alpha = 1/2$. For any $\delta, v > 0$ and for any natural number j, from the first inequality of Lemma it follows that

$$E\{(\hat{y}_j^{(k)} - y_j)^2\} \leq 2\{\Delta_j^{(k)}(\delta)\}^2 + 2L^2\Big[\frac{\phi^2(\delta)n_k^2}{\phi^2(vN_k^{-1/2})} + \Pr\{Z_j^{(k)}(vN_k^{-1/2}) = 0\}\Big].$$

Owing to condition (3.6), for any p_j and any $\epsilon > 0$ there exist a real number $v > 0$ and an integer k_0 such that

$$\Pr\{Z_j^{(k)}(v_0N_k^{-1/2}) = 0\} < \epsilon \quad \forall\ k > k_0.$$

Moreover, from condition (3.8) it follows that

$$\lim_{k\to\infty} \frac{\phi^2(\delta)n_k^2}{\phi^2(vN_k^{-1/2})} = 0. \tag{A.4}$$

Therefore, there exists an integer k_0 such that

$$E\{(\hat{y}_j^{(k)} - y_j)^2\} \leq 2\{\Delta_j^{(k)}(\delta)\}^2 + 4L^2\varepsilon \quad \forall\ k > k_0.$$

Finally, if y is continuous at p_j, then owing to (3.3) there exists a $\delta > 0$ such that

$$\sup_{k>k(j)} \{\Delta_j^{(k)}(\delta)\}^2 < \varepsilon$$

from which there exists an integer k_0 such that

$$E\{(\hat{y}_j^{(k)} - y_j)^2\} \leq (2 + 4L^2)\varepsilon \quad \forall\ k > k_0$$

and, from the arbitrariness of ϵ, it follows that

$$\lim_{k\to\infty} E\{(\hat{y}_j^{(k)} - y_j)^2\} = 0$$

that obviously entails (3.1).

The uniform consistency with respect to $\{\mathcal{U}_k\}$ can be proven in a similar way by using the second inequality of Lemma, relation (A.4) and conditions (3.7) and (3.4).

A.3. Proof of Result 2

Owing to the Lipschitz condition (3.11), from the first inequality of Lemma and from the use of a distance function of type $\phi(d) = d^{-\beta}$, for any j, any $\delta > 0$ and for any k such that $k \geq k(j)$, it follows that

$$\begin{aligned} E\{(\widehat{y}_j^{(k)} - y_j)^2\} &\leq 2(C\delta)^2 + 2L^2\delta^{-2\beta}n_k^2v^{2\beta}N_k^{-2\alpha\beta} + 2L^2\Pr\{Z_j^{(k)}(vN_k^{-\alpha}) = 0\} \\ &\leq 2(C\delta)^2 + 2L^2\delta^{-2\beta}n_k^{2-2\alpha\beta}v^{2\beta} + 2L^2\Pr\{Z_j^{(k)}(vN_k^{-\alpha}) = 0\}. \end{aligned}$$

Therefore, for

$$\delta^2 = \Big(\frac{L^2 n_k^{2-2\alpha\beta} v^{2\beta}}{C^2}\Big)^{\frac{1}{1+\beta}}$$

it follows that

$$\begin{aligned} E\{(\hat{y}_j^{(k)} - y_j)^2\} &\leq 2C^2\Big(\frac{L^2 n_k^{2-2\alpha\beta} v^{2\beta}}{C^2}\Big)^{\frac{1}{1+\beta}} + 2L^2 \Pr\{Z_j^{(k)}(vN_k^{-\alpha}) = 0\} \\ &= Dn_k^{\frac{2(1-\alpha\beta)}{1+\beta}} + 2L^2 \Pr\{Z_j^{(k)}(vN_k^{-\alpha}) = 0\} \end{aligned}$$

where $D = 2C^{-\frac{2\beta}{1+\beta}}(L^2 v^{2\beta})^{\frac{1}{1+\beta}}$.

Finally, taking $\alpha = 1/2$ and $v^2 = \text{diam}(\mathcal{A})^2 n_k^{\frac{\beta-2}{2\beta+1}}$, from condition (3.12) it follows that $E\{(\hat{y}_j^{(k)} - y_j)^2\} \leq Dn_k^{\frac{2-\beta}{1+\beta}} + 2L^2 c_j \text{diam}(\mathcal{A})^{-2} n_k^{\frac{2-\beta}{1+2\beta}}$ that proves the result.

A.4. Proof of Result 3

For any random variable X taking natural values $\Pr(X = 0) \leq \frac{CV^2(X)}{1+CV^2(X)}$. Therefore

$$\Pr\{Z_j^{(k)}(vN_k^{-\frac{1}{2}}) = 0\} \leq \frac{CV^2\{Z_j^{(k)}(vN_k^{-\frac{1}{2}})\}}{1 + CV^2\{Z_j^{(k)}(vN_k^{-\frac{1}{2}})\}} \leq CV^2\{Z_j^{(k)}(vN_k^{-\frac{1}{2}})\}$$

in such a way that proving condition (3.6) is the same as proving that for a p_j and any arbitrary $\varepsilon > 0$ there exist an integer k_0 and a real number $v > 0$ such that

$$CV^2\{Z_j^{(k)}(vN_k^{-\frac{1}{2}})\} < \varepsilon \quad \forall\ k > k_0. \tag{A.5}$$

Let $\tau_{ih}^{(k)} = \frac{\pi_{ih}^{(k)}}{\pi_i^{(k)}\pi_h^{(k)}} - 1$. It is at once apparent that

$$\begin{aligned} Var\{Z_j^{(k)}(vN_k^{-\frac{1}{2}})\} &= \sum_{i\in B_j^{(k)}(vN_k^{-\frac{1}{2}})} \pi_i^{(k)}(1-\pi_i^{(k)}) + 2\sum_{h>i\in B_j^{(k)}(vN_k^{-\frac{1}{2}})} \tau_{ih}^{(k)}\pi_i^{(k)}\pi_h^{(k)} \\ &\leq \sum_{i\in B_j^{(k)}(vN_k^{-\frac{1}{2}})} \pi_i^{(k)} + 2\sum_{h>i\in B_j^{(k)}(vN_k^{-\frac{1}{2}})} \tau_{ih}^{(k)+}\pi_i^{(k)}\pi_h^{(k)} \\ &\leq \sum_{i\in B_j^{(k)}(vN_k^{-\frac{1}{2}})} \pi_i^{(k)} + 2\max_{h>i\in\mathcal{U}_k} \tau_{ih}^{(k)+} \sum_{h>i\in B_j^{(k)}(vN_k^{-\frac{1}{2}})} \pi_i^{(k)}\pi_h^{(k)} \end{aligned}$$

and

$$CV^2\{Z_j^{(k)}(vN_k^{-\frac{1}{2}})\} \leq \frac{\sum_{i\in B_j^{(k)}(vN_k^{-\frac{1}{2}})} \pi_i^{(k)}}{\{\sum_{i\in B_j^{(k)}(vN_k^{-\frac{1}{2}})} \pi_i^{(k)}\}^2}$$

$$
\begin{aligned}
&+\frac{2\max_{h>i\in\mathcal{U}_k}\tau_{ih}^{(k)+}\sum_{h>i\in B_j^{(k)}(\upsilon N_k^{-\frac{1}{2}})}\pi_i^{(k)}\pi_h^{(k)}}{\{\sum_{i\in B_j^{(k)}(\upsilon N_k^{-\frac{1}{2}})}\pi_i^{(k)}\}^2} \\
&=\frac{1}{\sum_{i\in B_j^{(k)}(\upsilon N_k^{-\frac{1}{2}})}\pi_i^{(k)}}+2\max_{h>i\in\mathcal{U}_k}\tau_{ih}^{(k)+} \\
&\leq\frac{1}{\gamma\mathrm{Card}\{B_j^{(k)}(\upsilon N_k^{-\frac{1}{2}})\}}+2\max_{h>i\in\mathcal{U}_k}\tau_{ih}^{(k)+}
\end{aligned}
\tag{A.6}
$$

where $\gamma=\inf_k\min_{i\in\mathcal{U}_k}\pi_i^{(k)}$. Since $\mathcal{V}$ is regular and $\gamma>0$ owing to condition (4.2), then there exist a real number $u>0$ and an integer k_0 such that

$$\gamma\mathrm{Card}\{B_j^{(k)}(uN_k^{-\frac{1}{2}})\cap\mathcal{U}_k\}>\frac{2}{\epsilon}\ \ \forall\ k>k_0.$$

Moreover, owing to condition (4.3) there exists an integer k_0 such that

$$\max_{h>i\in\mathcal{U}_k}\tau_{ih}^{(k)+}\ \leq\frac{\varepsilon}{4}\ \ \forall\ k>k_0$$

in such a way that

$$CV^2\{Z_j^{(k)}(\upsilon N_k^{-\frac{1}{2}})]\leq\varepsilon\ \ \forall\ k>k_0$$

that proves (A.5).

A.5. Proof of consistency under SRSWOR and STRSPA

Consider the sequence of SRSWOR designs in which a constant fraction $0<p<1$ of units is selected from each population $\mathcal{U}_k$ of size N_k. In this case $\pi_j^{(k)}=p$ and $\pi_{jh}^{(k)}=(p^2N_k-p)/(N_k-1)$ for each $h\neq j\in\mathcal{U}_k$ and each k. Because (4.2) holds for any sequence of designs such that $\pi_j^{(k)}\geq p>0$ for all j and k, and condition (4.3) is immediately verified under SRSWOR stated that $\pi_{jh}^{(k)}-\pi_j^{(k)}\pi_h^{(k)}$ is invariably negative, SRSWOR ensures pointwise consistency.

Point-wise consistency is ensured also by STRSPA in which the sequence of populations of size N_k is partitioned into L strata of increasing sizes N_{lk} for $l=1,\dots,L$ and a constant fraction p of units is selected from each stratum by means of SRSWOR. In this case $\pi_j^{(k)}=p$ and $\pi_{jh}^{(k)}-\pi_j^{(k)}\pi_h^{(k)}$ turns out to be invariably negative if j and h belong to the same stratum or zero if they belong to different strata.

A.6. Proof of consistency under 3P sampling

Consider the sequence of 3P designs in which a random number n_k of units is selected from each population $\mathcal{U}_k$ of size N_k in such a way that each unit in the population enters the sample with probability $\pi_j^{(k)} = x_j/M$ for any $j \in \mathcal{U}_k$, where M is an upper bound of the survey variable Y ensuring that any acceptable prediction x_j be smaller than M so that $\pi_j^{(k)} \leq 1$ for any $j \in \mathcal{U}_k$ and any k. If there exists also a lower bound for Y, say $l > 0$, then $\pi_j^{(k)} \geq l/M > 0$ for any $j \in \mathcal{U}_k$ and any k. Therefore condition (4.2) holds and condition (4.3) is immediately verified stated that $\pi_{jh}^{(k)} = \pi_j^{(k)}\pi_h^{(k)}$ owing to the independence of drawings and hence $\pi_{jh}^{(k)} - \pi_j^{(k)}\pi_h^{(k)}$ is invariably zero. It is worth noting that a lower bound for the survey variable Y naturally arises in most forest and environmental surveys in which units with a mark (e.g tree height or basal area) smaller than a given threshold are not considered in the population.

A.7. Proof of Result 4

Owing to relation (A.6) and from condition (4.5), it follows that

$$\Pr\{Z_j^{(k)}(\upsilon N_k^{-\frac{1}{2}}) = 0\} \leq \frac{1}{\gamma \mathrm{Card}\{B_j^{(k)}(\upsilon N_k^{-\frac{1}{2}})\}} \leq \frac{1}{\gamma \mathrm{Card}\{B_j^{(k)}(\upsilon N_k^{-\frac{1}{2}}) \bigcap \mathcal{U}_k\}}.$$

Therefore, owing to condition (4.4) there exists a real number $c_j > 0$ such that

$$\Pr\{Z_j^{(k)}(\upsilon N_k^{-\frac{1}{2}}) = 0\} \leq (\gamma c_j)^{-1}\upsilon^{-2} \qquad \forall\ k > k(j),\ 0 < \upsilon \leq N_k^{1/2}\mathrm{diam}(\mathcal{A}).$$

Then, condition (3.12) is proven, that in turns, if y is a Lipschitz function on $\mathcal{V}$, implies Result 2.

A.8. Proof of Result 5

Since $(\tilde{y}_j^{(k)} - y_j)^2$ and $\hat{V}_j^{(k)}$ are bounded by L^2, from the Hölder-Schwarz inequality it follows that

$$\begin{aligned} MSE(\hat{y}_j^{(k)}) &= E\{\hat{V}_j^{(k)}\} + E\{(\tilde{y}_j^{(k)} - y_j)^2\} + 2E\{(\hat{y}_j^{(k)} - \tilde{y}_j)(\tilde{y}_j^{(k)} - y_j)\} \\ &\leq E\{\hat{V}_j^{(k)}\} + E\{(\tilde{y}_j^{(k)} - y_j)^2\} + 2[E\{\hat{V}_j^{(k)}\}]^{1/2}[E\{(\tilde{y}_j^{(k)} - y_j)^2\}]^{1/2} \\ &\leq E\{\hat{V}_j^{(k)}\} + L[E\{(\tilde{y}_j^{(k)} - y_j)^2\}]^{1/2} + 2L[E\{(\tilde{y}_j^{(k)} - y_j)^2\}]^{1/2} \end{aligned}$$

that proves (5.3).

References

[1] Avery, T. E. and Burkhart, H. E. (2002). *Forest Measurements*. 5th ed. New York: McGraw-Hill.

[2] Bruno, F., Cocchi D. and Vagheggini, A. (2013). Finite population properties of individual predictors based on spatial pattern. *Environ. Ecol. Stat.*, **20**, 467–494. MR3097174

[3] Cressie, N. (1993). *Statistics for spatial data*. New York: Wiley. MR1239641

[4] Falkowski, M. J., Wulder, M. A., White, J. C. and Gillis, M. D. (2009). Supporting large-area, sample-based forest inventories with very high spatial resolution satellite imagery. *Prog. Phys. Geog.*, **33**, 403–423.
[5] Fattorini, L. (2006). Applying the Horvitz–Thompson criterion in complex designs: a computer-intensive perspective for estimating inclusion probabilities. *Biometrika*, **93**, 269–278. MR2278082
[6] Fattorini, L. (2009). An adaptive algorithm for estimating inclusion probabilities and performing the Horvitz–Thompson criterion in complex designs. *Computation. Stat.*, **24**, 623–639. MR2566425
[7] Fattorini, L., Marcheselli, M. and Pratelli, L. (2018a). Design-based maps for finite populations of spatial units. *J. Am. Stat. Assoc.*, **113**, 686–697. MR3832219
[8] Fattorini, L., Marcheselli, M., Pisani, C. and Pratelli, L. (2018b). Design-based maps for continuous spatial populations. *Biometrika*, **105**, 419–429. MR3804411
[9] Gavrikov, V. and Stoyan, D. (1995). The use of marked point processes in ecological and environmental forest studies. *Environ. Ecol. Stat.*, **2**, 331–344.
[10] Grafström, A. (2012). Spatial correlated Poisson sampling. *J. Stat. Plan. Infer.*, **142**, 139–147. MR2827136
[11] Grafström, A., Lundström, N. L. P. and Schelin L. (2012). Spatially balanced sampling through the pivotal method. *Biometrics*, **68**, 514–520. MR2959618
[12] Grafström, A. and Tillé, Y. (2013). Doubly balanced spatial sampling with spreading and restitution of auxiliary totals. *Environmetrics*, **24**, 120–131. MR3067336
[13] Gregoire, T. G. and Valentine, H. T. (2008). *Sampling Strategies for Natural Resources and the Environment*. Boca Raton: Chapman and Hall. MR2510314
[14] Holden, L., Sannan, S. and Bungum, H. (2003). A stochastic marked point process model for earthquakes. *Nat. Hazard Earth Sys.*, **3**, 95–101.
[15] Iles, K. (2003). *A sampler of Inventory Topics*. Nanimo: Kim Iles & Associates.
[16] Illian, J., Penttinen, A., Stoyan, H. and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. Chichester: Wiley. MR2384630
[17] Isaki, C. T. and Fuller, W. A. (1982). Survey design under the regression superpopulation model. *J. Am. Stat. Assoc.*, **77**, 89–96. MR0648029
[18] Karr, A. F. (1986). Inference for stationary random fields given Poisson samples. *Adv. Appl. Probab.*, **18**, 406-422. MR0840101
[19] Lai Ping Ho and Stoyan, D. (2008). Modelling marked point patterns by intensity-marked Cox processes. *Stat. Probabil. Lett.*, **78**, 1194–1199. MR2441462
[20] Mase, S. (1996). The threshold method for estimating total rainfall. *Ann. I. Stat. Math.*, **48**, 201–213. MR1405927
[21] Penttinen, A., Stoyan, D. and Henttonen, H. M. (1992). Marked point processes in forest statistics. *Forest Sci.*, **38**, 806–824.

[22] Särndal, C. E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer Verlag. MR1140409

[23] Shiver, B. D. and Borders, B. E. (1996). *Sampling Techniques for Forest Inventory*. New York: Wiley.

[24] Stevens, D. L. and Olsen, A. R. (2004). Spatially balanced sampling of natural resources. *J. Am. Stat. Assoc.*, **99**, 262–278. MR2061889

[25] Tillé Y., Dickson, M. M., Espa, G. and Giuliani, D. (2018). Measuring the spatial balance of a sample: A new measure based on Moran's I index. *Spat. Stat. Neth.*, **23**, 182–192. MR3768182

[26] Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.*, **46**, 234–240.

[27] Vagheggini, A., Bruno, F. and Cocchi, D. (2016). A competitive design-based spatial predictor. *Environmetrics*, **27**, 454–465. MR3580912

[28] Van Laar, A. and Akça, A. (2007). *Forest Mensuration.* Dordrecht: Springer.